

# UNIVERSIDAD DE CANTABRIA

PROGRAMA DE DOCTORADO EN  
BIOLOGÍA MOLECULAR Y BIOMEDICINA



## TESIS DOCTORAL

### **Análisis descriptivo y funcional de las colonias microbianas visibles que crecen en la cueva de Altamira, enfocado al diseño de medidas de control**

Realizada por:

**Candela González-Riancho Fernández**

Dirigida por:

Juan María García Lobo

Escuela de Doctorado de la Universidad de Cantabria

**Santander 2021**









TESIS DOCTORAL  
ANÁLISIS DESCRIPTIVO Y FUNCIONAL DE LAS COLONIAS MICROBIANAS  
QUE CRECEN EN LA CUEVA DE ALTAMIRA, ENFOCADO AL DISEÑO DE  
MEDIDAS DE CONTROL

Memoria presentada por Candela González-Riancho Fernández:  
Licenciada en Biotecnología egresada de la Universidad de  
Francisco de Vitoria; para optar al grado de Doctora en Biología  
Molecular por la Universidad de Cantabria.

Candela González-Riancho Fernández

Autora

Juan María García-Lobo

Director y Tutor

Santander 1 de Marzo 2021



Esta tesis doctoral fue desarrollada en el grupo de Biología Molecular de la Patogenicidad de *Brucella* (Dir. Juan María García Lobo), en el departamento de Microbiología y Genómica del Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), que es un centro mixto de titularidad compartida entre la Universidad de Cantabria (UC), el Consejo Superior de Investigaciones Científicas (CSIC) y el Gobierno Regional de Cantabria a través de la sociedad para el desarrollo de Cantabria (SODERCAN).

El acceso a la cueva de Altamira y la toma de muestras necesarias para este trabajo se ha hecho con el permiso de la Comisión de Seguimiento del PCP, bajo la supervisión y colaboración del personal del Museo y Centro de Investigación de Altamira y en el contexto del contrato “SERVICIO DE CONTROL Y SEGUIMIENTO DE LAS CONDICIONES AMBIENTALES Y DEL BIODETERIORO DE LA CUEVA DE ALTAMIRA” suscrito entre la Universidad de Cantabria y el Ministerio de Cultura y Deporte.

Además, esta tesis ha tenido el apoyo financiero del proyecto “Aplicación de tecnologías OMICAS a la resolución de problemas microbiológicos en la cueva de Altamira” dentro de la convocatoria “Proyectos Puente 2018” cofinanciada por la Universidad de Cantabria y SODERCAN.



*“En cuestiones de ciencia, la autoridad de miles no vale más que el humilde razonamiento de un único individuo ”*

*Galileo Galilei*



# AGRADECIMIENTOS

Mis primeras palabras de agradecimiento son sin dudar dirigidas a mi gran director de tesis Juan M<sup>a</sup> García Lobo. Estoy agradecida de haber podido aprender de una magnífica persona en todos los aspectos. Quiero agradecerte por darme la oportunidad de conocer el mundo de la investigación que tanto me apasiona y de haber aprendido tantas y tantas y tantas cosas. Su pasión por la ciencia, su ambición por conocer, su tenacidad de trabajo y su gran apoyo me han ayudado en este largo camino. Un sensei es un término que se designa a una persona docta que atesora maestría y es lo que Juanma ha significado para mí estos años. En lo personal, me has apoyado y me has dado ánimos, y sorprendentemente has tenido tiempo de enseñarme nuevas cosas de tu diccionario antediluviano o de tu filmografía personal. Con todos los cargos, responsabilidades y obligaciones bajo la dirección de tu nombre, siempre has encontrado tiempo para recibirme y escucharme, y soy consciente que esto no es fácil. Por todo, desde esta persona que tanto te valora, MUCHAS GRACIAS.

Mi siguiente agradecimiento va dirigido con grandes y grandes palabras de gratitud y reconocimiento por el trabajo de dos personas que han facilitado la realización de esta tesis. Una de estas personas es Mónica Molina Edesa o SUPERMÓSNICA para mí. Has sido el viento que ha podido mover la vela de este trabajo de tesis y además en lo personal eres una persona de gran corazón de la que se puede aprender muchas cosas y con la que me he reído tantas veces. Y la otra persona a la que también agradezco tanto es a María Cruz Rodríguez González del servicio de secuenciación del IBBTEC, cuyo trabajo también ha sido encomiable para esta tesis. MUCHAS GRACIAS A LAS DOS.

Otra gran persona y tan importante en mi trayectoria científica, y quizás lejana en esta tesis pero iniciador indirectamente de ella es Jesús Agüero Balbin. Mis primeros pasos en la ciencia real fueron contigo cuando estaba en la universidad. Fueron tus consejos, tus ganas de enseñar y tu gran dedicación que me impresionaron para meterme en este mundo. A día de hoy, sigues

apoyándome y dándome ánimos tanto en lo personal como en lo profesional.  
GRACIAS a ti también.

Esta tesis tampoco habría sido posible sin la ayuda de todo el personal del Museo de Altamira, en especial de Carmen de las Heras, Pilar Fatás y Alfredo Prada. Gracias por facilitarme en todo momento la realización de esta tesis y vuestra disponibilidad. El trabajo que hacéis por la preservación de la cueva es excepcional, al igual que la ayuda que facilitáis a los demás profesionales que trabajamos también en la cueva. GRACIAS.

Quiero agradecer a mi compañera de laboratorio Yelina Ortiz por todo lo que hemos vivido en este tiempo. Has sido doctorada hace poco y tus consejos me han servido y ayudado bastante. Además, has estado durante este largo viaje ahí y por ello GRACIAS. También quiero agradecer a compañeros vecinos de otros laboratorios entre los que nos vamos apoyando en este viaje, algunos ya doctores y otros en ello, y que entre todos nos damos fuerzas y ánimos.  
GRACIAS a todos vosotros.

Quiero agradecer también al grupo de Ramón Rosselló Móra por su cálido recibimiento cuando llegué a su laboratorio para una corta estancia con ganas de aprender de su calidad de trabajo. Han sido todos ellos Tomeu, Carlota, Xisca y Joan quienes desde un primer momento me han hecho sentirme muy cómoda en el grupo, y siempre disponibles con una naturalidad tan sana de ayudarme y guiarme. Y no sólo en el laboratorio, sino fuera también desde el primer día que pisé Mallorca. Sois unas personas de admirar y el ambiente en el grupo es inmejorable. MUCHAS GRACIAS.

Y, por último, quiero agradecer a uno de los pilares principales en mi mundo, mi FAMILIA. A mis padres, siempre habéis estado ahí para lo que he y no he necesitado y seguiréis estando ciegamente. Sois un ejemplo para mi de trabajo y dedicación en esta vida y no tengo palabras para poder agradeceros todo lo que habéis hecho. A mis hermanos por estar ahí apoyando y ayudándome a desconectar cada día, algo tan necesario en este viaje. Os quiero y MUCHAS MUCHAS GRACIAS.

# TABLA DE CONTENIDO

<b>1. RESUMEN</b>	<b>1</b>
<b>2. INTRODUCCIÓN</b>	<b>9</b>
<b>2.1 LA CUEVA DE ALTAMIRA. UN BIEN CULTURAL EXCEPCIONAL</b>	<b>11</b>
2.1.1 DESCRIPCIÓN FÍSICA DE LA CUEVA	13
2.1.2 GESTIÓN DE LA CUEVA DE ALTAMIRA Y ALTERACIONES SUFRIDAS	20
<b>2.2 MICROBIOLOGÍA DE CUEVAS</b>	<b>24</b>
2.2.1 ESTRATEGIAS DE SUPERVIVENCIA DE LOS MICROORGANISMOS EN LAS CUEVAS	28
2.2.2 CASOS DE COLONIAS VISIBLES EN CUEVAS	31
2.2.3 COMUNIDADES MÁS FRECUENTES EN CUEVAS	36
2.2.4 TÉCNICAS DE CULTIVO DE BACTERIAS DIFÍCILES DE CRECER	43
2.2.5 LA IMPLICACIÓN NEGATIVA DE LA PROLIFERACIÓN DE MICROORGANISMOS EN CUEVAS	48
<b>2.3 MICROBIOLOGÍA SIN CULTIVO</b>	<b>54</b>
2.3.1 ANÁLISIS DE AMPLICONES	55
2.3.2 ANÁLISIS DE DNA TOTAL	59
2.3.3 AMPLIFICACIÓN DE GENOMAS DE CÉLULAS ÚNICAS.	65
<b>3. HIPÓTESIS Y OBJETIVOS</b>	<b>67</b>
3.1 HIPÓTESIS DE TRABAJO	69
3.2 OBJETIVOS DEL ESTUDIO DE TRABAJO	71
<b>4. MATERIALES Y MÉTODOS</b>	<b>73</b>
4.1 TOMA DE MUESTRAS	75
4.2 AMPLIFICACIÓN POR PCR Y SECUENCIACIÓN DEL GEN 16S rRNA	76
4.3 PROCESAMIENTO BIOINFORMÁTICO DE LOS DATOS DE AMPLICONES	78
4.4 PREPARACIÓN DE LIBRERÍAS Y SECUENCIACIÓN DEL DNA TOTAL	79
4.5 TRIMADO, ENSAMBLADO Y BINEADO DE LAS LECTURAS DE LOS METAGENOMAS; Y ANÁLISIS DE LA DIVERSIDAD	80
4.6 ASIGNACIÓN TAXONÓMICA DE LOS MAGS Y COMPARACIONES DE SECUENCIA DE DNA Y PROTEÍNA DE LOS COMPONENTES DE LAS COLONIAS	82
4.7 ABUNDANCIA RELATIVA DE LOS MAGs.	83
4.8 CONSTRUCCIÓN DE ÁRBOLES FILOGENÉTICOS BASADOS EN SECUENCIAS DE 16S rRNA DE LOS COMPONENTES MAYORITARIOS	84
4.9 DISEÑO DE ENSAYOS PARA LA DETECCIÓN DE COMPONENTES MAYORITARIOS DE LAS COLONIAS MEDIANTE EL DISEÑO DE PRIMERS ESPECÍFICOS	84
4.10 ANÁLISIS FUNCIONAL DE LOS METAGENOMAS	85
4.11 CONDICIONES DE CRECIMIENTO DE LOS AISLADOS DE COLONIAS VISIBLES	86
<b>5. RESULTADOS</b>	<b>89</b>
5.1 RESULTADOS TÉCNICA DE AMPLICONES	91
5.1.1 MÉTRICA GENERAL	91

5.1.2 ANÁLISIS DE LA BIODIVERSIDAD .....	92
5.1.3 ASIGNACIÓN TAXONÓMICA DE LAS OTUS.....	97
<b>5.2 RESULTADOS TÉCNICA DE SHOTGUN .....</b>	<b>114</b>
5.2.1 MÉTRICA GENERAL .....	114
5.2.2 ANÁLISIS DE LA BIODIVERSIDAD .....	119
5.2.3 ANÁLISIS TAXONÓMICO DE LOS MAGs .....	123
5.2.4 ANÁLISIS FILOGENÉTICO DE LOS COMPONENTES MAYORITARIOS DE LAS COLONIAS.....	149
5.2.5 MÉTODOS PARA DETECCIÓN DE COMPONENTES MAYORITARIOS DE LAS COLONIAS MEDIANTE EL DISEÑO DE PRIMERS.....	162
5.2.6 CAPACIDAD CODIFICANTE Y ANÁLISIS FUNCIONAL DE LOS METAGENOMAS .....	170
<b>5.3 CULTIVO DE COMPONENTES DE LAS COLONIAS .....</b>	<b>173</b>
<b><u>6. DISCUSIÓN.....</u></b>	<b><u>179</u></b>
6.1 LA CUEVA DE ALTAMIRA Y SU SITUACIÓN ACTUAL .....	181
6.2 COMPOSICIÓN DE LAS COLONIAS .....	183
6.3 NOVEDAD TAXONÓMICA DE LOS COMPONENTES DE LAS COLONIAS .....	195
6.4 LA FORMACIÓN DE COLONIAS VISIBLES COMO SOLUCIÓN UNIVERSAL PARA LOS MICROORGANISMOS DE CUEVAS. ....	197
6.5 EL METABOLISMO DE LAS COLONIAS DE LA CUEVA DE ALTAMIRA .....	199
6.6 RECUPERACIÓN EN CULTIVO PURO DE LAS BACTERIAS DE ALTAMIRA .....	209
<b><u>7. CONCLUSIONES.....</u></b>	<b><u>213</u></b>
<b><u>8. REFERENCIAS.....</u></b>	<b><u>217</u></b>
<b><u>9. ANEXO .....</u></b>	<b><u>235</u></b>

# Índice de figuras

<b>Figura 1.</b> Plano aéreo de la situación de la cueva de Altamira.....	11
<b>Figura 2.</b> Representación de arte rupestre paleolítico en la cornisa cantábrica.....	12
<b>Figura 3.</b> Plano actual de la cueva de Altamira.....	13
<b>Figura 4.</b> Variación de la temperatura para el exterior de la cueva y cada sala de la cueva durante los meses de Julio 2013 a Enero de 2017.....	14
<b>Figura 5.</b> Ejemplo de la variación anual de CO <sub>2</sub> en la cueva de Altamira durante el periodo de Marzo 2013 a Agosto del 2015.....	15
<b>Figura 6.</b> Ejemplo de la variación anual de Radón durante el periodo de 2013-2019 en cada una de las salas de la cueva de Altamira.....	16
<b>Figura 7.</b> Ejemplo de la variación anual de la humedad relativa en el año 2019 en el interior de la cueva de Altamira.....	16
<b>Figura 8.</b> Simulación de la implicación directa de las aguas de lluvia y escorrentía en la cobertera de la cueva.....	18
<b>Figura 9.</b> Acumulación de nitratos en puntos de goteo constantes en la cueva de Altamira para las salas de Cruce, Hoya y Pozo durante los periodos de año de Enero del 2003 a Julio del 2009.....	19
<b>Figura 10.</b> Evolución del plano original de la cueva actual tras las modificaciones sufridas hasta el día de hoy.....	21
<b>Figura 11.</b> Distribución de las regiones kársticas en la superficie terrestre.....	24
<b>Figura 12.</b> Ejemplo de los diferentes tipos de espeleotemas que resultan de la precipitación del carbonato cálcico en el interior de una cueva.....	25
<b>Figura 13.</b> Ejemplo de colonias visibles (amarillas, grises y blancas) en las paredes de la cueva de Altamira.....	31
<b>Figura 14.</b> Situación de la distribución de las colonias visibles en el interior de la cueva de Altamira.....	32
<b>Figura 15.</b> Estudios microbiológicos publicados sobre colonias visibles en cuevas de España.....	34
<b>Figura 16.</b> Estudios microbiológicos publicados sobre colonias visibles en ambientes de cuevas y otros entornos parecidos en el mundo.....	35

<b>Figura 17.</b> Estudios microbiológicos en cuevas del mundo.....	38
<b>Figura 18.</b> Fotografía de rizomorfos en la cueva de Altamira en muros artificiales.....	41
<b>Figura 19.</b> Imágenes obtenidas de la alga especie <i>Jenufa aeroterrestica</i> detectada en zonas verdosas en la cueva de Altamira.....	42
<b>Figura 20.</b> Ejemplos de cocultivo entre bacterias cultivables y bacterias VBNC. ....	45
<b>Figura 21.</b> Ejemplo de un dispositivo “ichip” situado en el ambiente para el cultivo de microorganismos incultivables de suelo.....	46
<b>Figura 22.</b> Ejemplo del diseño de un biorreactor de difusión para el cultivo de microorganismos incultivables, en este caso de suelo.....	47
<b>Figura 23.</b> Ejemplo de biodeterioro en la cueva de Lascaux (Francia).....	50
<b>Figura 24.</b> Zonas afectadas por la presencia de colonias blancas y grises en la sala de Polícromos.....	51
<b>Figura 25.</b> Representación de las diferentes regiones conservadas, variables e hipervariables que presenta el gen 16S rDNA .....	56
<b>Figura 26.</b> Pipeline del procesamiento de las lecturas de amplicones.....	57
<b>Figura 27.</b> Pipeline de los pasos de procesamiento de las lecturas por la técnica de shotgun.....	60
<b>Figura 28.</b> Enfoques empleados en la asignación taxonómica de los borradores de genomas (MAGs) obtenidos en este estudio metagenómico.....	64
<b>Figura 29.</b> Curvas de rarefacción para las muestras de colonia visibles y control analizados en la técnica de shotgun.....	93
<b>Figura 30.</b> Representación de los valores de $\alpha$ -diversidad en las muestras de colonia y control analizado por la técnica de amplicones.....	95
<b>Figura 31.</b> $\beta$ -Diversidad en las muestras de colonia y control analizadas por la técnica de amplicones .....	97
<b>Figura 32.</b> Representación de los filos más abundantes (con una abundancia >1%) en las muestras de las colonias (A) y control (B).....	98
<b>Figura 33.</b> Lista de las 10 OTUs con mayor frecuencia en las muestras de colonias visibles y en el control analizadas por la técnica de amplicones.....	100

<b>Figura 34.</b> Heatmap de abundancia de las OTUs que conforman el core en las muestras de colonias visibles y control de la cueva de Altamira analizadas por la técnica de amplicones.....	105
<b>Figura 35.</b> Comparación de los filos más representativos para las muestras de pared y suelo y muestras de colonia en distintos puntos de la cueva de Altamira analizadas por la técnica de amplicones.....	110
<b>Figura 36.</b> Lista de las 10 OTUs con mayor frecuencia en las muestras de suelo y pared en comparación con las muestras de colonias visibles en la cueva de Altamira analizadas por la técnica de amplicones.....	112
<b>Figura 37.</b> Mapabilidad de las lecturas para cada metagenoma.....	116
<b>Figura 38.</b> Abundancia relativa de los MAGs en cada metagenoma.....	117
<b>Figura 39.</b> Representación de la calidad en los MAGs.....	118
<b>Figura 40.</b> Curvas de rarefacción en los metagenomas de colonias visibles.....	121
<b>Figura 41.</b> $\beta$ -Diversidad presente en los metagenomas de colonias visibles.	122
<b>Figura 42.</b> Valores ANI para los MAGs que agruparon en el cluster I contra los genomas disponibles en la familia <i>Nitrosopumilaceae</i> .....	126
<b>Figura 43.</b> Valores ANI para los MAGs que agruparon en el cluster II contra los genomas disponibles de la clase <i>Blastocatellia</i> .....	128
<b>Figura 44.</b> Valores ANI para los MAGs que agruparon en el cluster III contra los genomas más cercanos disponibles dentro del orden <i>Pseudonocardiales</i> ....	129
<b>Figura 45.</b> Valores ANI para los MAGs que agruparon en el cluster VII contra los genomas más cercanos disponibles dentro del género <i>Ferrovum</i> .....	132
<b>Figura 46.</b> Valores ANI para los MAGs que agruparon en el cluster VIII contra los genomas más cercanos disponibles dentro del género <i>Lysobacter</i> .....	134
<b>Figura 47.</b> Valores ANI para los MAGs que agruparon en el cluster IX contra el genoma de referencia más cercano ( <i>Chromatiales bacterium USCg_Taylor</i> ).....	135
<b>Figura 48.</b> Valores ANI para los MAGs que agruparon en el cluster X contra los genomas más cercanos disponibles dentro del género <i>Nitrospira</i> .....	136
<b>Figura 49.</b> Valores ANI para los MAGs que agruparon en el cluster XI contra los genomas más cercanos disponibles dentro del filo Candidate division NC10.....	138

<b>Figura 50.</b> Representación de la secuencia del 16S rRNA identificado en el MAG006 en colonias G1 que clasifica con una <i>Parcubacteria</i> .....	139
<b>Figura 51.</b> Salida gráfica de la comparación con Blastn del 16S rRNA (nativo) del MAG006 contra secuencias más cercanas en la base de datos nr/nt del NCBI.....	140
<b>Figura 52.</b> Representación de la secuencia del 23S rRNA correspondiente al MAG006 en colonias G1 que clasifica con una <i>Parcubacteria</i> .....	142
<b>Figura 53.</b> Valores ANI para los MAGs que agruparon en el cluster XII contra los genomas más cercanos disponibles de los fillos <i>Candidatus</i> Nomurabacteria y Campbellbacteria.....	143
<b>Figura 54.</b> Valores ANI para los MAGs que agruparon en el cluster XIII contra los genomas más cercanos disponibles dentro del filo <i>Candidatus</i> Dadabacteria.....	144
<b>Figura 55.</b> Valores ANI del MAG055 contra los ensamblados de referencia de la especie <i>Inquilinus</i> disponibles en la base de datos “genoma” del NCBI y también el genoma obtenido de la secuenciación del genoma del cultivable recuperado del cultivo de colonias.....	147
<b>Figura 56.</b> Valores de ANI y AAI de los componentes mayoritarios en las colonias que agruparon en el cluster III.....	153
<b>Figura 57.</b> Comparación de las secuencias del 16S rRNA, valores ANI y AAI para los componentes mayoritarios agrupados en el cluster III.....	154
<b>Figura 58.</b> Árbol de máxima verosimilitud que muestra las relaciones filogenéticas de los genes del 16S rRNA de los componentes mayoritarios agrupados en el cluster III con las secuencias disponibles de los géneros descritos en la familia <i>Pseudonocardiaceae</i> .....	155
<b>Figura 59.</b> Distancias estimadas entre los genomas de los componentes mayoritarios que agruparon en el cluster III y al menos un representante de los genomas de referencia de los géneros descritos dentro de la familia <i>Pseudonocardiaceae</i> .....	157
<b>Figura 60.</b> Comparación de las secuencias del 16S rRNA, valores ANI y AAI para los componentes agrupados en el cluster IV.....	158
<b>Figura 61.</b> Valores de ANI y AAI de los componentes mayoritarios en las colonias que agruparon en el cluster IV.....	159

<b>Figura 62.</b> Árbol de máxima verosimilitud que muestra las relaciones filogenéticas de los genes del 16S rRNA de los componentes mayoritarios agrupados en el cluster IV con las secuencias disponibles de los géneros descritos en la clase <i>Nitriliruptoria</i> .....	161
<b>Figura 63.</b> Distancias estimadas entre los genomas de los componentes mayoritarios que agruparon en el cluster IV y los genomas de referencia de los géneros descritos dentro de la clase <i>Nitriliruptoria</i> .....	162
<b>Figura 64.</b> Representación de la estructura del gen 23S rRNA en el genoma del MAG082, componente mayoritario de las colonias W2 intervenido por 2 regiones intrónicas.....	167
<b>Figura 65.</b> Imagen de los resultados del ensayo de primers diseñados para cada componente mayoritario con las distintas muestras de colonias por la técnica de PCR.....	169
<b>Figura 66.</b> Lista de las categorías funcionales de los grupos de ortólogos anotados en los metagenomas de las colonias visibles .....	172
<b>Figura 67.</b> Análisis filogenético molecular por el método de máxima verosimilitud con el software de MEGAN de los aislados cultivables de las colonias visibles junto con las bacterias de referencias más cercanas de la base de datos del NCBI.....	177
<b>Figura 68.</b> Propuesta de componentes finales para las colonias beis.....	191
<b>Figura 69.</b> Propuesta de componentes finales para las colonias amarillas....	192
<b>Figura 70.</b> Propuesta de componentes finales para las colonias blancas subtipos W1 y W2.....	194
<b>Figura 71.</b> Propuesta de componentes finales para las colonias grises subtipos G1 y G2.....	195
<b>Figura 72.</b> Modelo propuesto de la composición general de una colonia visible del tipo observado habitualmente en cuevas.....	198
<b>Figura 73.</b> Hipótesis del potencial metabólico sugerido para las colonias.....	208
<b>Figura S1.</b> Representación de los filos más abundantes (con una abundancia >1%) en las muestras de suelo y soporte .....	245
<b>Figura S2.</b> Representación de los parámetros indicadores de $\alpha$ -diversidad en las muestras de suelos y soportes analizados por la técnica de amplicones.....	246

**Figura S3.** Correlación de los índices de  $\alpha$ -diversidad obtenidos por la técnica de amplicones (índice de Shannon) y shotgun (índice Nd) en las muestras de metagenomas.....247

# Índice de tablas

<b>Tabla 1.</b> Datos generales obtenidos tras el procesamiento de las secuencias a través de la plataforma de QIIME2.....	92
<b>Tabla 2.</b> Métrica general obtenida para cada muestra tras los pasos de ensamblado y bineado de las lecturas de los metagenomas.....	113
<b>Tabla 3.</b> Resultado de los análisis de $\alpha$ -diversidad en los metagenomas de colonias visibles.....	120
<b>Tabla 4.</b> Parámetros principales de calidad y tamaño de los componentes mayoritarios de las colonias.....	150
<b>Tabla 5.</b> Targets seleccionados para la detección primaria de los componentes mayoritarios que corresponden a cada tipo de colonia.....	164
<b>Tabla 6.</b> Tabla del número total de ORFs con/sin función en cada metagenoma de colonia .....	171
<b>Tabla 7.</b> Lista de marcadores de genes autotróficos y número de genes detectados en cada metagenoma.....	206
<b>Tabla S1.</b> Metadatos de las muestras analizadas por la técnica de amplicones.....	237
<b>Tabla S2.</b> Metadatos de las muestras de suelo y soporte analizadas por la técnica de amplicones.....	239
<b>Tabla S3.</b> Números de secuencias obtenidas para las muestras analizadas por la técnica de amplicones.....	241
<b>Tabla S4.</b> Taxonomía asignada a las OTUs que forman parte del core en las muestras de colonias visibles analizadas por la técnica de amplicones .....	243
<b>Tabla S5.</b> Resumen de datos de los MAGs obtenidos de los metagenomas de las colonias.....	248
<b>Tabla S6.</b> Clusters de MAGs en base a su similitud (AAI y ANI).....	257
<b>Tabla S7.</b> Genomas de referencia utilizados para la comparación a nivel de de los MAGs agrupados o no en Clusters y sus números de acceso en GenBank. ....	260
<b>Tabla S8.</b> Genomas de referencia utilizados para los análisis de estimación de distancias con los genomas de los componentes mayoritarios y su número de acceso en GenBank.....	263

<b>Tabla S9.</b> Lista de primers y su secuencia diseñados para identificar los componentes mayoritarios de las colonias.....	265
<b>Tabla S10.</b> Lista de los aislados cultivables de las colonias visibles.....	266
<b>Tabla S11.</b> Lista de las bacterias de referencias y su número de acceso a GenBank más cercanas a los aislados cultivables de las colonias visibles utilizados en la construcción del árbol filogenético de los aislados.....	278



# 1. Resumen



La cueva de Altamira, conocida como “la Capilla Sixtina” del arte rupestre destaca en su ámbito por la calidad de sus pinturas, y eso hace que sea de las cuevas más estudiadas y con mayor seguimiento de trabajos de investigación y mantenimiento, actualmente, la cueva se encuentra coordinada por una comisión de expertos y se encuentra en vigor el Plan de Conservación Preventiva (PCP) para la monitorización de los riesgos identificados en el “Programa de investigación para la conservación preventiva y régimen de acceso a la cueva de Altamira”. Dentro de estos riesgos destaca el riesgo biológico identificado con la presencia de colonias visibles coloreadas que colonizan las paredes y techos de la cueva.

Desde el punto de vista microbiológico las cuevas son ambientes subterráneos desarrolladas la gran mayoría en rocas solubles de carbonato, principalmente piedra caliza. Las condiciones físicas que encontramos en su interior las definen como entornos aislados con unas condiciones muy estables de temperatura, poca luz y un grado de intercambio de nutrientes con el exterior dependiente de su grado de apertura. Por esta constancia ambiental, constituyen hábitats naturales con presencia de microorganismos específicos bien adaptados y que han evolucionado hasta alcanzar un equilibrio estable pero delicado, siendo muy sensible a intervenciones que pueden producir efectos impredecibles y no deseados.

En el caso de Altamira, al encontrarse cerrada y sometida a seguimiento estrecho, las condiciones son de escasa disponibilidad de nutrientes lo que obliga a la formación de consorcios (biofilms) entre varias especies diferentes cuyo sentido es obtener un fitness más fuerte en comunidad para su supervivencia, beneficio que se comparte entre las distintas especies que lo forman.

Este tipo de biofilms, en forma de lo que hemos llamado colonias visibles microbianas coloreadas, son 4 en el caso de Altamira, denominadas colonias blancas, grises, amarillas y beis. Su descripción comienza cerca del año 1999 y las colonias beis se detectan más recientemente, en 2018. Las colonias

visibles son mucho más abundantes cerca de la entrada y su presencia se considera el principal riesgo para el biodeterioro de las pinturas.

Inicialmente, en los años 70 y 80 del siglo veinte se impulsaron los primeros estudios microbiológicos basados en cultivos en la cueva de Altamira, que se dirigieron a la caracterización de las colonias inmediatamente tras la detección de las mismas. Rápidamente se observó que las técnicas habituales de cultivo no eran satisfactorias para la descripción precisa de los componentes de las colonias. Por ello, estudios posteriores llevados a cabo a partir del año 2000, se hicieron con técnicas de secuenciación sin necesidad de cultivo. Estos estudios de colonias describieron los filos *Actinobacteria* y *Proteobacteria* claros protagonistas en estas muestras, y en menor incidencia *Acidobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Candidate division SPAM*, *Candidate phylum SBR1093*, *Gemmatimonadetes*, *Firmicutes*, *Candidate NC10*, *Nitrospirae*, *Planctomycetes* y *Verrucomicrobia*.

Las nuevas técnicas de secuenciación masiva empleadas en esta tesis (técnica de amplicones y shotgun) suponen una diferencia cualitativa bastante grande en comparación con las primeras técnicas utilizadas en Altamira, proporcionando una cantidad de secuencia muy superior al total de la disponible hasta ahora, y aportando un valor creciente a los resultados que se han obtenido en esta tesis. No sólo estos resultados podrían mejorar nuestro conocimiento sobre la composición de los componentes de las colonias de Altamira, sino que podrían servir como herramienta en ambientes similares en otras cuevas del mundo. Colonias visibles parecidas a las que se están desarrollando en la cueva de Altamira se observan en otras cuevas de España o del mundo con una composición muy similar. Asimismo, esta tesis es una aportación absoluta a la caracterización de la microbiota de Altamira gracias a la secuenciación total del DNA que nunca antes se había llevado a cabo, y una de las primeras en términos generales sobre microbiota de cuevas.

En este trabajo, se aporta evidencia de la existencia en las colonias de unos componentes predominantes (uno o dos por colonia), al menos 10 veces más abundantes que el resto, a los que hemos denominado componentes

mayoritarios (CMs) y que son característicos en cada colonia. Estos componentes más abundantes, que por la técnica de amplicones resultaron ser tres, clasificaron dentro del filo *Actinobacteria* con la familia *Euzebyaceae* (OTU1) y el género *Crossiella* (OTUs 2 y 4). Su distribución desigual en las colonias han sido la base para diferenciar en 2 subtipos a las colonias grises y blancas. Esta heterogeneidad detectada en algunas muestras de colonias aparentemente del mismo color no se ha mencionado anteriormente en otros estudios, indicando que la clasificación por color no es suficiente para describir la diversidad de las colonias.

Los análisis de biodiversidad de las colonias revelaron una complejidad media-baja, con la dificultad de distinguir entre componentes poco abundantes de las colonias de contaminantes provenientes del material de soporte que fue arrastrado en la toma de muestra. Por esta razón y por una tendencia a considerar las colonias como consorcios no excesivamente complejos, se seleccionó un core de 36 OTUs que resultarían los componentes más frecuentemente encontrados dentro de los distintos tipos de colonias de Altamira. En esta selección de core, destacaron las OTUs correspondientes a los CMs y un componente secundario (OTU3) compartido en casi todas las muestras con menor abundancia. La OTU3, asignada taxonómicamente a la clase *Gammaproteobacteria* y orden *Wb1-P19* con la base de datos de SILVA, filogenéticamente se agrupa con el orden *Chromatiales* y es detectado con regularidad en ambientes de cueva.

Con la técnica de shotgun se han obtenido secuencias de DNA total de los seis tipos de colonias identificados. Con esta técnica solo los genomas más abundantes van a recuperarse como MAGs y por esta razón las bacterias contaminantes del soporte cuya abundancia es minoritaria en las colonias no podrán ser recuperadas como MAGs. La clasificación de las secuencias de los seis metagenomas en “bins” nos ha permitido recomponer 90 borradores de genomas (MAGs) correspondientes hipotéticamente a otros tantos componentes de las colonias. La presencia de unos componentes mayoritarios muchos más abundantes que el resto se confirmó también por este enfoque. El número de componentes finales que se recuperó por colonia fue similar al core

descrito con la anterior técnica y que osciló entre 6 y 34 MAGs, siendo las colonias grises (ambos subtipos) las más complejas de todas. Los taxones que se encontraron en los MAGs recuperados de las colonias fueron *Acidobacteria*, *Actinobacteria*, *Candidate division NC10*, *Candidatus Dadabacteria*, *Candidatus Rokubacteria*, *Chloroflexi*, *Parcubacteria*, *Proteobacteria* (*alpha*, *beta*, *gamma* y *delta*), *Nitrospirae* y *Thaumarchaeota*.

Usando como criterio el grado de identidad de secuencia de nucleótidos (ANI) y de los polipéptidos predichos por Prodigal (AAI) entre los borradores de los genomas obtenidos (MAGs) para pertenecer al mismo género, alrededor de un tercio (36) agruparon en 13 clusters compartidos en las colonias.

Especialmente, los genomas de los CMs agruparon en dos clusters separados y asignados taxonómicamente al filo *Actinobacteria*, uno a la familia

*Pseudonocardiaceae* (cluster III) y el otro a la clase *Nitriliruptoria* (cluster IV).

Los CMs no tuvieron homólogos cercanos en las bases de datos de secuencias completas o taxonómicamente bien clasificadas en base a la comparación de los genomas completos por la secuencia de DNA, las secuencias codificantes y la secuencia completa del 16S rRNA.

Su asignación taxonómica no mostró una buena coincidencia con las referencias disponibles dando lugar a una novedad en la filogenia de las Actinobacterias. Por lo tanto, sugerimos que los cinco genomas (CMs de las colonias YC, G1, G2, W1 y BC) agrupados en el cluster III podrían constituir un nuevo género candidato dentro de la familia *Pseudonocardiaceae* con la probable presencia de 5 especies distintas. En el caso del cluster IV, en el que agruparon los componentes mayoritarios de las colonias W2 y BC, formaron un grupo separado de los otros órdenes descritos de la clase *Nitriliruptoria*, por lo que igualmente se justificaría la propuesta de un nuevo orden con dos especies distintas en el que situaríamos estos dos MAGs. Además, las secuencias presentes en las bases de datos más próximas a estos CMs provienen de estudios de suelos o de colonias de cuevas, lo que indica que estos nuevos taxones comprenderían solamente bacterias adaptadas y específicas de cuevas. Los borradores de genomas de alta calidad que proporciona esta tesis

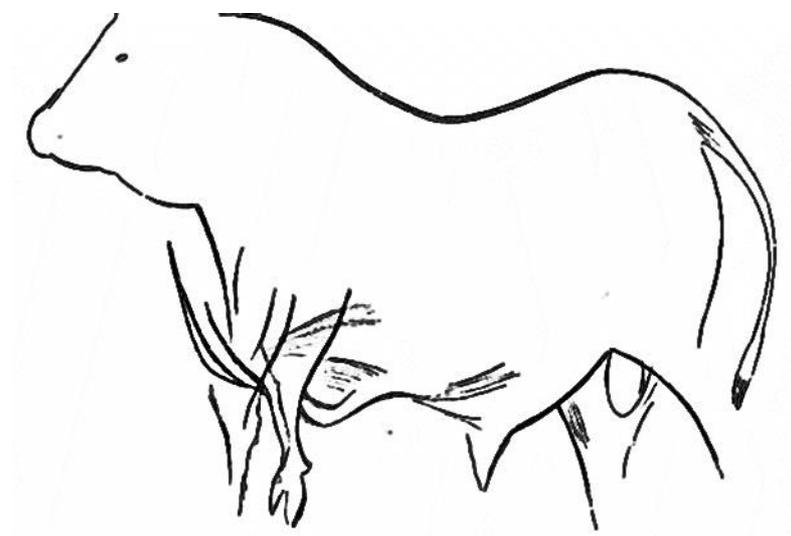
ofrecen la única información disponible sobre las capacidades nutricionales de estos “nuevos” taxones.

Además, los borradores obtenidos de los componentes de las colonias proporcionan un catálogo de genes disponibles para sobrevivir en el ambiente de Altamira. La comparación funcional del atlas de genes obtenido agrupado en clusters de grupos de ortólogos (COGs) muestra una gran uniformidad en todos los metagenomas. Entre las funcionalidades detectadas destaca la abundancia de genes involucrados en procesos del metabolismo y transporte, y concretamente, la presencia de genes capaces de soportar un crecimiento quimioautótrofo basado en la fijación de CO<sub>2</sub> y la oxidación de compuestos como hidrógeno, monóxido de carbono, amonio, etc. Este hallazgo sugiere que además de una forma de vida heterotrófica basada en compuestos orgánicos aportados por el agua de infiltración, las colonias podrían crecer de forma autotrófica en épocas de baja disponibilidad de materia orgánica.

Otra estrategia utilizada para un mayor conocimiento de los componentes de las colonias fue su recuperación en cultivos puros mediante el empleo de diferentes medios y condiciones de cultivo. Los experimentos de cultivo en esta tesis han proporcionado 117 cultivos puros a partir de los tipos de colonias analizadas. Sin embargo, ninguno de ellos correspondió claramente con un componente de las colonias descrito por las técnicas moleculares empleadas. Por tanto, este resultado concuerda con otros en poner de manifiesto la incultivabilidad de estos componentes. Un análisis exhaustivo de los borradores de genomas obtenidos podrá ayudar a plantear medios más apropiados para el cultivo de estas bacterias.

Finalmente, los trabajos microbiológicos de la cueva de Altamira no pueden ser meramente descriptivos, sino que deben estar necesariamente orientados al diseño de medidas de conservación preventiva para salvaguardar sus pinturas. La obtención de borradores de buena calidad de los componentes predominantes en las colonias ha permitido el desarrollo de primers específicos contra estos componentes, y con ellos, de ensayos de PCR que detectan de una manera directa rápida, sencilla y poco costosa la presencia de dichos

componentes. Estos ensayos de detección suponen una mejora sustancial en los métodos de seguimiento microbiológico del biodeterioro de cuevas y otros bienes culturales. Igualmente, la información genómica aportada en esta tesis puede ser la base para diseñar medidas de control eficaces para combatir eventuales problemas de biodeterioro causado por colonias visibles.



## 2. Introducción



## 2.1 La cueva de Altamira. Un bien cultural excepcional

La cueva de Altamira se encuentra situada en Santillana del Mar (Cantabria), a 2 km del casco urbano del pueblo. Dentro del recinto, donde se encuentra la cueva, encontramos también el Museo y Centro de Investigación de Altamira, todo ello bajo la gestión del Ministerio de Cultura y Deporte de España (Figura 1). La cueva fue descubierta originalmente en 1868 por Modesto Cobielles Pérez al ir de caza, pero al tratarse de un terreno kárstico con presencia de muchas cavidades este descubrimiento no trascendió a más. Fue poco después, en 1870, cuando Marcelino Sanz de Sautuola y su hija descubrieron las famosas pinturas en el interior de la cueva.

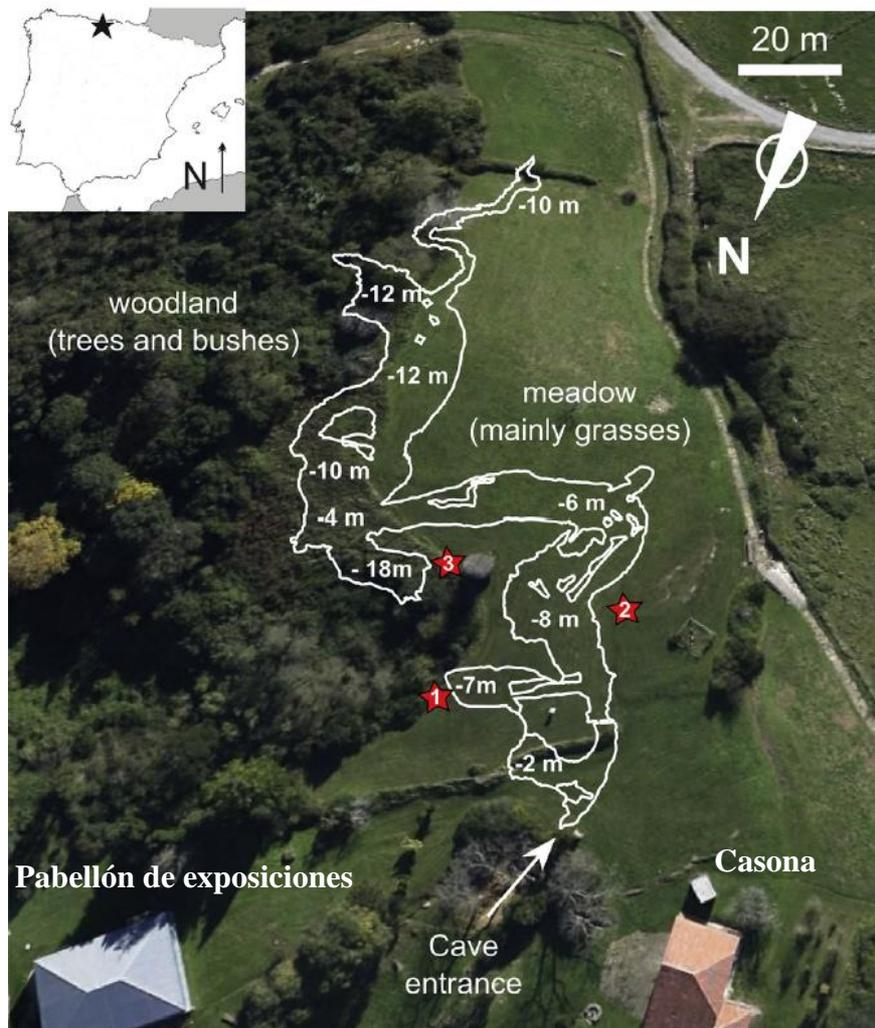


Figura 1. Plano aéreo de la situación de la cueva de Altamira. Figura editada de Gázquez et al., 2016.

La cueva de Altamira es un icono del arte rupestre universal por la presencia de pinturas representativas de la prehistoria y por su trayectoria desde que fueron descubiertas y puestas en valor. Por su calidad e importancia, es considerada la Capilla Sixtina del arte rupestre. Fue declarada Patrimonio Mundial por la Unesco en 1985 y forma parte del conjunto “Arte Rupestre Paleolítico de la Cornisa Cantábrica”. Bajo esta denominación se encuentran un conjunto de 18 cuevas que se sitúan en el Norte de España (Asturias, Cantabria y País Vasco) también con representaciones de arte rupestre (Figura 2).

Las pinturas que encontramos en la cueva de Altamira incluyen algunas de las más antiguas que podemos encontrar en Europa correspondientes al periodo Auriñacense (38.000 a 30.000 años atrás) y a los periodos Solutrense (entre 22.000 y 18.000 años atrás) y Magdalenense (18.000 a 8.000 años atrás). Estas pinturas se encuentran localizadas mayoritariamente en una zona de la cueva denominada la sala de Polícromos (Figura 3). A parte de las pinturas, podemos encontrar también otro tipo de arte como grabados y símbolos distribuidos por toda la cueva.

### Arte rupestre paleolítico de la cornisa cantábrica



Figura 2. Representación de arte rupestre paleolítico en la cornisa cantábrica. Situación de las 18 cuevas que agrupan bajo la denominación “Cueva de Altamira y Arte Rupestre Paleolítico de la Cornisa Cantábrica” declaradas Patrimonio de la Humanidad por la UNESCO a partir del 2008.

### 2.1.1 Descripción física de la cueva

La cueva de Altamira se localiza en la comunidad de Cantabria que presenta un 30% de roca karstificable (aproximadamente 1.530 km<sup>2</sup>) y es por ello, que en esta provincia se han detectado 6.500 cuevas de las cuales más de 40 tienen arte rupestre (Labrada et al., 2010). La cueva es una cavidad de formación natural de tipo kárstico y la composición de la roca es carbonatada formada mayoritariamente por calizas y calcarenitas. No es muy larga, tiene una longitud de 270 metros y posee una única entrada a 156 metros sobre el nivel del mar. Presenta una profundidad mínima de 2 y máxima de 26 m. Se encuentra dividida en las siguientes salas: Vestíbulo, Cocina, Cruce, Polícromos, Sala de los Muros, la Hoya, el Pozo y finalmente la Cola de Caballo (Figura 3). Sólo la sala del Vestíbulo se encuentra separada de las demás salas por la construcción de una puerta interior.

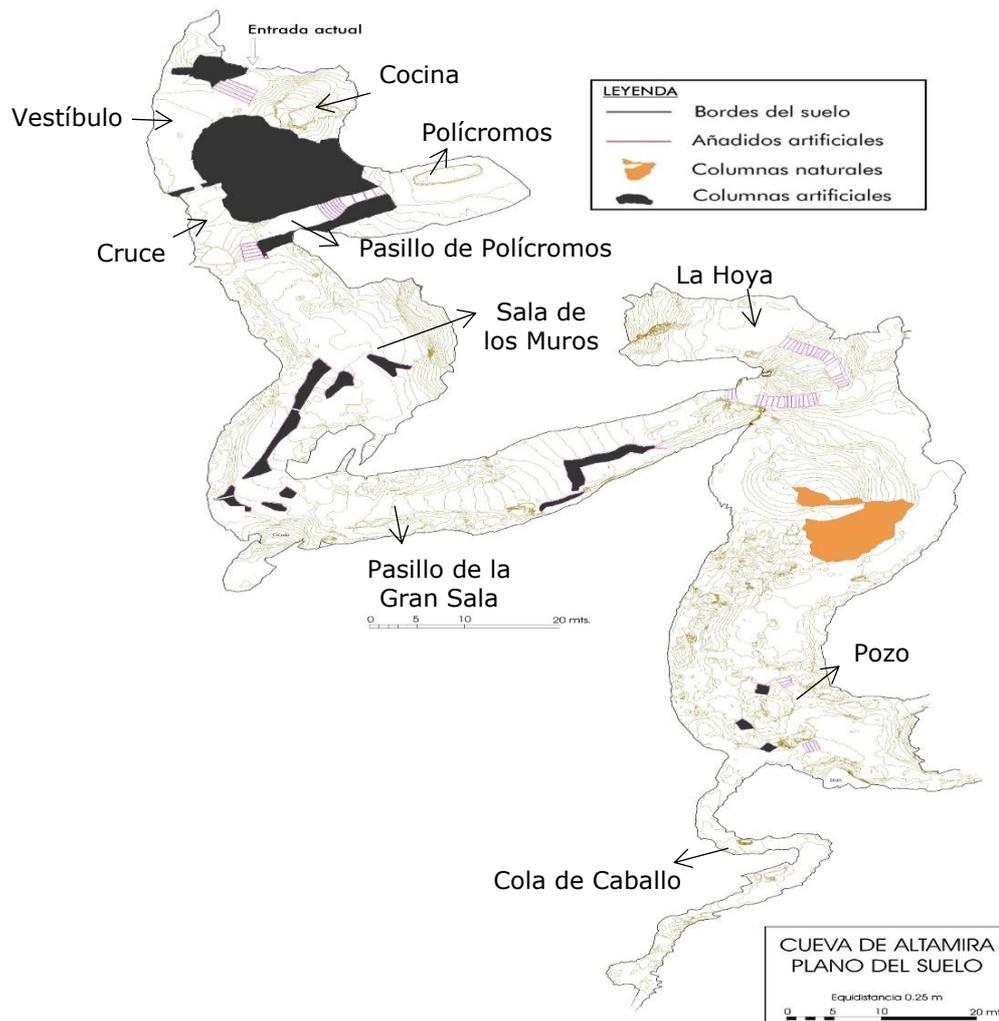


Figura 3. Plano actual de la cueva de Altamira. Se representa la división de salas en su interior junto con las modificaciones artificiales presentes.

Las condiciones ambientales son bastante constantes a lo largo del año. Mientras que la temperatura en el exterior tiene una gran variación en grados de verano a invierno, la temperatura en el interior de la cueva solamente varía entre 12,5-16,8°C (Figura 4). La zona de Vestíbulo, al ser la sala más cercana al exterior, es la más variable y aquí la temperatura oscila alrededor de 4°C durante el periodo anual, siendo menos variable en zonas más interiores. Además, la temperatura que se alcanza en la sala del Vestíbulo en los meses de verano puede significar 3 grados de diferencia respecto a zonas más interiores como el Pozo. En polícromos, por ejemplo, la temperatura anual oscila entre 13,1 (mínimo en verano) y 14,9 grados (máximo en invierno).

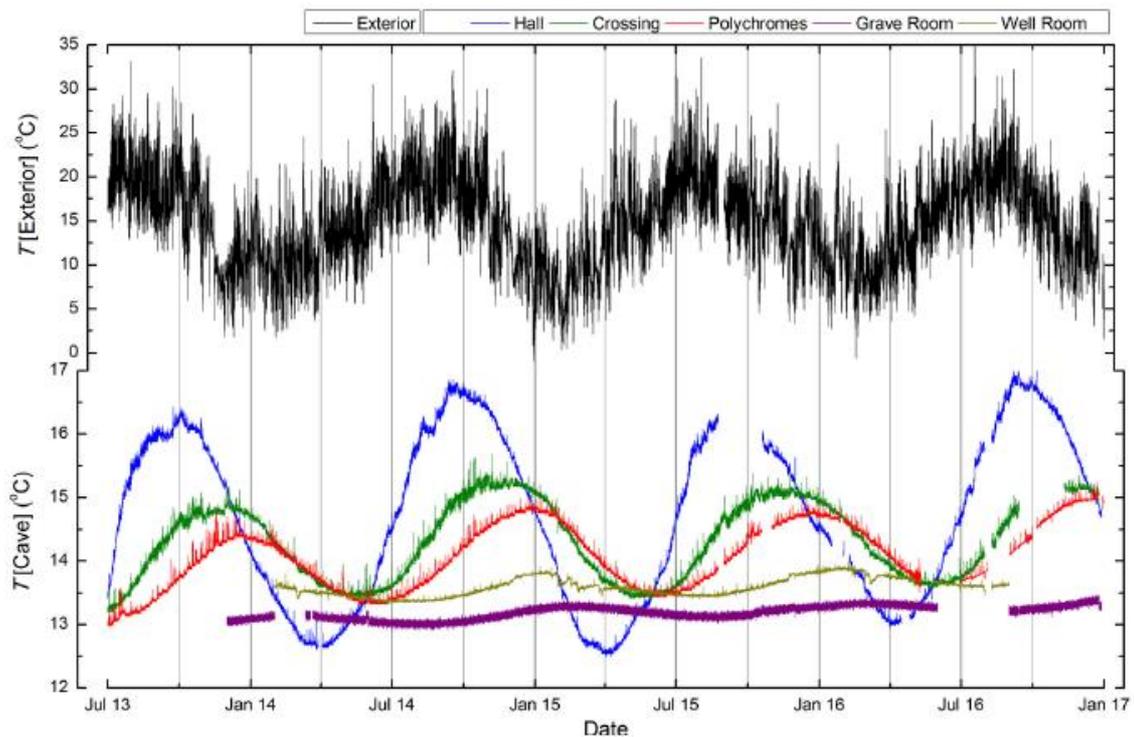


Figura 4. Variación de la temperatura para el exterior de la cueva y cada sala de la cueva durante el periodo de Julio 2013 a Enero de 2017. Figura obtenida de Sainz et al., 2018. Hall (Entrada); Crossing (Cruce); Polychromes (Polícromos); Grave Room (Sala de la Hoya); Well Room (Sala del Pozo).

La concentración de CO<sub>2</sub> en la cueva a menudo alcanza concentraciones máximas superiores a 4000 ppm, diez veces más altas que las concentraciones atmosféricas normales (Gázquez et al., 2016). La concentración de CO<sub>2</sub> sigue un ciclo anual con valores máximos en invierno y mínimos en verano, respondiendo a la dinámica de ventilación típica de la cueva (Figura 5). El gas radón que se produce naturalmente, se acumula en la cueva en condiciones de escasa ventilación y sus variaciones se usan como trazador de la ventilación y sus cambios en el interior (Sainz et al., 2018). El radón anual promedio durante el período 2003-2019 fue de alrededor de 3500 Bq m<sup>-3</sup> con valores máximos en invierno y mínimos en verano como se muestra en la figura 6 (Sainz et al., 2020). La humedad que presenta el suelo de la superficie debido a las precipitaciones hace un efecto de sellado evitando una difusión de los gases de la cueva al exterior. Este efecto se asocia al comportamiento anual del CO<sub>2</sub> y radón mencionados y se relaciona con los ciclos estacionales de desgasificación y recarga en la cueva (Figura 5).

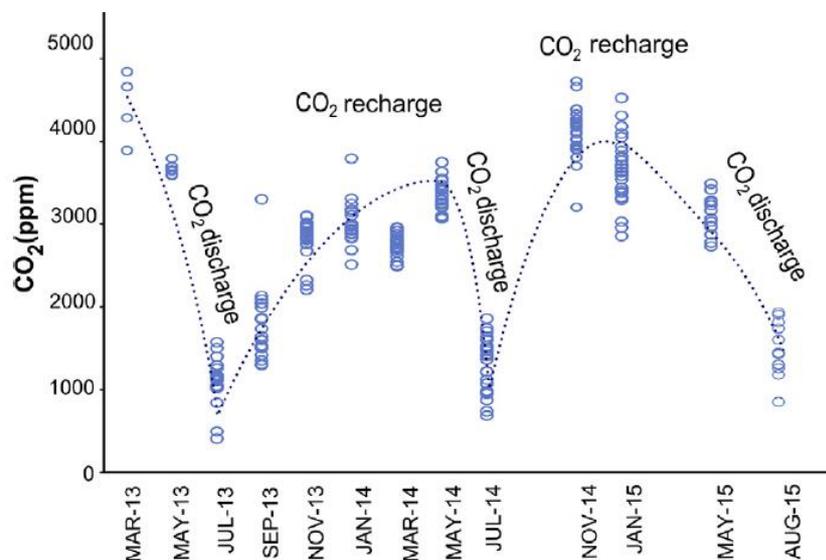


Figura 5. Ejemplo de la variación anual de CO<sub>2</sub> en la cueva de Altamira durante el periodo de Marzo 2013 a Agosto del 2015. Figura obtenida de Gázquez et al., 2016.

La humedad relativa en el interior de la cueva es siempre próxima al punto de saturación >75% (Figura 7). Esto favorece la aparición de condensaciones de agua sobre las paredes y equipos o estructuras artificiales. No se producen

variaciones estacionales adicionales dentro de la cueva, por lo que se considera un microclima muy estable.

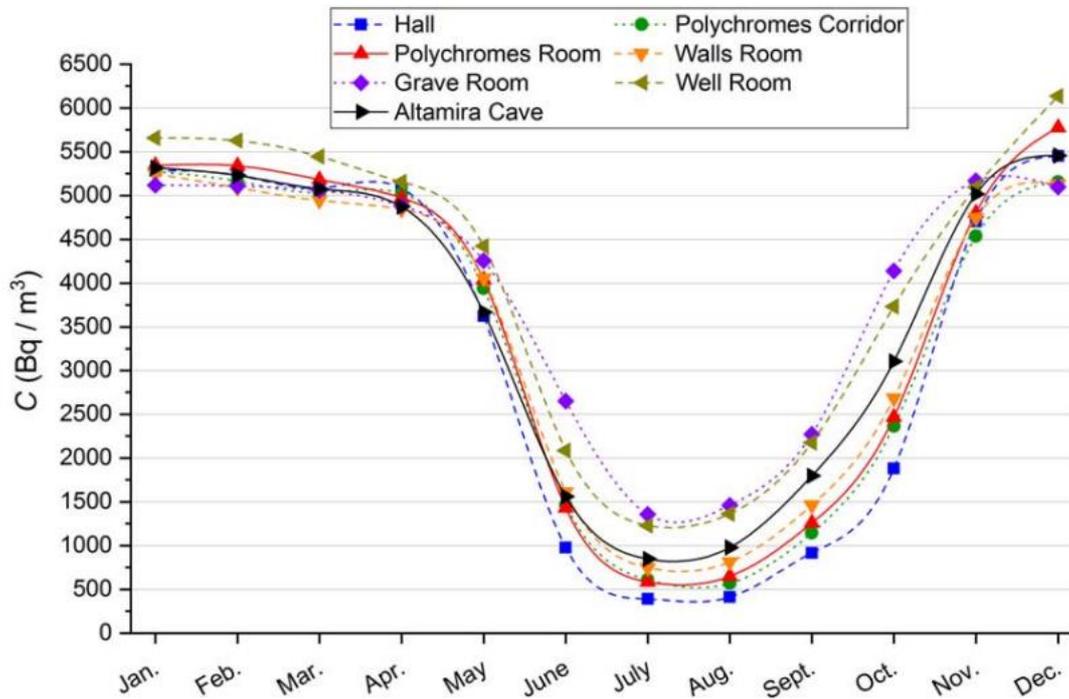


Figura 6. Ejemplo de la variación anual de Radón durante el periodo de 2013-2019 en cada una de las salas de la cueva de Altamira. En negro la variación promedio de todas las salas. Hall (Entrada); Polychromes Room (Sala de Polícromos); Grave Room (Sala de la Hoya); Polychromes Corridor (Pasillo de Polícromos); Walls Room (Sala de los Muros); Well Room (Sala del Pozo). Figura obtenida de Sainz et al., 2020.

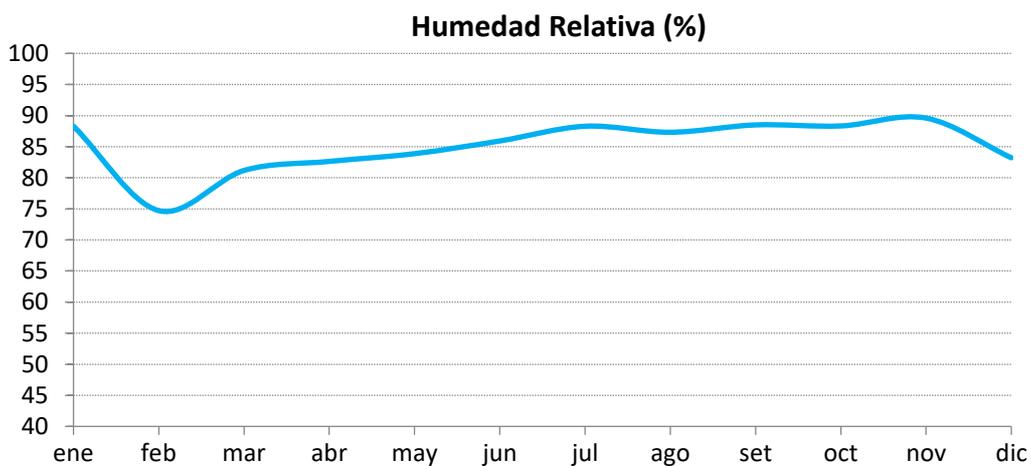


Figura 7. Ejemplo de la variación anual de la humedad relativa en el año 2019 en el interior de la cueva de Altamira.

La cueva se encuentra además situada en la zona lateral de una colina como se observa en la figura 1, y eso hace que la única infiltración natural de agua sea directamente de la lluvia o de escorrentía que acaba filtrando en la cobertera de la cueva (Figura 8). Esta agua suele transportar nutrientes como compuestos inorgánicos (nitratos y sulfatos), materia orgánica disuelta (mezcla de sustancias como la lignina y derivados, ácidos húmicos, aminas, urea, polisacáridos y proteínas), y aniones y cationes derivados de la lixiviación de los minerales que conforman la roca carbonatada entre otros. Estos compuestos acaban entrando en la cueva y alimentan a las bacterias que encontramos en ella.

Según el estudio de Saiz-Jimenez and Hermosin (1999) en aguas de infiltración recogidas en puntos de goteo de las salas del Vestíbulo y Polícromos en la cueva de Altamira, los compuestos más abundantes detectados en estas aguas fueron los ácidos grasos hexadecanoico (o ácido palmítico C16:0), octadecenoico (o ácido oleico C18:1<sup>9</sup>), octadecanoico (o ácido esteárico C18:0) y tetradecanoico (o ácido mirístico C14:0), así como derivados de la degradación de lignina por acción microbiana. Las membranas bacterianas son esencialmente lipídicas, y en la que podemos encontrar ácidos grasos como ácido mirístico, ácido palmítico y ácido esteárico. Por ello, se sugiere que los ácidos grasos encontrados en las aguas provengan de la cubierta de estas bacterias.

La presencia de cationes como  $\text{Ca}^{2+}$ ,  $\text{Na}^+$ ,  $\text{K}^+$  y  $\text{Mg}^{2+}$  y de aniones como  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$  y  $\text{HCO}_3^-$  (Laiz et al., 1999) también es detectable en las aguas de infiltración en la cueva de Altamira. Algunos de estos compuestos resultan de la descomposición mineral de las rocas carbonatadas presentes entre los que destaca la calcita ( $\text{Ca}^{2+}$ ,  $\text{HCO}_3^-$ ) y dolomita ( $\text{Mg}^{2+}$ ).



Figura 8. Simulación de la implicación directa de las aguas de lluvia y escorrentía en la cobertera de la cueva. Figura obtenida y editada del Informe CSIC 2005-2007, pag.45.

La presencia de nitratos en las aguas que entran en la cueva es también una importante fuente de nutrientes para las bacterias, ya que lo pueden reducir a amonio para la incorporación del nitrógeno del amonio a moléculas orgánicas en forma de grupos amino. A su vez también las bacterias son capaces de oxidar el amonio obteniendo energía y poder reductor. Se han medido concentraciones de nitratos en varios puntos de la cueva. La sala del Cruce (punto EN-2) corresponde con la sala donde se obtienen valores más altos como se representa en la figura 9 durante el periodo 2005-2009 (Informe CSIC 2005-2007, pag.79-94). En estas medidas se observa que, en el año 2003, las concentraciones de nitratos en el Cruce fueron muy altas, y se explicó a acumulaciones de material vegetal y crecimiento de la vegetación en la cobertera de la cueva. Por ello, actualmente se cuida limpiar los alrededores de la cueva y no permitir un crecimiento exagerado de vegetación. En la sala de los muros (no mostrado en la figura 9) es otro punto donde la medida de concentración de nitratos en el periodo de 2006 a 2009 fue también alta de 23,83 mg/l. No obstante, la concentración de nitratos en el agua de infiltración sigue una neta tendencia descendente probablemente como efecto del cese de actividades ganaderas en la cercanía de la cueva.

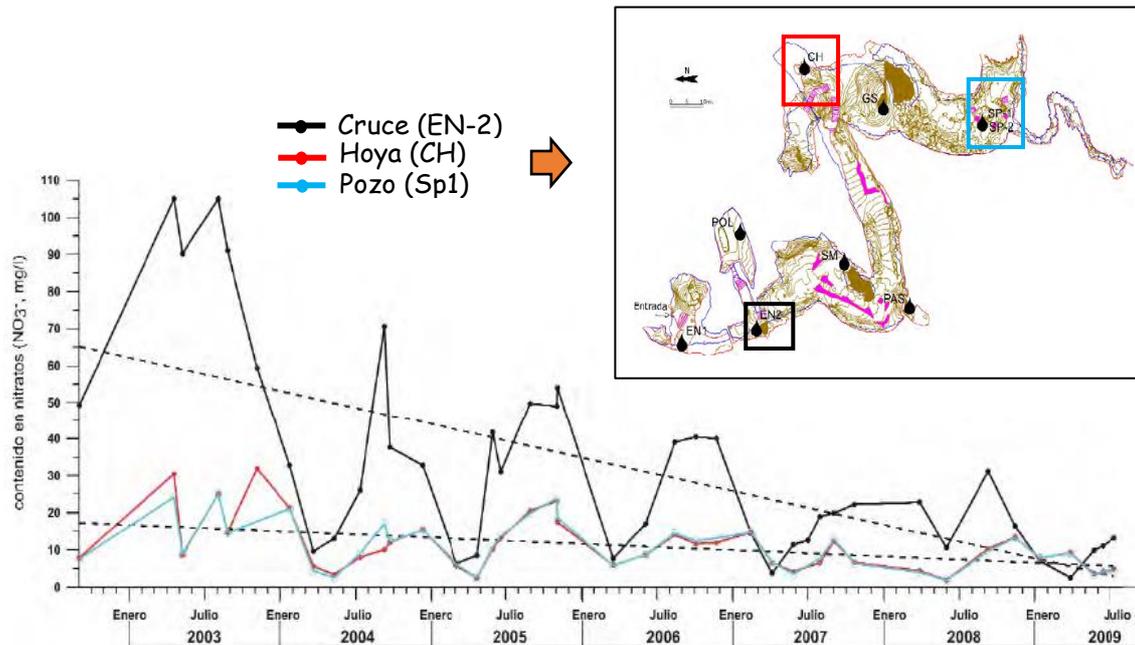


Figura 9. Acumulación de nitratos en puntos de goteo constantes en la cueva de Altamira para las salas de Cruce, Hoya y Pozo durante los periodos de año de Enero del 2003 a Julio del 2009. Figura obtenida y editada del Informe CSIC 2005-2007, pag.93.

Otras especies nitrogenadas que pueden servir de alimento a las bacterias como por ejemplo nitritos, amonio, óxido de nitrógeno u óxido nitroso siguen una tendencia similar a los nitratos. El nitrógeno Kjeldahl es un indicador cuantitativo de la cantidad de nitrógeno orgánico en sus diversas formas (proteínas y ácidos nucleicos en diversos estados de degradación, urea, aminas) y el ion amonio  $\text{NH}_4^+$ , presentes en una muestra de agua. Para su determinación se rompen todos los enlaces de nitrógeno de los compuestos nitrogenados presentes en la muestra y se convierte todo el nitrógeno unido orgánicamente en iones amonio ( $\text{NH}_4^+$ ). Después los iones amonio se convierten en amoniaco ( $\text{NH}_3$ ) en un proceso de destilación y finalmente se determina la concentración total de nitrógeno. Este valor medido en diferentes puntos en el suelo exterior de la cueva presentó concentraciones altas (Informe CSIC 2005-2007, pag.51-62).

Estos compuestos nitrogenados, aparte de provenir de la propia composición del suelo, se sospecha que puedan ser consecuencia de las actividades ganaderas y agrarias que se llevaron a cabo en los alrededores de la cueva

antes de su adquisición para el conjunto Cueva-Museo. Es posible que como consecuencia de aquella actividad existan depósitos de residuos orgánicos que están siendo arrastrados por el agua y cayendo con los años hacia la cueva alcanzando la sala del Vestíbulo, el Cruce y Polícromos, aunque se cree que su concentración también muestra una tendencia decreciente como ocurre con los nitratos.

### **2.1.2 Gestión de la cueva de Altamira y alteraciones sufridas**

La cueva de Altamira debido a la calidad de sus pinturas y a que fue la primera cueva española en dónde se descubrió arte paleolítico, a diferencia de otras con arte similar, es probablemente la cueva que más sometida ha estado a la intervención por el hombre. Numerosas excavaciones arqueológicas y el boom de visitas que sufrió durante sus primeros años abierta al público son algunos de los acontecimientos que ha sufrido en su historia reciente, y han provocado que la cueva actualmente presente modificaciones respecto al plano original (Figura 10). Los más relevantes se detallan a continuación:

- 1868-1880: Desprendimientos de rocas del techo, instalación de una puerta de madera, así como primeras visitas.
- 1880-1925: Modificaciones en la entrada que conllevan la sustitución de la puerta de madera por una verja de hierro y se agranda la entrada. Construcción de un muro en la entrada a la sala de Polícromos, primeros signos de deterioro de pinturas por filtraciones de agua, uso de antorchas y velas y marcas de dedo sobre los pigmentos, excavaciones y nuevos desprendimientos.
- 1925-1940: Apertura de caminos en el interior de la cueva, construcción de un nuevo muro y rebaje del suelo de Polícromos, sellado de grietas y fisuras en el interior y exterior de la cueva, construcciones en los alrededores del exterior de la cueva de la carretera a Santillana y al museo, vivienda del guarda y aseos, y excavaciones arqueológicas.

- 1940-1944: Colocación de apeos de madera en la sala de Polícromos.
- 1944-1977: Retirada de los apeos en la sala de Polícromos, construcción de muros y pilares de hierro y hormigón, instalación de una nueva puerta entre Vestíbulo y Cruce, acondicionamiento interno para las visitas, instalaciones luminosas, visitas masivas, construcciones en los alrededores de un aparcamiento, edificios de acogida a los visitantes, viviendas y explotaciones agrarias y ganaderas en el entorno de la cueva.
- 1977-2014: Creación del museo, aperturas y cierres temporales de las visitas a la cueva, construcción de un nuevo edificio del museo, retirada de las luminarias y otras estructuras artificiales del interior de la cueva, creación de nuevas puertas e investigaciones arqueológicas.

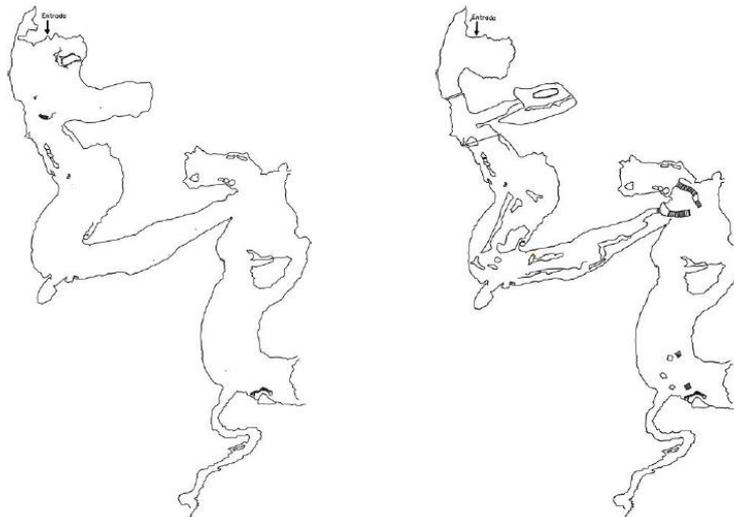


Figura 10. Evolución del plano original de la cueva actual tras las modificaciones sufridas hasta el día de hoy. A la izquierda plano original de la cueva y a la derecha plano actual tras las modificaciones sufridas. Figura obtenida del Informe del Programa de Investigación para la Conservación Preventiva y Régimen de acceso a la cueva de Altamira (2012-2014) Volumen I.

No cabe duda de que, desde su descubrimiento, la cueva ha tenido en su interior una actividad permanente, ya sea de investigación arqueológica, de construcción o mantenimiento o de visitas públicas.

Entre 1870 y 1925, la investigación arqueológica en la cueva fue muy notable. Durante los años 1880 a 1925 la gestión de la cueva estaba a cargo del Ayuntamiento de Santillana del Mar y de 1925 a 1940 la gestión pasó a la Comisión de Administración y Exploración de la cueva de Altamira. En esta etapa, en concreto en los años 1924-1935 se llevó a cabo un proyecto de reparación y consolidación de la cueva coordinado por el ingeniero Alberto Corral. A partir del año 1875 fue cuando empezaron las primeras visitas de los vecinos del pueblo y primeros investigadores, no hay datos de visitantes hasta 1952 en el que el se empezaron a registrar el número de visitantes. En este año el número de visitas alcanzó la cifra de 30.000 visitantes al año hasta llegar a 174.000 en el año 1972. Estas visitas siguieron altas en 1974 con 162.072 y en 1975 con 151.326 personas al año.

De 1940 a 1977 la gestión de la cueva pasó a cargo del Patronato Provincial y de 1976 a 1978 la comisión investigadora que coordinó la conservación de la cueva de Altamira fue dirigida por Eduardo Ripoll (Martín and Corruchaga, 2009). En 1976 se acordó rebajar el número de visitas a 500 personas al día, alcanzando la cifra de 86.634 visitantes al año. En 1977 la titularidad de la gestión de la cueva pasó a ser una gestión estatal y sigue siéndolo hasta el día de hoy. De 1979 a 1984 se encomendó a la Universidad de Cantabria, en concreto al Profesor Eugenio Villar, la coordinación de los primeros estudios para evaluar el estado de conservación de la cueva. Estos análisis se debieron a la observación de los primeros indicios de decoloración en las pinturas.

El primer cierre de la cueva fue durante el periodo de 1978-1981 y, a partir de 1982 se retomaron de nuevo las visitas con grupos de 5 personas y un guía. El estudio de Villar terminó en 1983, y en 1993 el Ministerio de Cultura traslada el encargo de estos análisis de conservación de la cueva al Consejo Superior de Investigaciones Científicas (CSIC). De 1997 a 2001 se diseñó una estrategia de conservación para la cueva y su entorno dirigido por Juan Antonio Lasheras (director en funciones en ese momento del Museo de Altamira), en los que se eliminaron estructuras construidas en el interior de la cueva como infraestructuras eléctricas o canalizaciones, y alrededor de la cueva como

instalaciones ganaderas o viviendas/edificios. La cifra anual de visitas en este periodo fue de 11.320 personas al año.

No fue hasta 2002 cuando el personal del museo de Altamira recomendó un segundo cierre temporal de la cueva cuando observaron el crecimiento de microorganismos fototróficos en el techo causado por el empleo de equipos de iluminación. En esta misma época también se detectaron las primeras apariciones de colonias visibles. Aunque permaneciera cerrada al público, la cueva seguía siendo accesible para los investigadores. Durante el periodo de 2003 a 2012, se realizaron diversos proyectos para la conservación de la cueva dirigidos por Sergio Sánchez-Moral (CSIC).

Más adelante, de 2012 a 2014 se emprendió el programa de investigación para la conservación preventiva y régimen de acceso a la cueva de Altamira por el Ministerio de Educación, Cultura y Deporte (MECD), y encargado al profesor Gaël de Guichen como director científico. Este programa fue coordinado por el Museo y Centro de Investigación de Altamira y el Instituto del Patrimonio Cultural de España (IPCE) y se encontraba dividido en 6 proyectos: (1) control del biodeterioro, (2) seguimiento ambiental, (3) conservación del soporte y la policromía, (4) accesibilidad, (5) valor social y (6) comunicación y memoria. Estos proyectos tienen en común que el objetivo fundamental del programa trataba de identificar los riesgos de deterioro específicos de la cueva.

El riesgo biológico y dentro del mismo la proliferación de microorganismos, recogido en el proyecto de control del biodeterioro, se consideró como uno de los riesgos más importantes a analizar debido a los indicios previos de colonización que había sufrido ya la cueva. La presencia de colonias visibles se encuentra integrada en este riesgo y actualmente constituye el mayor peligro para la cueva debido a su rápida dispersión en estos años.

La cueva abrió sus puertas de nuevo al público en 2014 de forma experimental a un grupo semanal de 5 personas, manteniéndose desde entonces este plan de visitas para las visitas actuales. Esta decisión representa un compromiso entre la conservación preventiva de la cueva y el disfrute del bien cultural. Las

evaluaciones de riesgo realizadas hasta la fecha no han mostrado incrementos significativos de parámetros a las visitas, por lo que el régimen de visitas se mantiene activo en las mismas condiciones. La cueva de Altamira sigue siendo gestionada por el Museo Nacional y Centro de Investigación de Altamira con titularidad del Ministerio de Cultura y Deporte. Además, la cueva se encuentra coordinada por una comisión de seguimiento formada por expertos multidisciplinares y se encuentra en vigor el Plan de Conservación Preventiva (PCP) para la monitorización de los riesgos identificados en el programa de investigación para la conservación preventiva y régimen de acceso a la cueva de Altamira.

## 2.2 Microbiología de cuevas

Una cueva es una cavidad natural originada en rocas de tipo sedimentario por procesos de disgregación por factores como el agua, viento, o hielo (meteorización) y en rocas magmáticas por la solidificación de coladas volcánicas. El proceso más común de formación de cuevas es la meteorización química, en la cual el material de la roca se descompone alterando su composición. Un ejemplo bien conocido de meteorización química es el proceso de karstificación, por el cual se originan cuevas como consecuencia de la acción del agua circulante sobre evaporitas carbonatadas como calizas y dolomías. En torno a un 90% de cuevas en el mundo se localizan sobre este tipo de rocas. Las evaporitas carbonatadas representan un 15,2% de la superficie terrestre como se muestra en la figura 11 (Goldscheider et al., 2020).

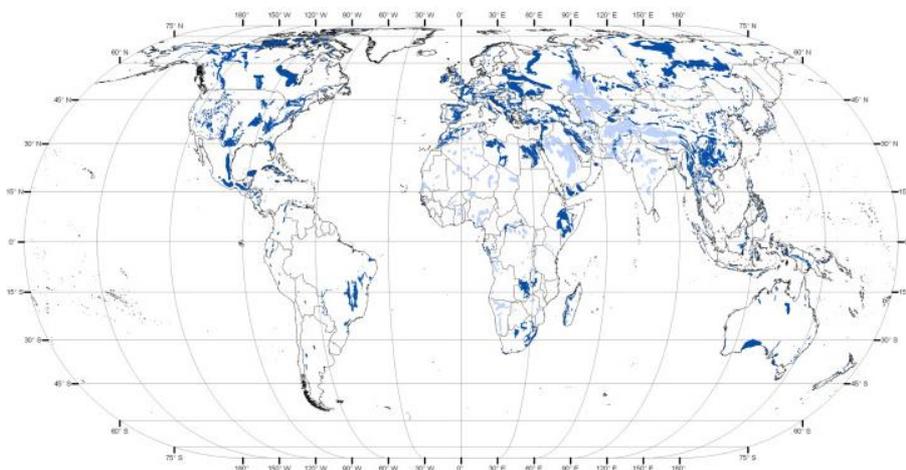
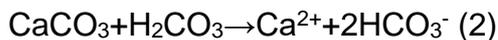
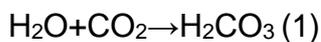


Figura 11. Distribución de las regiones kársticas en la superficie terrestre. En color azul oscuro se muestran las regiones que presentan roca carbonatada. Figura obtenida de Jones and White, 2019.

La formación de karst es un proceso de disolución muy lento, aproximadamente 125 mm de avance en la roca cada 1.000 años (Martínez, 1984). Este proceso se genera cuando el agua, que entra en la cueva, se acidifica primero cuando pasa a través del suelo al adquirir dióxido de carbono, generado por procesos como la respiración de las plantas y su descomposición (1). Esta agua acidificada interactúa con la roca carbonatada disolviéndola (2) en un proceso lento y que es el que da lugar a la formación de las cuevas.



Cuando esta agua, cargada de gas y piedra caliza disuelta en forma de bicarbonato cálcico, llega a una cavidad más grande que las fisuras por las que ha pasado, puede evaporarse lentamente y las sales disueltas en el agua llegan a cristalizarse, proceso conocido como precipitación de carbonato cálcico ( $\text{CaCO}_3$ ) que en estas cuevas da lugar a la formación de espeleotemas (3). Por ejemplo, las sales cristalizan cuando esta agua gotea desde el techo de una cueva hacia el suelo, formando estalactitas en el techo y estalagmitas en el suelo o columnas cuando estas dos formas llegan a unirse (Figura 12).

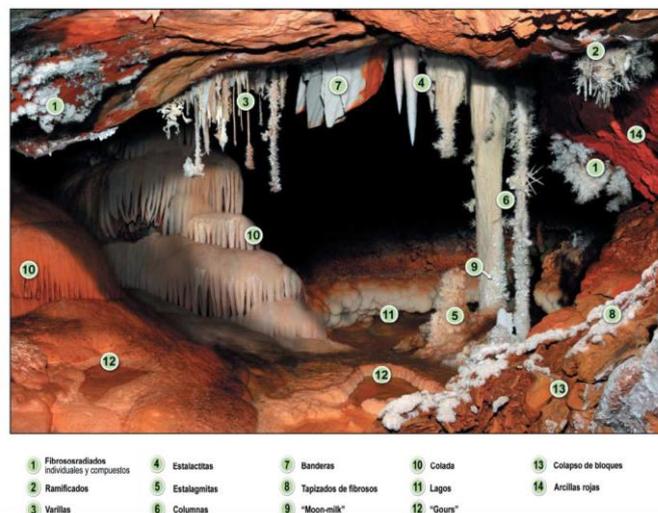


Figura 12. Ejemplo de los diferentes tipos de espeleotemas que resultan de la precipitación del carbonato cálcico en el interior de una cueva. Figura obtenida de Monumento Natural Cueva del Castañar - "Un laboratorio natural", 2015

España contiene 4 dominios litológicos: silíceo, calizo o calcáreo, arcilloso y volcánico, entre los que destaca el dominio calcáreo al que frecuentemente se menciona como "la España caliza" debido a la extensión que presenta (Codrón, 1983). En torno a un 30% de la superficie peninsular (110.000 km<sup>2</sup>) es karstificable. Esto explica que haya un gran número de cuevas, de las cuáles más de 10.000 han sido exploradas y estudiadas. Además, muchas de estas presentan arte rupestre y muchas de ellas han sido habilitadas al público para su visita, constituyendo un recurso socioeconómico muy notable para España.

Por lo general, las cuevas que no están abiertas al exterior, como es el caso de la cueva de Altamira, son ambientes muy interesantes ya que suelen tratarse de entornos aislados con unas condiciones ambientales constantes todo el año, humedad, temperatura homogénea, con poco aporte externo de nutrientes, y dónde apenas hay luz, especialmente a medida que nos adentramos hacia zonas más interiores. Por tanto, los seres vivos que habitan especialmente en estos ambientes se ven obligados a sobrevivir y adaptarse bajo estas condiciones. Por ejemplo, los animales de cuevas (troglóbios) suelen ser pequeños, ciegos y faltos de pigmentación etc... como consecuencia a esta adaptación. De la misma manera, los microbios que habitan en las cuevas van a ser específicos siendo el resultado de la selección impuesta por vivir en estas condiciones.

La caracterización de estas comunidades microbianas de cuevas ha sido por tanto objeto de estudio desde los inicios de la microbiología. En los primeros estudios de cuevas, los microorganismos no se detectaban y su presencia se deducía de la incapacidad de explicar muchas reacciones geoquímicas a través de procesos puramente químicos, necesitándose para entenderlos una implicación directa de la actividad microbiana (Hess, 1900; Hill, 1981). Desde entonces distintas técnicas tradicionales de cultivo se han utilizado para la detección de estos microorganismos. Por ejemplo, en el caso de la cueva de

Altamira los primeros estudios microbiológicos, que se llevaron a cabo entre 1978-1999, consistieron en el plaqueo en medios de cultivo ricos para el crecimiento de bacterias y hongos (Somavilla et al., 1978). En la misma línea se realizaron estudios posteriores entre 1981-1983 por F. Uruburu y C. Hardisson de las universidades del País Vasco y de Oviedo, respectivamente. Estos análisis microbiológicos que se hicieron en la cueva de Altmira fueron impulsados por los primeros indicios de biodeterioro observados en las pinturas.

Con el paso del tiempo la percepción de que el 99% de la diversidad bacteriana presente en muestras ambientales permanecía incultivable, impulsó la necesidad de utilizar técnicas moleculares independientes de cultivo para el análisis de poblaciones bacterianas naturales. Estas técnicas fueron posibles por el desarrollo de las tecnologías de clonado y secuenciación de DNA y después la PCR, cuyo uso empezó a generalizarse a partir de 1980. Su uso supuso una revolución en el campo de la taxonomía bacteriana (Hugenholtz et al., 1998; Sharma et al., 2005).

Los primeros usos de tecnología molecular para identificación de bacterias consistían principalmente en amplificar por PCR marcadores moleculares conservados entre especies como los genes ribosomales (ej. 16S, 23S y 5S rRNA en procariontes y 18S, 5,8S y 28S en eucariotas), genes “housekeeping” (ej. *recA*, *gyrA*) o genes implicados en pasos concretos del metabolismo (ej. *amoA*, *nirA*) para la detección de grupos específicos de bacterias. Tras la PCR, se purificaban o clonaban los productos de amplificación para ser secuenciados y las secuencias obtenidas se comparaban con las disponibles en las bases de datos. Con el uso de la PCR y el diseño de primers universales fue posible la caracterización de poblaciones bacterianas sin necesidad de cultivo. Para ello se realiza la amplificación simultánea de una misma diana (16S rDNA) de todos los individuos presentes en una mezcla. La colección de fragmentos resultantes se usaba para generar librerías de clones para secuenciar o se empleaban en combinación con otros métodos de huella digital. En concreto el método más utilizado en ecología microbiana fue la amplificación de 16S rDNA seguida por separación de productos en

electroforesis en gel con gradiente desnaturante (PCR-DGGE), donde el producto de interés amplificado se corre en un gel de poliacrilamida en condiciones desnaturantes con un gradiente de urea o formamida (Muyzer et al., 1993). Las moléculas de DNA migran en el gel hasta alcanzar su punto de desnaturalización, que va depender de su secuencia, obteniéndose una banda específica por cada una de las especies bacterianas presentes en la muestra estudiada.

Más adelante, la invención de técnicas de secuenciación masiva supuso otra revolución en el campo de la microbiología. Estas técnicas iniciales fueron la pirosecuenciación desarrollada por la compañía 454 y luego por Roche, producía una longitud de lecturas entre 700-800 pb (Hernández et al., 2020). Aunque con diferente fundamento, era similar la tecnología de Ion-Torrent que actualmente es explotada por ThermoFisher. La tecnología de Illumina proporciona lecturas de 150 a 300 nt, y es probablemente la más utilizada. El desarrollo de esta tecnología acompañado de un descenso en los precios de secuenciación y un aumento en la productividad, y unidos al desarrollo de herramientas bioinformáticas (Mothur, Qiime) y de bases de datos cada vez más comprehensivas (SILVA, GreenGenes, RDP), han permitido una mejor caracterización y estimación de la diversidad presente en las muestras ambientales. A día de hoy las tecnologías de secuenciación siguen avanzando y cada vez son más potentes y económicas con una tecnología aplicada a moléculas únicas capaces de dar productos de secuenciación más grandes que las anteriores plataformas (>20 kb). Estas plataformas se suelen denominar de tercera generación, y entre ellas encontramos Pacific Biosciences (Pootakham et al., 2017; Yang et al., 2020) y Oxford Nanopore (Kerkhof et al., 2017; Liem et al., 2020).

### **2.2.1 Estrategias de supervivencia de los microorganismos en las cuevas**

Los microbios que se han adaptado a vivir bajo condiciones energéticas desfavorables y limitantes a nivel nutricional han desarrollado diversas estrategias para poder subsistir, que casi siempre consisten en detener o ralentizar su crecimiento.

La producción de esporas es uno de los principales ejemplos de suspensión de crecimiento en condiciones desfavorables en espera de condiciones con una mayor disponibilidad de nutrientes que les permita comenzar de nuevo a desarrollarse. De hecho, algunas de estas esporas pueden llevar a cabo una germinación espontánea, evaluando la capacidad que las condiciones del medio presentan para poder crecer (Sturm and Dworkin, 2015).

Sin embargo, no todos los microbios esporulan y aun así, consiguen adaptarse y sobrevivir alcanzando lo que se conoce como “estado de crecimiento oligotrófico” (Gray et al., 2019). Este estado consiste en adoptar un proceso de inanición de largo plazo ralentizando su crecimiento y haciéndose más tolerantes frente a diferentes tipos de estrés, manteniendo algunas actividades celulares básicas, como la fuerza protón-motriz para seguir siendo viables. Las células en este estado pueden reanudar rápidamente el crecimiento normal cuando las condiciones de nutrientes mejoran. Células que entran en este estado y persisten en el tiempo pueden volverse no cultivables, estado conocido como “viables pero no cultivables” (VBNC). En este segundo estado las células no consiguen llegar a un número de divisiones suficientes para dar lugar a un crecimiento visible en medio de crecimiento no selectivo pudiendo permanecer largos períodos de tiempo en este estado, de ahí que no sea una respuesta transitoria.

Las bacterias VBNC, a diferencia de las células muertas, presentan una membrana intacta y un constante nivel de transcripción génica en procesos como la respiración, la morfología celular, la virulencia, la incorporación de aminoácidos, la resistencia a diferentes tipos de estrés o la adhesión. Una característica común de las VBNC es su capacidad para reducir su tamaño y redondear su forma, con el objetivo de minimizar energía como se ha descrito en bacterias VBNC de *Campylobacter spp*, *Burkholderia pseudomallei*, *Vibrio Cholerae* o *Escherichia coli O157:H7* (Li et al., 2014; Wei and Zhao, 2018). Tienen una menor tasa metabólica a la que podrían estar expresando en una fase exponencial de crecimiento.

Una situación que podemos encontrar en cuevas aisladas es una competición de crecimiento haciendo que la disponibilidad de nutrientes sea aún más limitante y, las condiciones sean insuficientes para permitir el crecimiento de ningún microorganismo de forma individual. Al ser esta condición persistente, obliga a la aparición de asociaciones entre especies como única solución para mantener un crecimiento activo (Barton and Jurado, 2007). De esta manera siempre va a haber especies que sean capaces de obtener energía de las fuentes disponibles en el medio y a su vez producir compuestos secundarios que van a ser fuente de energía para soportar el crecimiento del resto. Estas relaciones recurso-recurso es lo que se conoce como mutualismo y dan lugar a numerosas soluciones en microbiología como la formación de biofilms y otros consorcios estables.

Un ejemplo de mutualismo son los biofilms. Suelen ser consorcios entre diferentes organismos (hongos, bacterias) y fotosintetizadores (cianobacterias, algas) que crecen embebidos en una matriz orgánica polimérica adherida sobre una superficie. Dicha matriz confiere protección y resistencia frente a fenómenos adversos, aumenta la disponibilidad y mejora del aprovechamiento de nutrientes, y un mayor transporte e intercambio de enzimas. La limitación de nutrientes en el exterior, así como las perturbaciones presentes obligan a las células a responder rápidamente interaccionando entre ellas y favoreciendo un nicho de transferencia horizontal de material genético de las que obtener un fitness más fuerte en comunidad. La secreción constante de sustancias extracelulares poliméricas ricas en carbohidratos (ej. polisacáridos) o adhesinas que conforman esta matriz es característico para sustentar estas agrupaciones de comunidades (Branda et al., 2005; Jennings et al., 2015; Paharik and Horswill, 2016). Además, estas sustancias son capaces de almacenar alrededor de 20 veces su peso en agua, incrementando de esta manera la capacidad de retención de agua en el biofilm (Rao et al., 2020).

Este tipo de consorcios se pueden observar en las cuevas como colonias coloreadas visibles con un tamaño inferior a 5 mm que se desarrollan sobre las paredes y techos. Suelen denominarse simplemente en función del color que presentan, así encontramos colonias amarillas, grises, blancas o beis (Figura

13). Estas colonias son una mezcla de diferentes componentes procarióticos (bacterias y arqueas) con una estructura muy organizada, y cuyo metabolismo se encuentra relacionado al ciclo del nitrógeno y carbono de la cueva. Las condiciones ambientales y nutricionales sobre las que se desarrollan hacen difícil la recuperación de los componentes individuales en forma de cultivos puros en el laboratorio, comportándose individualmente en la mayoría de los casos como bacterias VBNC como se ha descrito anteriormente.



Figura 13. Ejemplo de colonias visibles (amarillas, grises y blancas) en las paredes de la cueva de Altamira.

### **2.2.2 Casos de colonias visibles en cuevas**

La presencia de colonizaciones visibles es una característica natural de los ecosistemas de cuevas. En la cueva de Altamira, objeto de esta tesis, observamos también este tipo de colonizaciones. Encontramos varios tipos de colonias visibles en función del color que presentan: amarillo, gris, blanco y beis. Las primeras colonias que se detectaron a partir de 1999 se encontraban en la entrada y eran de color amarillo, gris y blanco. Con el tiempo, su progresión hacia zonas interiores ha sido rápida, llegando incluso a la sala de las pinturas (sala de Polícromos). Actualmente se sitúan principalmente en el Vestíbulo, Cruce, Pasillo de Polícromos y en menor medida en la sala de Polícromos (Figura 14). En 2018 se detectó un nuevo tipo de colonias visibles

de color beis cerca de la puerta que conecta la sala del Vestíbulo con el Cruce y también en el pasillo de Polícromos.

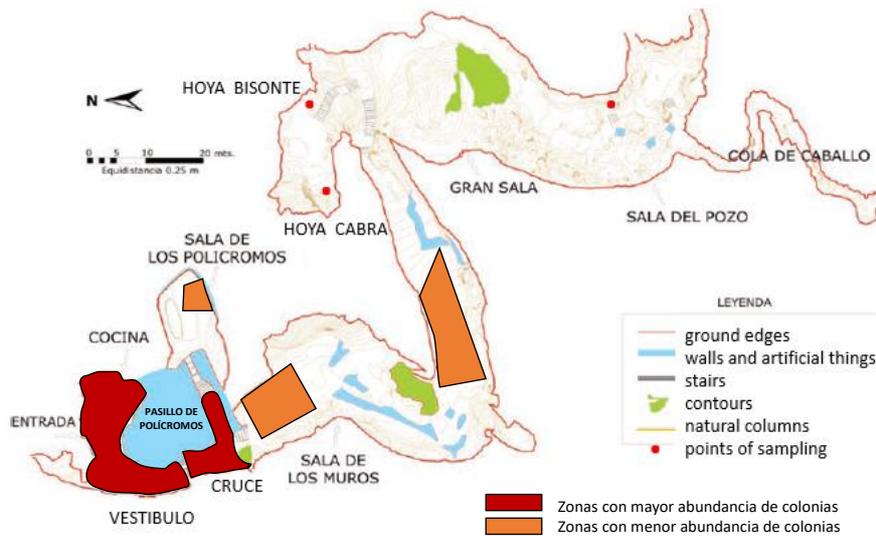


Figura 14. Situación de la distribución de las colonias visibles en el interior de la cueva de Altamira.

La primera publicación sobre colonias visibles en la cueva de Altamira y a la vez en la cueva asturiana de Tito Bustillo fue en 1999. A partir de dichas colonias se trataron de cultivar actinomicetos en una variedad de medios de cultivo (Groth et al., 1999). La mayoría de los aislados fueron del género *Streptomyces*, siendo también el más representativo en otro estudio de Altamira en muestras de agua de goteo (Laiz et al., 1999). Ahora es bien conocido que los componentes presentes en este tipo de colonizaciones son difíciles de crecer y por ello, estudios posteriores sobre colonias de Altamira se basaron en técnicas independientes de cultivo. Las primeras publicaciones con este enfoque en colonias de Altamira fueron hechas primero de colonias amarillas y grises (Portillo et al., 2008), y más tarde de blancas (Portillo et al., 2009). En estos estudios se empleó la técnica de PCR-DGGE donde el producto de interés amplificado en este caso es la región variable V3 del gen 16S rDNA. Por otra parte, también se hicieron librerías de 16S rRNA obteniendo una región más grande, en torno a 900 pb que los 200 pb utilizados en la anterior técnica. Los resultados que se obtuvieron indicaron que *Proteobacteria* fue el filo más abundante y más activo en las 3 colonias, con predominancia de la clase *Gammaproteobacteria* en colonias amarillas y blancas, y *Alphaproteobacteria* en colonias grises. Otros fillos de menor

abundancia detectados en las colonias en estos estudios fueron *Acidobacteria*, *Bacteroidetes*, *Chloroflexi*, *Gemmatimonadetes*, *Firmicutes*, *Nitrospirae*, *Planctomycetes*, *Verrucomicrobia*, y especialmente en colonias blancas *Cyanobacteria* y *Candidate division SPAM*.

Con posterioridad, se publicaron otros dos estudios basados en librerías de clones del gen 16S rDNA, un primer estudio comparativo de colonias amarillas en 3 cuevas diferentes (Porca et al., 2012) y uno segundo estudiando la estructura y composición de la colonias grises y su papel en la solución y precipitación de carbonato (Cuezva et al., 2012). En ambos estudios en este caso el filo más abundante fue *Actinobacteria* con una abundancia alrededor del 77%. En el estudio comparativo de Porca, se detectaron 3 cores de secuencias comunes a las 3 cuevas compuesto por el suborden *Pseudonocardinae* y los órdenes *Chromatiales* y *Xanthomonadales*. La presencia de estos cores sugiere la existencia de comunidades bacterianas de composición fija, especialmente desarrolladas para prosperar en condiciones de cueva. En estos últimos estudios se detectaron nuevos filios en menor abundancia en colonias amarillas como los candidatos NC10 y SBR1093.

Trabajos más recientes, no centrados en la cueva de Altamira sino en otras cuevas de España sobre colonias visibles similares como se presentan en la figura 15 (Gonzalez-Pimentel et al., 2018; Stomeo et al., 2008), amplían detalles sobre los componentes que forman estas colonias. De nuevo tanto el estudio de Pimentel (cueva Honda del Bejenado en la isla de la Palma) sobre colonias amarillas como en el de Stomeo (cueva Santimamiñe en Vizcaya y cueva Doña Trinidad en Málaga) de colonias blancas, coinciden en que el filo más abundante fue *Actinobacteria*. En concreto en la cueva de la Palma, análisis de librerías de clones de RNA y DNA indicaron que el filo más abundante y activo era *Actinobacteria*, en concreto el orden *Euzebyales*, mientras que por la técnica de amplicones, que no implica los pasos de clonaje, fue la clase *Gammaproteobacteria* y orden *Chromatiales*. Sin embargo, en el estudio de Stomeo mediante el análisis de librerías de clones de RNA y DNA, el filo más abundante y activo fue *Actinobacteria*, en concreto el género *Pseudonocardia*. Por el contrario, en los estudios de Scharabereiter en colonias

blancas, amarillas y grises en la cueva de Tito Bustillo (Schabereiter-Gurtner et al., 2002) y de colonias blancas en la Garma y Llonín (Schabereiter-Gurtner et al., 2004) usando la técnica de huella digital PCR-DGGE indicaron que en estos casos *Proteobacteria* fue el filo más abundante seguido de *Actinobacteria* y *Acidobacteria*.

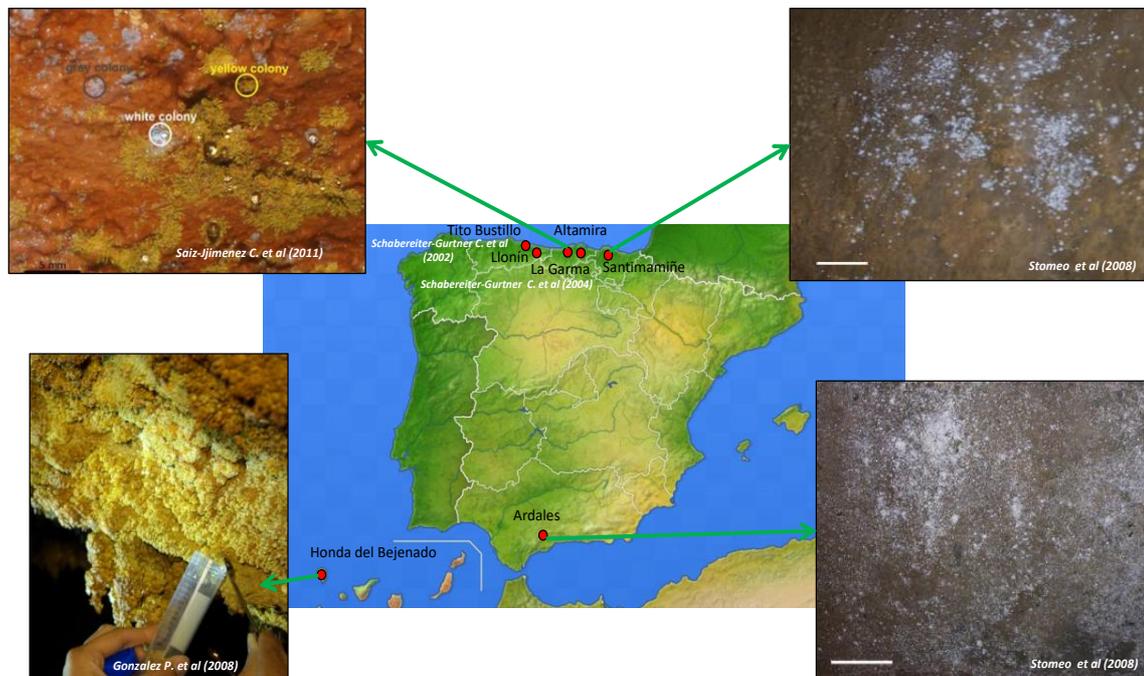


Figura 15. Estudios microbiológicos publicados sobre colonias visibles en cuevas de España.

También encontramos este tipo de colonias en cuevas similares pero con otra situación geográfica diferente, algunos ejemplos de ello se ilustran en la figura 16 (Mulec et al., 2015; Pasić et al., 2010). En el estudio de Pasic se analizaron muestras que consistían en mezclas de colonias amarillas, blancas, grises azuladas y rosas mediante el análisis de librerías de clones de DNA donde *Gammaproteobacteria* fue el taxón más abundante seguido de *Actinobacteria* y *Nitrospira* y, en menor abundancia *Alphaproteobacteria*, *Betaproteobacteria*, *Deltaproteobacteria*, *Acidobacteria*, *Verrucomicrobia*, *Planctomycetes*, *Chloroflexi* y *Gemmatimonadetes*. Por el contrario, en el estudio de Mulec sobre colonias amarillas en 3 cuevas en Eslovenia utilizando la técnica de MALDI-TOF, *Pseudomonas* fue el género obtenido con mayor frecuencia en las muestras.

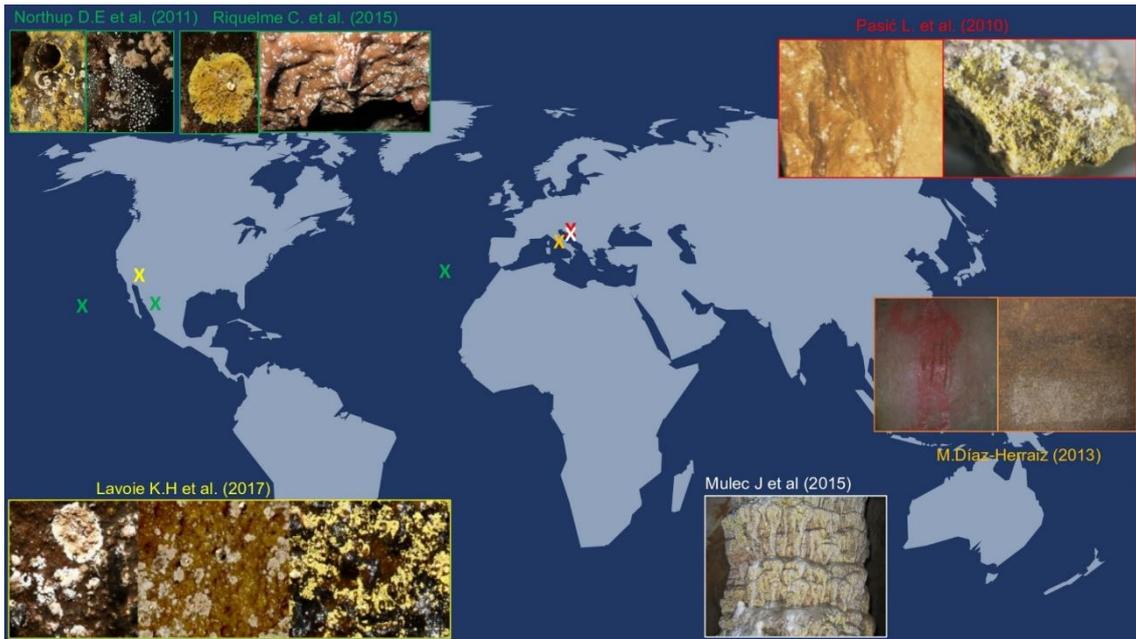


Figura 16. Estudios microbiológicos publicados sobre colonias visibles en ambientes de cuevas y otros entornos parecidos en el mundo.

El contexto geológico no parece jugar un papel crucial para el desarrollo de estas colonias, ya que también se ha detectado su presencia en cuevas volcánicas (Lavoie et al., 2017; Northup et al., 2011; Riquelme et al., 2015). En el estudio de Northup en colonias amarillas blancas y marrones mediante el análisis de librerías de clones de DNA, *Actinobacteria* fue el filo más dominante siendo *Pseudonocardiaceae* el más representativo seguido de *Proteobacteria*, donde *Gammaproteobacteria* fue la clase más prevalente. Otros filos detectados con menor abundancia en este estudio fueron *Alphaproteobacteria*, *Nitrospirae*, *Acidobacteria*, *Bacteroidetes*, *Betaproteobacteria*, *Chloroflexi*, *Deltaproteobacteria* y *Gemmatimonadetes*. En el estudio de Riquelme en colonias blancas y amarillas en varias cuevas volcánicas (Islas Azores, Islas Canarias, Portugal, Hawaii, México y Canadá) por la técnica de amplicones, *Actinobacteria* fue de nuevo el filo más abundante, en concreto las familias *Pseudonocardiaceae* y *Euzebyaceae* y el orden *Acidimicrobiales*. En el último estudio de Lavoie en algunas de las cuevas anteriores también por la técnica de amplicones estudiaron diferencias en composición de colonias visibles y depósitos secundarios de minerales de color que presentan estas cuevas, y una de las mayores diferencias fue la ausencia de *Actinobacteria* en estos

depósitos siendo *Proteobacteria*, *Nitrospirae* y *Acidobacteria* los filos más frecuentes.

También se ha prestado atención a este tipo de colonizaciones microbianas halladas en ambientes subterráneos conteniendo bienes culturales con valor patrimonial como tumbas o catatumbas (Díaz Herráiz et al., 2013; Krakova et al., 2015). En estos estudios utilizando las técnicas de PCR-DGGE y librería de clones de DNA y RNA, *Actinobacteria*, principalmente *Nocardia* y *Pseudonocardia* fueron los taxones más representativos en las colonias blancas que colonizan estos ambientes.

Claramente, los estudios mencionados anteriormente en colonias visibles que se han llevado a cabo en una variedad de ecosistemas tanto en cuevas kársticas como volcánicas o en ambientes subterráneos subrayan la importancia de alta presencia de filotipos los cuales se podrían considerar típicos de cueva como *Actinobacteria* o *Proteobacteria* junto con otros en menor presencia como *Acidobacteria* o *Nitrospirae*.

### **2.2.3 Comunidades más frecuentes en cuevas**

En una cueva no sólo encontramos diversidad biológica en estos biofilms característicos, sino que también los suelos, las paredes, el agua de goteo, el aire o los espeleotemas sirven de nichos para sustentar el desarrollo o dispersión de bacterias, hongos, arqueas, virus o algas. Su presencia es vital para la cueva puesto que su metabolismo influye en la dinámica ecológica de la misma, jugando papeles esenciales en los ciclos del nitrógeno y del carbono (De Mandal et al., 2017; Wiseschart et al., 2019). Estas comunidades suelen ser naturales de estos ecosistemas o en ocasiones accidentales, ya sean transmitidas al interior de la cueva vía aire, agua, personas o animales.

Uno de los aspectos que mejor define las comunidades microbianas es su clasificación en grupos nutricionales según su fuente de carbono y energía. Dos variables ambientales condicionan el tipo nutricional de los microorganismos que nos interesan en la cueva de Altamira, la falta de luz y la

presencia de oxígeno. Como se ha mencionado en apartados anteriores en la cueva hay una baja disponibilidad de compuestos orgánicos preformados. Las puertas de la cueva impiden una entrada masiva de materia orgánica, que sin embargo, acaba entrando disuelta en el agua de infiltración que procede del exterior o en pequeñas partículas dispersas en el aire como se ha demostrado en estudios previos en Altamira (Cañaveras et al., 1999; Laiz et al., 1999; Saiz-Jimenez and Hermosin, 1999; Sánchez-Moral et al., 1999). Otra vía de entrada, pero no tan consistente, es el aporte accidental de material vegetal al interior de la cueva en los pies de los visitantes que entran en la cueva o transportados por animales.

A pesar de las condiciones oligotróficas (<2mg de carbono orgánico por litro) que presentan estos entornos, estas cantidades limitadas de nutrientes podrían ser suficientes para sustentar el desarrollo de la biodiversidad microbiana de la cueva por un metabolismo de tipo heterotrófico, que podría obtener la energía por oxidación de esa materia orgánica con el oxígeno como aceptor final de electrones. Existe la opinión mayoritaria de que ésta es la forma principal de obtención de energía que utiliza la microbiota de Altamira.

Otra posibilidad, sin embargo, sería la presencia de organismos autótrofos capaces de sintetizar *de novo* sus compuestos de carbono a partir del abundante CO<sub>2</sub> presente en la atmósfera de Altamira, constituyendo fuentes de materia orgánica para el resto de microorganismos que no son capaces de sintetizarlos. Cuando no hay luz que energice este proceso de fijación, la fotoautotrofia no es una opción, pero existen vías alternativas como la actividad derivada de organismos quimiolitótrofos. Estos son capaces de obtener energía y poder reductor a base de oxidar compuestos inorgánicos reducidos, como por ejemplo hidrógeno, sulfuros, amonio, etc... y luego utilizar esa energía y poder reductor para energizar la fijación en la oscuridad del CO<sub>2</sub> atmosférico.

En la literatura encontramos numerosos estudios en cuevas del mundo para tratar de caracterizar los microorganismos que habitan en estos ecosistemas, una muestra de ellos se puede encontrar en la figura 17. Estos estudios definen la biodiversidad microbiana detectada tanto en las rocas de paredes

y/o techos como depósitos de minerales, suelos, agua de goteo, zonas que presentan pinturas de arte rupestre, biofilms o zonas inundables conocidas como zonas epifáticas. Los taxones más frecuentemente encontrados en estos entornos son *Proteobacteria*, *Actinobacteria*, *Nitrospirae*, *Acidobacteria*, *Bacteroidetes*, *Firmicutes*, *Verrucomicrobia*, *Plantomycetes*, *Chloroflexi*, *Gemmatimonadetes*, *Spirochetes*, *Cyanobacteria*, *Chlorobi* y *Fibrobacteres* (Tomczyk-Żak and Zielenkiewicz, 2016).

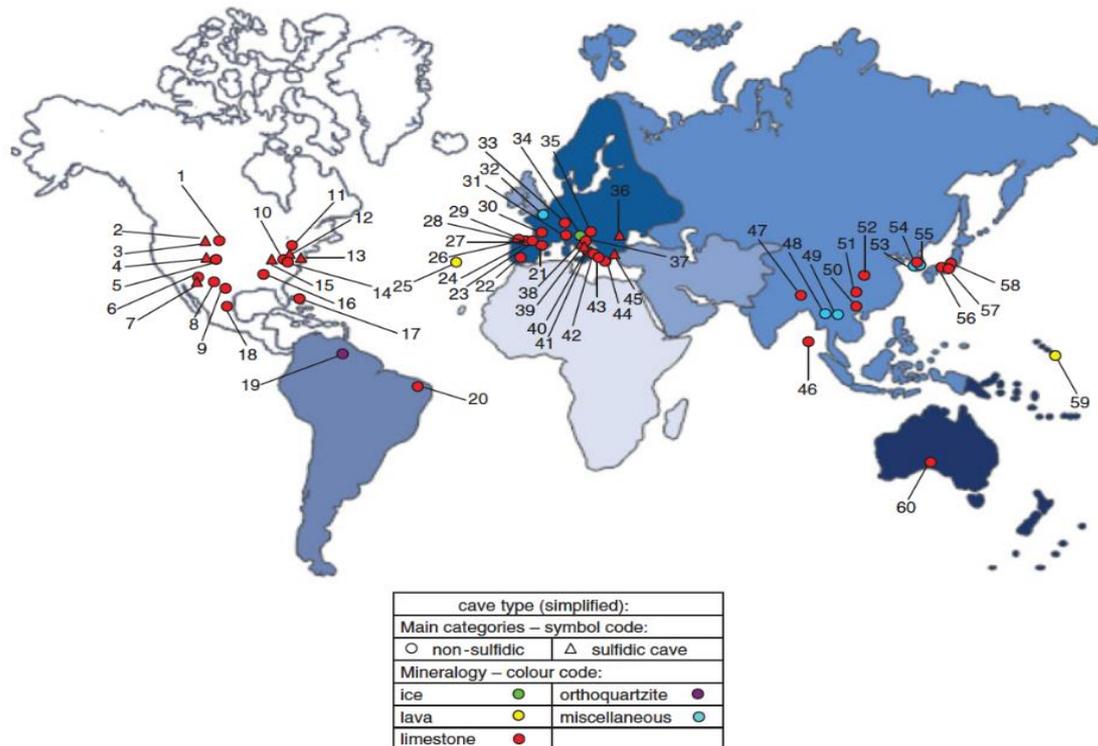


Figura 17. Estudios microbiológicos en cuevas del mundo. Figura obtenida de Lee et al., 2012.

Entre ellos, *Proteobacteria* es uno de los grandes grupos de bacterias más comunes en cuevas. Las bacterias de este grupo tienen un metabolismo muy diverso pudiendo ser fotótrofos, heterótrofos y quimiolitótrofos. Este filo representa 6 clases de las cuales *alpha*, *beta* y *gamma* son las más abundantes en cuevas. Estas tres clases más abundantes incluyen formas oligotróficas que fijan nitrógeno y degradan urea a CO<sub>2</sub> y amonio, asociados a procesos de oxidación del amonio (Alfreider et al., 2009; Hathaway et al., 2014; Sauro et al., 2018). Las otras dos clases, *delta* y *epsilon*, son más frecuentes en cuevas con un contenido alto en azufre involucrados en procesos de

oxidación y reducción (Elshahed et al., 2003; Engel et al., 2004; Jones et al., 2010; Macalady et al., 2008; Rossmassler et al., 2012).

*Actinobacteria* es otro de los grandes filotipos abundantemente encontrados en cuevas sobre todo muy relacionado a la precipitación de carbonato cálcico, conocido este proceso también como biomineralización (Cañaveras et al., 1999; Cao et al., 2016; Cuezva et al., 2009, 2012). Las bacterias de este grupo resultan muy interesantes puesto que poseen un metabolismo secundario muy diverso y prometedor tanto en la industria bioenergética como en el campo clínico (Lewin et al., 2016; Rangseekaew and Pathom-aree, 2019). Suelen presentar un metabolismo heterótrofo siendo capaces de utilizar como fuente de carbono orgánico la celulosa, hemicelulosa o quitina que suelen ser compuestos muy abundantes en la naturaleza provenientes de material vegetal en descomposición, incluida la hojarasca, el compost o el estiércol. Aunque por otra parte, cuando las condiciones nutricionales lo requieren también pueden presentar un metabolismo oligotrófico fijando CO<sub>2</sub> lo que se conoce como mixotrofia (Ji et al., 2017; Kouki et al., 2011; Niederberger et al., 2015).

Las arqueas también son un importante componente dentro de estas comunidades de cuevas. Su metabolismo también es muy diverso desde aerobias a anaerobias y desde quimiolitótrofas a quimioheterótrofas (Tomczyk-Żak and Zielenkiewicz, 2016). Las clases más encontradas son *Crenarchaeota*, *Euryarchaeota* y *Thaumarchaeota* (Barton et al., 2014; Chronáková et al., 2009; Gonzalez et al., 2006). Su papel en las cuevas normalmente es oxidar amonio contribuyendo al proceso de nitrificación en el ciclo global de nitrógeno (He et al., 2019; Zhao et al., 2017).

Las comunidades de eucariotas, en especial hongos, son parte también de la biodiversidad en cuevas. En todas las cuevas existe una población de hongos más o menos estable que no se hace visible, probablemente debido a las condiciones de baja disponibilidad de nutrientes en condiciones habituales, y que en general son de origen desconocido, aunque muchos de los hongos identificados en cuevas se consideran habitantes especializados de cuevas (Vanderwolf et al., 2013). De hecho, un estudio hecho en la cueva Jarkowicka

(Polonia), recién descubierta y no muy frecuentada por el hombre, describe una composición similar de especies de hongos a la encontrada en cuevas que presentan un mayor flujo de visitas (Pusz et al., 2015). Aun así, la colonización de estos en las paredes y techos de cuevas no suele ser aparente y solamente pasa a ser visible a consecuencia de factores abióticos a menudo causados por el hombre, como pueden ser el empleo de biocidas como pasó en la cueva de Lascaux (Bastian et al., 2010), perturbaciones que alteren el microclima (Fernández Cortés et al., 2011; Jurado et al., 2009a) o una disponibilidad de nutrientes extraordinaria (Nováková et al., 2018). Por consiguiente, muchos de los hongos identificados están asociados a la degradación de restos lignocelulósicos, o son comensales o patógenos de plantas o entomófagos.

El aire es la principal vía de dispersión de los hongos, o mejor de sus esporas, del exterior al interior de la cueva (García Antón et al., 2013; Porca et al., 2011). Los taxones más frecuentemente aislados de muestras de aire son *Aspergillus*, *Cladosporium* y *Penicillium* (Docampo et al., 2011; Fernández Cortés et al., 2011; Pusz et al., 2015).

Entre los hongos presentes en la cueva de Altamira destaca la presencia de rizomorfos correspondientes a hongos saprófitos degradadores de lignocelulosa, generalmente del filo *Basidiomycota*. Estos rizomorfos son muy evidentes con un fuerte impacto visual, y a diferencia de los brotes esporádicos efímeros observados sobre cadáveres de insectos o restos vegetales esporádicos, son estructuras que permanecen visibles por largos periodos de tiempo. Los rizomorfos son órganos complejos formados por los micelios de algunos hongos que permiten el transporte de nutrientes a largas distancias, permitiendo el desarrollo del hongo lejos de su fuente de alimento. En la cueva de Altamira hemos identificado rizomorfos pertenecientes al basidiomiceto *Coprinellus*, en varias localizaciones, siempre asociados a muros artificiales de construcción humana (Figura 18). No se ha podido establecer si estos rizomorfos conectan con poblaciones de *Coprinellus* presentes en el exterior, en la cobertera vegetal (lo más probable) o si por el contrario provienen del hongo que colonizaba y ha degradado las estructuras de madera utilizadas para la construcción de los muros y luego abandonadas en su interior.

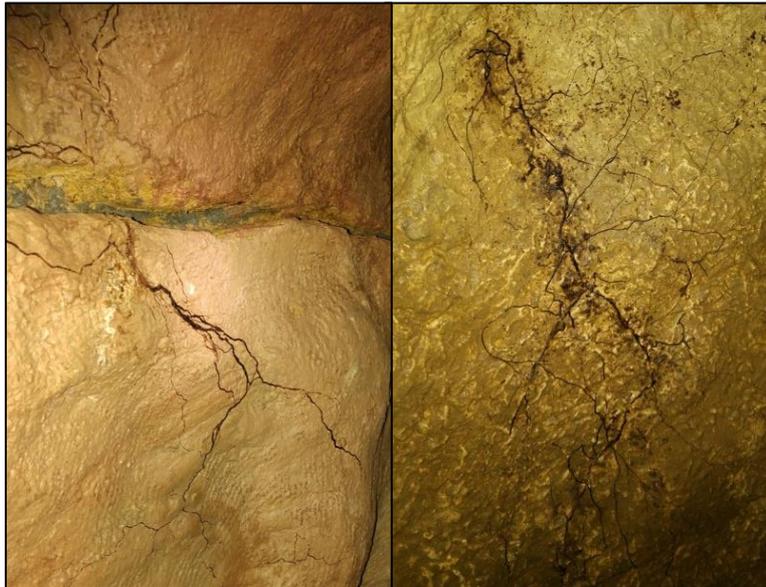


Figura 18. Fotografía de rizomorfos en la cueva de Altamira en muros artificiales. La foto de la izquierda está hecha en el muro de la sala del Pozo y la de la derecha en la sala de los Muros tomada el 29.11.16. El rizomorfo de los Muros pertenece al basidiomiceto *Coprinellus disseminatus* (mismo resultado obtenido en otras ocasiones anteriores) y el del Pozo presenta mejor identidad con *Coprinellus micaceus* (hongo similar al anterior, pero especie diferente).

La presencia de comunidades fotosintéticas como algas o cianobacterias es también muy común en cuevas (Czerwik-Marcinkowska and Mrozińska, 2011). La alta humedad, una fuente de luz constante, filtración de agua que suelen contener un aporte de micronutrientes y altas concentraciones de CO<sub>2</sub> son factores que ayudan en el crecimiento de estas comunidades (Grobbelaar, 2000). Por tanto, suelen desarrollarse en la zona de la entrada dónde la luz es más accesible. En cuevas abiertas al público, la iluminación por la luz de las linternas y focos permite el desarrollo de organismos fotosintéticos en zonas dónde previamente no podían hacerlo siendo esta una de las causas mejor conocidas de biodeterioro. La cueva de Altamira ha sufrido casos de colonización visible en el periodo de 2003 a 2005 por cianobacterias del orden *Chroococcales* y *Oscillatoriales*, y de algas unicelulares eucariotas en concreto de los géneros *Cyanidium*, *Chladmydomonas* y *Chlorococcum* en menor frecuencia (Informe CSIC 2003-2005 pag 88-91). Su aparición en la cueva se debió a la presencia de puntos de iluminación fijos en la sala de las pinturas

acondicionados para las visitas que se acabaron retirando. En septiembre de 2016 se detectó otro caso de colonización de microorganismos similares (Figura 19), en este caso se identificó por secuenciación un alga perteneciente a la especie *Jenufa aeroterrestica*, cerca de unos trazos de pintura negra, dónde hasta el 2000 hubo un foco de luz fijo. Desde entonces no hay luz en este punto, salvo en el momento de las visitas públicas. A pesar de esto, se sigue recuperando el alga de la pared, permitiendo pensar que esta alga presenta un metabolismo mixotrófico capaz de actuar como un organismo tanto autotrófico como heterotrófico según las circunstancias ambientales.



Figura 19. Imágenes obtenidas de la alga especie *Jenufa aeroterrestica* detectada en zonas verdosas en la cueva de Altamira. La foto de la izquierda corresponde con manchas verdosas detectadas en la zona del pasillo de camino a la Hoya, cerca de zonas con trazos de pintura negra, como consecuencia de colonización de organismos fotosintéticos. La foto de la derecha es una imagen obtenida con el microscopio de luz visible en las muestras recogidas de zonas verdosas con el objetivo de 100X dónde se detecta la presencia única de esta especie.

Los protozoos son otro de los grupos taxonómicos cuya presencia en el suelo y aguas superficiales es muy abundante y esporádica, aunque no con mucha frecuencia, su presencia también es descrita en cuevas (García Sánchez et al., 2013; Walochnik and Mulec, 2009). Nuestros resultados de secuenciación de DNA total de colonias o soportes en la cueva de Altamira no han revelado la presencia de DNA originario del reino, salvo en el caso de una muestra de moonmilk tomada en una columna artificial que permitió detectar un DNA circular conteniendo un operón rDNA que pudo ser identificado con un 97% de

identidad con la ameba de vida libre *Allovahlkampfia spelaea* (García Lobo, JM, datos no publicados). Esta ameba localizada en múltiples ambientes ha sido identificada en varios casos de queratitis humana (Tolba et al., 2016).

#### **2.2.4 Técnicas de cultivo de bacterias difíciles de crecer**

En los últimos años se ha dedicado mucho esfuerzo al cultivo de bacterias, hasta el momento no cultivables, que son las que se localizan con más frecuencia en ambientes extremos y también en suelos y cuevas. En el caso de comunidades bacterianas que han crecido en condiciones naturales, como es el caso que nos ocupa de las “colonias visibles” de las cuevas, tenemos que considerar que son perfectamente capaces de crecer y que, si no conseguimos hacerlo en el laboratorio, es porque no conocemos o no somos capaces de reproducir las condiciones naturales de crecimiento.

Una causa frecuente de fracaso es el intento recurrente de recuperar las bacterias del medio natural por crecimiento directo en medios de cultivo estándar. El uso de medios ricos en el laboratorio impide por lo general la recuperación de bacterias de interés, ya que, al sembrar mezclas complejas de bacterias, se favorece el crecimiento de bacterias de crecimiento rápido impidiendo el de aquellas con un crecimiento más lento a desarrollarse, las cuales también podrían estar inhibidas por concentraciones altas de nutrientes. Una solución frente a este problema sería usar medios diluidos, o los llamados medios oligotróficos, en los que las condiciones nutricionales son más adecuadas para que bacterias de crecimiento más lento, con una baja tasa de metabolismo y baja densidad de población, puedan desarrollarse. Además, junto con el uso de medios pobres, hay que considerar que serán necesarias las condiciones ambientales adecuadas y largos periodos de incubación para tener éxito en los cultivos.

A veces, la simple preparación de los medios bacteriológicos produce inhibición del crecimiento de algunas bacterias. Está comprobado que durante el proceso de autoclavado, algún componente del agar reacciona con el fosfato en el

medio de cultivo dando lugar a un descenso del número de colonias obtenidas. Este efecto se atribuye a la producción de peróxido de hidrógeno que inhibe el crecimiento de bacterias que no producen catalasa, enzima que cataliza la descomposición del peróxido de hidrógeno en oxígeno y agua (Tanaka et al., 2014). Para evitar este problema se recomienda autoclavar separadamente el agar y los nutrientes del medio o sustituir el agar por otro agente solidificante para los medios sólidos (Pulschen et al., 2017; Vartoukian et al., 2010). En otros casos, algunas bacterias necesitan de la adición al medio de ciertos compuestos químicos como lacasas, AMP cíclico (cAMP) o N-Acil Homoserina Lactonas, o simplemente de extractos de suelo o sustratos naturales, empleando además cambios de temperatura o diferentes concentraciones de medio (Bruns et al., 2002; Kong et al., 2014; Liu et al., 2018).

Otros intentos de recuperación en cultivo se enfocan al cocultivo con bacterias “asistentes” cultivables o en presencia de sus productos, ya que estas pueden secretar factores específicos, como por ejemplo sideróforos, que estimulen el crecimiento de bacterias VBNC como se muestra en la figura 20B (D’Onofrio et al., 2010; Stewart, 2012). A veces no se precisa el cocultivo sino que simplemente, añadiendo el factor purificado o el medio libre de células en el que se ha cultivado previamente el “asistente” es suficiente para promover este crecimiento. Una de las moléculas estimuladoras del crecimiento más conocidas hasta el momento entre bacterias del filo *Actinobacteria* son las llamadas proteínas Rpf (“Resuscitation Promoting Factor”), descubiertas por primera vez en la bacteria *Micrococcus luteus* (Mukamolova et al., 2002). El factor Rpf promueve la reanimación de otras actinobacterias que se encuentran en un estado durmiente a bacterias capaces de formar colonias. Con el paso del tiempo este factor se ha encontrado en bacterias Gram positivas con alto contenido en GC dentro de este filotipo como *Mycobacterium spp.*, *Corynebacterium spp.* o *Streptomyces spp* (Gupta and Srivastava, 2012). Se sabe que las proteínas Rpf son peptidoglucano glucosidasas, aunque el mecanismo y su papel como mediador en la resucitación de células durmientes sigue sin estar claro.

Este tipo de cooperación se ha logrado entre bacterias de diferentes filos como se muestra nuevamente en la figura 20A (Cross et al., 2019; St. John et al., 2019; Wurch et al., 2016) o con otro tipo de células como eucarióticas (HT-29, Caco-2, T84, HeLa, Intestine 407, y CHO) en el caso de patógenos humanos como las bacterias VBNC de *Vibrio cholerae non-O1/non-O139*, *V. parahaemolyticus*, enterohemorrágico *Escherichia coli*, enterotoxigénico *E. coli*, enteropatógeno *E. coli*, *Shigella flexneri* y *Salmonella entérica* (Senoh et al., 2012).

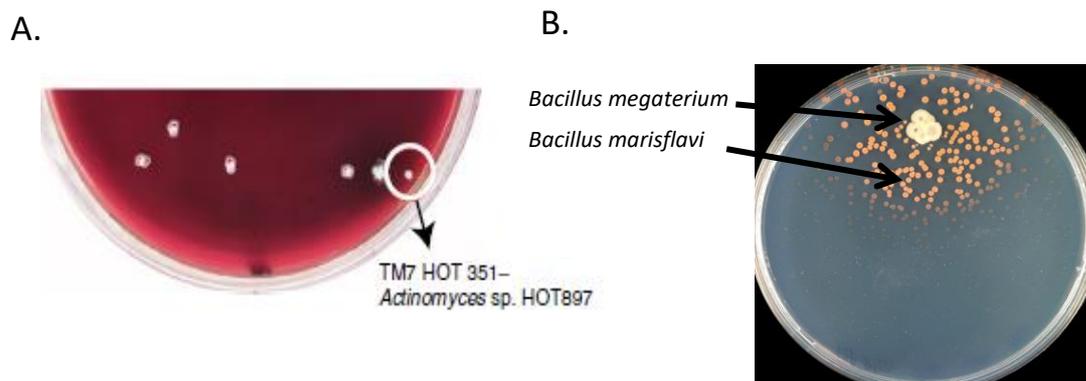


Figura 20. Ejemplos de cocultivo entre bacterias cultivables y bacterias VBNC. A. Cocultivo de bacteria VBNC *TM7* HOT351 y la bacteria cultivable *Actinomyces* sp. HOT897 en agar sangre BHI (Figura obtenida de Cross et al., 2019). B. Cocultivo de VBNC *Bacillus marisflavi* con la ayuda del plaqueo de *Bacillus megaterium*, en este caso esta bacteria se considera “asistente”, ya que secreta al medio sideróforos para que la otra bacteria pueda crecer (Figura obtenida de Stewart, 2012).

Dada la gran variedad de condiciones ambientales o nutricionales que habría que combinar hasta dar con las condiciones permisivas para un cultivo *in vitro*, recientemente se han diseñado tecnologías de alto rendimiento (High throughput) para simplificar esta búsqueda. Uno de los primeros intentos en esta dirección consistía en intentar simular *in vitro* los ambientes en los que se encuentran originalmente las bacterias no cultivables o intentar crecer estas bacterias nuevas *in situ*, a través del diseño de cámaras de difusión o bioreactores que permiten el paso de sustancias esenciales del ambiente original al medio de cultivo a través de membranas semipermeables (Figura 21). Uno de estos dispositivos bien conocido es el llamado “Isolation Chip” o simplemente “ichip” el cual se puede colocar fácilmente *in situ* en el ambiente

natural y permite capturar un solo tipo celular (Berdy et al., 2017). Las células se diluyen para que se obtenga una célula por agujero y se suspenden en un medio líquido de agar fundido. Cada microcanal va a capturar un microvolumen de esta suspensión de tal modo que cuando el agar solidifica, las células quedan inmovilizadas dentro del agar quedando así atrapadas individualmente en cada microcanal y el crecimiento resultante será un monocultivo. Otras cámaras de difusión similares se han diseñado para el cultivo de bacterias incultivables de suelo (Ferrari et al., 2005; Kakumanu and Williams, 2012) o de sedimentos y muestras de agua dulce de estanques (Bollmann et al., 2007).



Figura 21. Ejemplo de un dispositivo “ichip” situado en el ambiente para el cultivo de microorganismos incultivables de suelo.

Otro ejemplo de una aplicación de este tipo es el diseño de biorreactores de difusión que emplean una membrana que separa el ambiente natural del medio de cultivo inoculado con el ambiente natural (Figura 22). Por ejemplo, en el estudio de Chaudhary et al. 2019, el biorreactor diseñado logró la recuperación de 35 nuevos aislados de cepas previamente no cultivadas en comparación a utilizar sólo el medio de cultivo con suelo añadido.

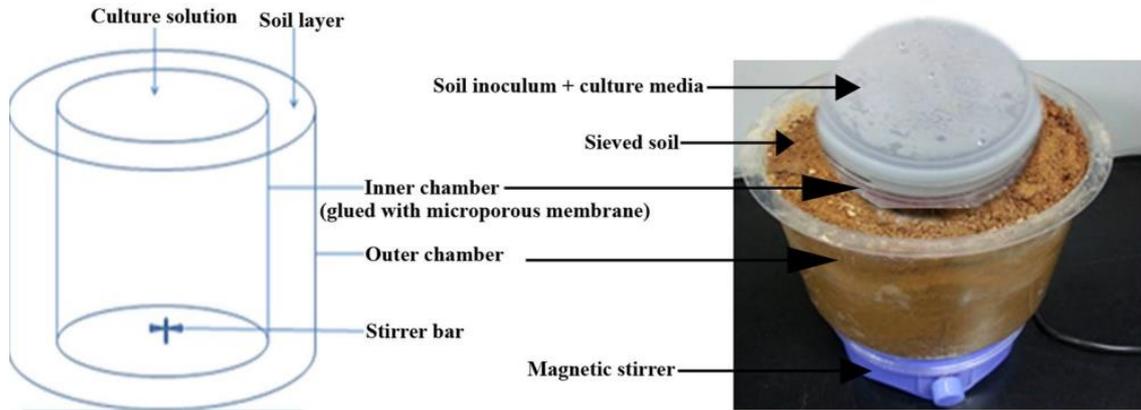


Figura 22. Ejemplo del diseño de un biorreactor de difusión para el cultivo de microorganismos incultivables, en este caso de suelo. Figura obtenida de Chaudhary et al., 2019.

Los procesos descritos hasta aquí intentan la recuperación de componentes aislados como cultivos puros. Sin embargo, es posible que esta estrategia sea equivocada para tratar de analizar consorcios como los que forman las colonias visibles en los que previsiblemente varias especies bacterianas o de arqueas pueden cooperar entre sí compartiendo productos o señales. En el caso de consorcios complejos es probable que sea actualmente imposible recuperar componentes individuales en forma de cultivos puros.

En estos casos la única opción consistiría en un intento de reconstrucción de los consorcios naturales a partir de consorcios preexistentes o bien a partir de sus componentes individuales. La construcción de consorcios a partir de sus componentes es un objetivo actual, aun no resuelto para resolver diferentes cuestiones biotecnológicas (Rosero-Chasoy et al., 2021). La reproducción de consorcios recuperados de su hábitat natural es una opción más directa, aunque ofrece menos opciones de control. El problema central puede centrarse en la reproducción de las condiciones de desarrollo en su hábitat natural, lo que podría conseguirse por alguno de los procedimientos descritos para cultivos puros como el uso de biorreactores, o de sistemas tipo ichip que permitan simular las condiciones naturales.

### **2.2.5 La implicación negativa de la proliferación de microorganismos en cuevas**

En una cueva, como ecosistema natural, la colonización microbiana debe considerarse algo natural y si no se introducen alteraciones graves de las condiciones ambientales o nutricionales, se va a mantener dentro de los límites marcado por los equilibrios biológicos naturales. Cuando una cueva contiene un valor añadido, como es el caso de Altamira, la proliferación microbiana es una preocupación por el riesgo que supone, en este caso para la conservación de las pinturas. Los materiales empleados por los artistas paleolíticos eran en su mayor parte minerales (óxidos de hierro) o carbón, de los que no tenemos evidencia de que sirvan específicamente como nutrientes que soporten la proliferación microbiana. Por lo tanto, en una cueva no hay a priori un riesgo específico de colonización de las pinturas. En Altamira, afortunadamente, la proliferación bacteriana es mucho más grave en áreas sin pintura.

Además del deterioro visual o estético que se produciría por la aparición sobre las pinturas de zonas colonizadas por consorcios microbianos pigmentados, el biodeterioro del arte parietal en cuevas por proliferación microbiana puede derivar de dos efectos: a) Debilitamiento o disgregación de la roca soporte por acción de los ácidos producidos como resultado del metabolismo microbiano. Este efecto puede provocar descascarillados y arrastre y pérdida de pigmento. b) Precipitación o cristalización de sales minerales sobre la pintura. En nuestro caso la alteración del pH podría alterar el equilibrio del sistema carbonato/bicarbonato y provocar deposición de cristales de carbonatos que producen veladuras sobre grabados o zonas pintadas que pueden ocultar o destruir las figuras.

Este proceso en su conjunto es conocido como biodeterioro y no solo afecta a cuevas sino que también es evidente en numerosas superficies como edificios arquitectónicos o monumentos, piezas de arte en museos, manuscritos y/o libros en bibliotecas, murales en iglesias o catacumbas etc... (Barton et al., 2014).

La pérdida del material en bienes culturales puede ser causada por ejemplo por la acción combinada por la corrosión de ácidos orgánicos y ácidos minerales (ácido nítrico, sulfúrico o carbónico), la degradación enzimática y/o mecánica, el incremento de la solubilidad de compuestos poco solubles por agentes quelantes y la producción de compuestos nocivos como óxidos de nitrógeno (Sand, 1997). Casos estudiados de ácidos orgánicos como ácidos glucónico, cítrico y oxálico producidos por hongos como *Aspergillus niger* (Biswas et al., 2013) o ácidos fumárico y succínico por especies de *Penicillium* y *Fusarium* (Gómez-Alarcón et al., 1994) o sulfuros y ácido sulfúrico por bacterias reductoras de sulfato (Enning and Garrelfs, 2014; Spruit and Wanklyn, 1951) son ejemplos estudiados de compuestos biológicos con un efecto corrosivo en entornos naturales.

Otro factor importante sobre todo en cuevas abiertas al público es el impacto de una presencia humana excesiva que podría contribuir a la propagación e introducción de microorganismos no autóctonos que comprometan la integridad de la cueva (Chelius et al., 2009; Mulec et al., 2015). Un factor de riesgo aún más importante en estas cuevas es la alteración de las condiciones ambientales (temperatura, humedad, concentración de CO<sub>2</sub>) y el aporte de materia orgánica extra al interior de la cueva producido por un régimen elevado de visitas. Estas alteraciones van a producir cambios en los equilibrios biológicos y pueden dar lugar a proliferación inesperada de microorganismos y por tanto causar problemas de biodeterioro.

El caso mejor conocido de biodeterioro en cuevas con arte rupestre producido por desequilibrios provocados por el hombre fue el ocurrido en la cueva de Lascaux (Francia). La cueva durante años sufrió tres tipos de crisis microbiana que acabaron por causar un efecto irreversible en las pinturas. La primera crisis microbiana fue en 1960 causada por el uso de luz artificial a consecuencia del turismo masivo que sufría la cueva llamada "La Maladie Verte", y ocasionado por el crecimiento primero del alga *Chlorobotrys* y más tarde de *Bracteacoccus minor*. Como consecuencia, en 1963, emplearon un spray combinado de estreptomina y penicilina para bacterias y un tratamiento de formaldehído para algas. Estas acciones condujeron a la aparición consecutiva de

selecciones de comunidades no autóctonas en los posteriores años que acabaron en la pérdida de las pinturas. En 2001 apareció la segunda crisis microbiana llamada “La Maladie Blanche” por el hongo *Fusarium solani* y la bacteria *Pseudomonas fluorescens* (Figura 23B). A consecuencia, en 2001 usaron un tratamiento basado en soluciones de cloruro de benzalconio (Vitalub QC 50) más estreptomicina y polimixina (Figura 23A). Finalmente, después en 2006 surgió la tercera crisis “La Maladie Noire” por los hongos *Scolecobasidium tshawytschae* y *Ochroconis lascauxensis* y el artrópodo *Folsomia Candida*.



Figura 23. Ejemplo de biodeterioro en la cueva de Lascaux (Francia). A. Tratamiento de biocidas en las paredes de la cueva; B. Colonización por el hongo *F.solani* en una de las salas de la cueva.

En Altamira, si nos fijamos en el estado actual del techo de Polícromos, que es la sala de las pinturas, podemos preguntarnos cómo ha evolucionado con el tiempo y cuales han sido los factores que han determinado su evolución. La superficie del techo es de 159 m<sup>2</sup> y le separan en torno a 10 metros de la superficie exterior. La mayor parte del techo estaba cubierta originalmente por pinturas, de las que a día de hoy se conserva la mitad mientras que la otra mitad se ha perdido principalmente por lavado con arrastre de pigmento o desprendimiento de placas de sustrato superficial con arrastre de pintura (Figura 24). El techo presenta una importante fractura longitudinal que va desde la entrada a la parte final de la sala que fue sellada con cemento en 1928 debido a la infiltración abundante de agua directamente de la lluvia.

Además de la fisura central, existe una red relativamente densa de fisuras transversales.

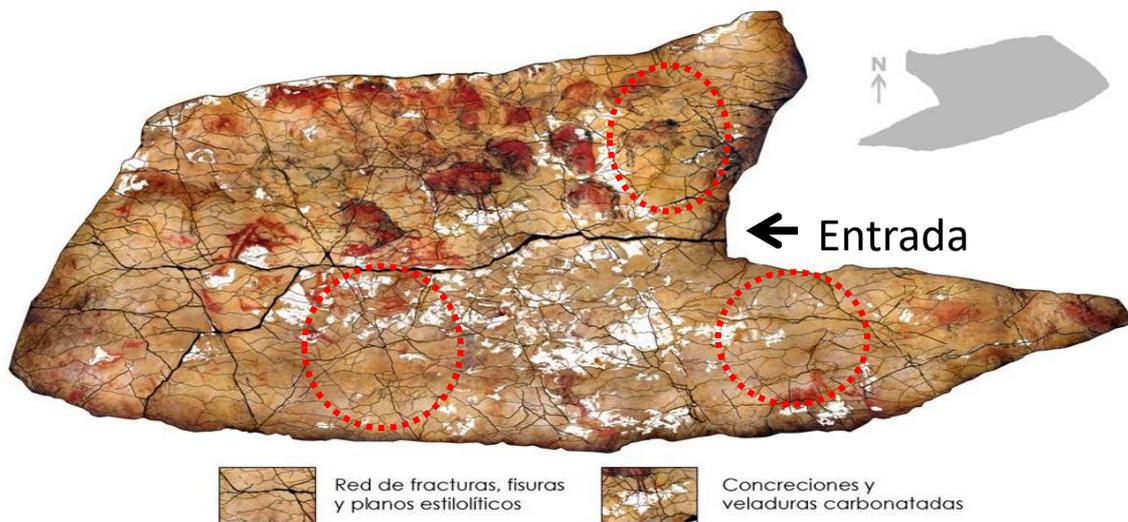


Figura 24. Zonas afectadas por la presencia de colonias blancas y grises en la sala de Polícromos. Estas zonas se muestran con un círculo rojo. Imagen editada del Informe del Programa de Investigación para la Conservación Preventiva y Régimen de acceso a la cueva de Altamira (2012-2014) Volumen I.

Así, en condiciones naturales, existe una entrada de agua permanente a través de las fisuras creando un sistema de cuencas bien definido con unos puntos de goteo estables. Es de notar una importante disminución del caudal de agua de infiltración en los últimos años. La infiltración y la humedad permanente presente en la sala producen una fuerte corrosión, una aceleración de la precipitación de carbonatos (como se observa en la figura 24) y un fuerte arrastre y lavado de pigmentos. Podemos decir que la mayor causa de deterioro en Altamira ha sido el paso del tiempo y los procesos físicos e hidrológicos naturales.

También se han sufrido episodios de biodeterioro en Altamira. Quizás los primeros detectados fueron generados por proliferaciones de algas verdes causadas por la iluminación artificial y que fueron controladas con su eliminación. También se han observado casos recientes de aparición de veladuras de carbonato en algunas figuras mayores, que afortunadamente han sido reversibles. Sobre la aparición de colonias visibles, no se documentan

hasta los últimos años del siglo XX. En la actualidad son muy abundantes, especialmente en las zonas más exteriores de la cueva, Vestíbulo y Cruce. La sala de Polícromos no se encuentra tan colonizada por colonias visibles como lo están el Vestíbulo, Cruce y pasillo de Polícromos (Figura 14). Sin embargo, sí que se observan puntos de presencia de colonias blancas y grises en zonas del techo de la sala marcadas como círculos rojos en la figura 24.

Una cuestión fundamental en la gestión de la cueva es determinar el impacto que el régimen de visitas puede jugar en la aparición de cambios ambientales o de colonización por microorganismos. Grupos de geólogos y microbiólogos realizaron diversos estudios para ver el impacto de las visitas en el interior de la cueva y, en especial en la sala de las pinturas (Sala de Polícromos). Factores como la temperatura o la concentración de CO<sub>2</sub> aumentaban tras la visita de 8 grupos de 6 personas al día y con un tiempo estimado de 30 min en el interior (Saiz-Jimenez et al., 2011). Cada grupo representaba un aumento en el interior de la cueva en la temperatura de +0.14°C y +110 ppm of CO<sub>2</sub>. Especialmente, se demostró que las visitas en la sala de Polícromos donde la temperatura de la roca de la pared es ligeramente inferior a la temperatura del aire circulante, se aceleraba el proceso de microcorrosión en las paredes y techos (Sánchez-Moral et al., 1999). En otras palabras, un aumento en la temperatura por un grupo de visitas aumenta la condensación en las paredes, así como la disolución del CO<sub>2</sub> del aire en el agua que es una reacción de pocos minutos, y también la formación de ácido carbónico, con lo que el pH desciende y aumenta la disolución de la calcita presente en la roca. Otro estudio (Sánchez-Moral et al., 1999) revisado en la tesis de Portillo (2008), menciona que este régimen de visitas conllevaba a la formación de una película de agua 0,3 mm sobre el techo de Polícromos incrementando el proceso de corrosión entre 80 y 250 veces en relación a la producida en condiciones naturales.

En la actualidad la cueva de Altamira mantiene un régimen de visitas públicas limitado a la entrada de un grupo semanal de 5 personas (más dos guías) que permanecen 35 minutos en la cueva, de los cuales, 5 minutos son en Polícromos. Además, la cueva recibe visitas con objeto de conservación experimental que generalmente se mantienen por debajo de las 100 horas

persona mensuales de las que solo una pequeña parte es en Polícromos. Los parámetros de control no muestran señales apreciables de alteración por estas visitas. Los efectos sobre la temperatura y nivel de CO<sub>2</sub> se recuperan en poco tiempo y este régimen de visitas, siempre sujeto a monitorización, se considera de nuevo un compromiso aceptable entre conservación y disfrute de un bien cultural.

Otro riesgo considerado es el posible efecto de los microorganismos de la cueva sobre la salud de las personas que visiten la cueva. En algunos casos se ha descrito la presencia en cuevas de patógenos específicos para algunas especies animales. Un caso muy estudiado es el del síndrome de la nariz blanca en murciélagos que hibernan en cuevas, causado por el hongo *Pseudogymnoascus destructans* en muchas de las cuevas de Estados Unidos (Blehert et al., 2009; Lorch et al., 2016). Este patógeno que provoca la muerte de millones de murciélagos se ha visto asociado a ambientes naturales como sedimentos de cueva (Reynolds and Barton, 2014). No se conoce un caso equivalente de patógenos humanos asociados a cuevas, si bien animales que viven en cuevas pueden actuar como vectores para transmisión de enfermedades graves. Los murciélagos pueden ser foco de histoplasmosis o de rabia, sin olvidarnos de su posible papel en la evolución del coronavirus SARS-Cov2. La cueva de Altamira, libre de murciélagos y de roedores, no se puede considerar un problema sanitario desde este punto de vista.

Una cuestión diferente es que muchos de los microorganismos que se pueden encontrar en Altamira puedan ser patógenos oportunistas y causar infecciones en personas debilitadas. Sin embargo, se puede considerar que el riesgo microbiológico para la salud humana de la visita a Altamira sea equivalente o menor del que sufriríamos al entrar en cualquier otro recinto cerrado o si nos bañáramos en el mar.

Para hacer una recapitulación y sistematizar los riesgos de todo tipo que corre la cueva de Altamira y sus visitantes, durante el periodo de 2012-2014 se llevó a cabo un Programa de investigación para la conservación preventiva y régimen de acceso de la cueva de Altamira. Los objetivos de ese programa

fueron determinar las medidas de conservación preventiva necesarias, conocer la incidencia de la dinámica natural actual, determinar el impacto de la presencia humana y establecer las medidas preventivas y protocolos de gestión en relación al acceso de personas a la cueva. Entre las evaluaciones realizadas se determinó entre otros el riesgo biológico y en especial, el crecimiento de las colonias visibles como el mayor riesgo de biodeterioro para las pinturas de la cueva. Como resultado del mencionado proyecto se elaboró el Plan de Conservación Preventiva (PCP) vigente actualmente y que define los procedimientos de control y monitorización de los riesgos determinados, y las medidas a adoptar en caso de que los parámetros medidos indiquen la existencia de una situación de riesgo.

## **2.3 Microbiología sin cultivo**

Hace tiempo que se sabe que de todos los microorganismos presentes en una muestra solo unos pocos son capaces de formar colonias en una placa, por rico que sea el medio de cultivo y por mucho que se reproduzcan las condiciones ambientales óptimas. Este hecho, que en sus inicios fue llamado “Great Plate Count Anomaly” (Staley and Konopka, 1985), nos recuerda que solamente recuperamos en cultivos ordinarios enteros entre un uno y un 5% de la diversidad bacteriana, especialmente cuando se trata de muestras de origen ambiental. Esta limitación obliga al uso de técnicas específicas para tener una imagen real de una población microbiana. Ya he mencionado antes los métodos de alto rendimiento para optimizar la recuperación de microorganismos en cultivo.

La capacidad que hemos adquirido de secuenciar el DNA de los microbios, sin necesidad de cultivarlos de muestras ambientales en el laboratorio ha dado paso a una nueva especialidad de la ciencia, la metagenómica, que se ha convertido en la herramienta más útil para acceder a la elevada biodiversidad de las muestras ambientales, obviando las dificultades encontradas en el cultivo en el laboratorio. Dentro de este campo, hay una serie de técnicas cuya aparición han significado un importante avance en el esfuerzo de esclarecer la diversidad bacteriana.

### 2.3.1 Análisis de amplicones

Esta técnica consiste en la amplificación y secuenciación a partir de una población microbiana de una o varias regiones discretas del genoma de cada organismo presente en la muestra original. Del análisis de la secuencia de esta región obtendremos una imagen de la distribución de especies de la muestra inicial. Un buen “target” para el análisis de amplicones debe estar universalmente presente, y ser al mismo tiempo suficientemente conservado como para que sea posible su amplificación a partir de todos y cada uno de los individuos de una población y al mismo tiempo suficientemente diverso como para reflejar la diversidad de la población.

El gen universalmente adoptado para este análisis ha sido el del 16S rRNA. Las bacterias al igual que las arqueas presentan de 1 a 15 copias de este gen por genoma que mide alrededor de 1.600 pb. Pocos taxones bacterianos idénticos albergan copias idénticas del gen 16S rDNA, y el grado de heterogeneidad aumenta con el número creciente de copias (Ibal et al., 2019; Větrovský and Baldrian, 2013). En el campo de la microbiología sin cultivo, la propuesta de utilizar los genes ribosomales para el estudio de la filogenia y taxonomía microbiana fue rápidamente adoptada por la comunidad científica (Lane et al., 1985). Estos genes se encuentran presentes en todos los procariotas de una manera muy conservada y esto facilita el diseño de primers universales.

El gen 16S rDNA es el más utilizado porque contiene regiones altamente conservadas y a la vez regiones variables para identificar las características filogenéticas de los diferentes taxones (Figura 25). Además, al tratarse de un gen que puede ser multicopia se aumenta la sensibilidad de detección.

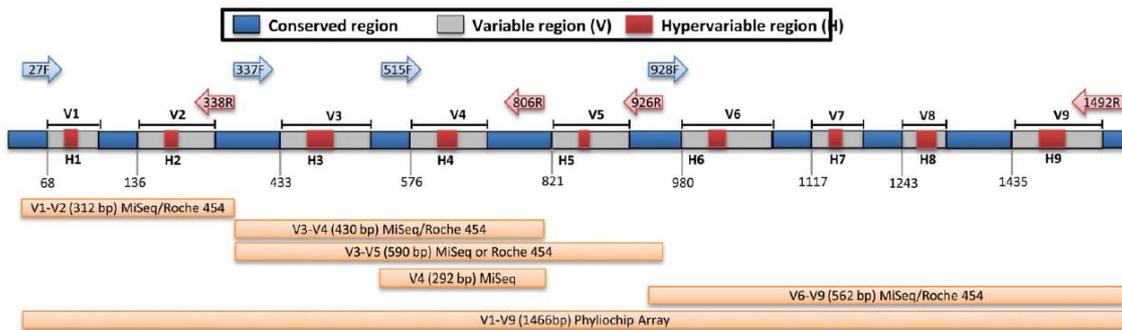


Figura 25. Representación de las diferentes regiones conservadas, variables e hipervariables que presenta el gen 16S rDNA. También se representa los primers universales más empleados para la amplificación de cada región y la plataforma de secuenciación que lo soporta. Figura obtenida de Shahi et al., 2017.

El gen 16S rDNA contiene 9 regiones variables (V1-V9) que presentan diferencias entre especies (Chakravorty et al., 2007). La elección de la región a analizar en el gen 16S rDNA puede afectar significativamente a las estimaciones de la diversidad taxonómica presente en la muestra. La inclusión de varias regiones de este gen en el mismo amplicon es la estrategia más empleada para obtener una mejor resolución. En el estudio de Chakravorty se analizaron las regiones variables de 110 especies bacterianas y determinaron que las regiones variables V2 y V3 son las que mayor número de polimorfismos (SNPs) presentan para la diferenciación entre especies bacterianas a nivel de género. Otro estudio más reciente compara las regiones V2-V3 con las regiones V3-V4, teniendo mayor resolución para taxones de menor rango (géneros y especies) estas primeras regiones (Bukin et al., 2019). Por el contrario, las regiones más conservadas son útiles para determinar los taxones de mayor rango.

Con el uso más frecuente de las plataformas de secuenciación NGS para el estudio del gen 16S rDNA, se ha logrado que cada vez haya un mayor perfeccionamiento y estandarización de la técnica, así como una reducción progresiva de los costos. Para aprovechar la potencia de secuenciación de estas plataformas fue decisivo el uso de etiquetas de secuencia o también conocido con el término “tags”, que se añaden a los amplicones en el proceso de elaboración de las librerías y que nos van a permitir secuenciar juntas varias

muestras que luego serán separadas por el software. En sus inicios, la técnica/plataforma dominante para la secuenciación de amplicones fue la pirosecuenciación en los aparatos de 454-Roche. Otras metodologías dominantes fueron la de la plataforma IonTorrent y la de Illumina que finalmente se hizo dominante en el campo. Con la tecnología de Illumina es posible secuenciar desde ambos extremos el fragmento amplificado del gen 16S rDNA.

Las lecturas crudas que se obtienen de los fragmentos amplificados que han sido secuenciados posteriormente se tratan bioinformáticamente a través de procesos (pipeline) que agrupan todos los pasos de tratado, filtrado y clusterizado de las lecturas (Figura 26). Los programas más utilizados que contienen este pipeline son Mothur, QIIME2 o DADA2. El primer paso en estos “pipeline” es un proceso de control de calidad y trimado de las lecturas, importante para filtrar lecturas cortas o de baja calidad y quimeras (artefactos producidos por la técnica de PCR que originan secuencias con distintas procedencias). El siguiente paso también importante es la agrupación de lecturas con una identidad del 97% en unidades operacionales taxonómicas (OTUs). Cada OTU teóricamente corresponde con una especie diferente y representa un cluster de secuencias casi idénticas. Una vez generados los clusters, se utiliza una secuencia representante de cada uno para compararla contra las bases de datos específicas de 16S rRNA y obtener su taxonomía. Para medir diferencias entre distintas comunidades biológicas, se estudia la diversidad presente en las muestras (la riqueza de especies y su distribución).

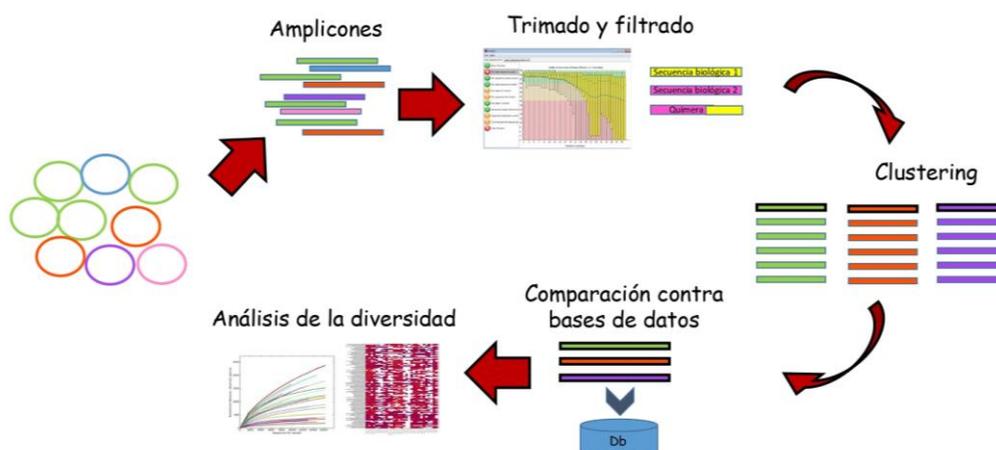


Figura 26. Pipeline del procesamiento de las lecturas de amplicones. Desde el procesado de las lecturas crudas hasta su agrupación en OTUs y estimación de la diversidad.

Esta técnica es de una innegable potencia y de uso casi universal, sin embargo, no está exenta de problemas que pueden conducir a resultados sesgados y a problemas en la correcta identificación de una población. Una limitación puede provenir del uso de primers que no son del todo universales y que pueden excluir algunos taxones importantes de ser amplificados y detectados. Por ejemplo, este método sería ineficaz para detectar un grupo de taxones conocido como “Candidate Phyla Radiation” (CPR) que se conoce que presentan intrones en estas regiones altamente conservadas del 16S rDNA, y que impedirían la amplificación o producirían amplicones de tamaño inesperado (Brown et al., 2015). Este grupo de bacterias se estima que puedan representar >15% de la diversidad bacteriana y consten alrededor de 70 filos (Geesink et al., 2020; Méheust et al., 2019). Los CPRs suelen ser bacterias ultrapequeñas que carecen de numerosas rutas biosintéticas o las tienen incompletas, y esto hace que se sospeche que muchos de ellos sean simbiontes obligados (Castelle and Banfield, 2018). Estas bacterias las encontramos en una variedad de entornos como cuevas, muestras de agua marina, acuíferos, aguas termales, areniscas, plantas de tratamiento de agua potable, permafrost o microbioma humano (Herrmann et al., 2019; Vigneron et al., 2020).

Además de desviaciones atribuibles a los primers utilizados y a su eficiencia de amplificación sobre diferentes dianas, puede ocurrir que la amplificación sea más rápida y eficaz sobre componentes más abundantes produciendo también alteraciones de los resultados cuantitativos o pérdidas de componentes muy poco abundantes.

No está exento de sesgos el procesamiento de las secuencias, y en concreto la agrupación de secuencias muy similares en clusters por identidad, puede ofrecer una idea de la diversidad presente en la muestra lejos de la realidad (Mysara et al., 2017). En este caso, se debería utilizar una identidad más

estricta (>97%) para clusterizar estas secuencias. Pero, por el contrario, subir demasiado el umbral de identidad por encima del 97% puede llevar a sobreestimar el número de OTUs presentes debido a que los errores de secuenciación introducidos por la técnica dan lugar a OTUs artificiales. Por lo tanto, estas regiones elegidas para la amplificación son cruciales y pueden tener un gran impacto sobre la diversidad microbiana observada resultando en la fusión de dos especies taxonómicas distintas o división en varios taxones de los miembros de una misma especie.

Todos estos factores de error deben tratar de evitarse con la mejor práctica de laboratorio posible, aunque es difícil de evitarlos por completo. El uso de controles internos de secuenciación obtenidos de analizar muestras de composición conocida es una práctica que sirve para verificar la fidelidad del proceso.

### **2.3.2 Análisis de DNA total**

La secuenciación total del DNA es otro método de estudio que consiste en secuenciar completamente el DNA obtenido de todos los organismos presentes en una muestra ambiental sin necesidad de cultivar. Es posible en teoría conseguir los genomas completos de todos y cada uno de los organismos presentes sean bacterias, arqueas, hongos, protozoos... Como ventaja frente a la anterior técnica, además de proporcionar una caracterización taxonómica de los genomas recuperados y una idea de la diversidad presente en la muestra, recoge toda la información genética permitiendo una anotación funcional metabólica de la comunidad y obteniendo un perfil de genes para cada genoma y para la comunidad.

En esta aproximación, llamada también secuenciación a ciegas o “Shotgun sequencing”, es predominante la secuenciación por la técnica de Illumina usando generalmente secuencias cortas de 150 pb obtenidas de los dos extremos de fragmentos de DNA que pueden tener tamaños medios que oscilan entre los 300 y los 500 bp. La capacidad de los nuevos secuenciadores ha hecho descender el precio de la secuenciación hasta los 10 céntimos de

dólar por megabase (<https://www.genome.gov/about-genomics/factsheets/DNA-Sequencing-Costs-Data>). La mayor facilidad y economía de esta técnica se compensa con una mayor complejidad y exigencia del análisis bioinformático tanto por el volumen de los datos a analizar como por las dificultades en los pasos individuales del procedimiento.

Tras un paso de control de calidad de las lecturas, en el que se eliminan zonas de baja calidad frecuentes en los extremos de las secuencias, se procede con el ensamblado de las lecturas (Figura 27). Consiste en conectar las lecturas individuales en secuencias continuas más largas denominadas contigs usando diferentes estrategias como la similitud entre las bases. Cuando se trata de ensamblar genomas que tienen referencias en las bases de datos, las lecturas se mapean alineándolas a dichas referencias con programas como por ejemplo Bowtie2 (Langmead y Salzberg, 2012) y a continuación se genera una nueva secuencia consenso. A falta de genomas de referencia el proceso es complejo y en su defecto, se utiliza sólo la información obtenida de la secuenciación para reconstruir el genoma en cuestión, sin conocimiento a priori de la organización del mismo (ensamblado *de novo*).

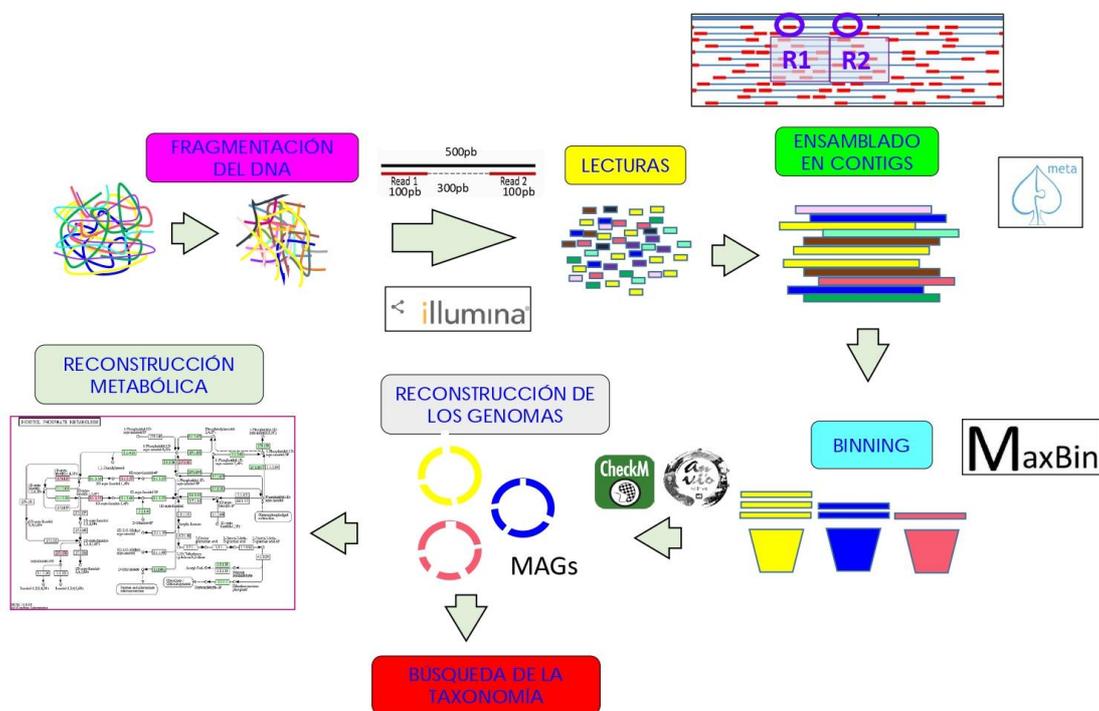


Figura 27. Pipeline de los pasos de procesamiento de las lecturas por la técnica de shotgun.

El empleo de lecturas cortas para ensamblar genomas nuevos se convierte en un proceso laborioso, que se complica en caso de existir numerosas repeticiones dentro del genoma, ya que las lecturas son tan cortas que no se sabe con exactitud de que copia provienen. A parte, cada ensamblador tiene su algoritmo para hacer el trabajo, siendo los más populares los basados en el método de gráficos de Bruijn utilizado por los ensambladores más usados (Velvet, SOAP de novo, Spades, etc). Para proyectos de metagenómica, metaSPAdes (Nurk et al., 2017) es el ensamblador más empleado. Aunque idealmente el producto de un ensamblador deberían ser secuencias con el genoma completo, en la práctica el resultado está bastante alejado de esto, sobre todo si estamos analizando muestras muy complejas. No todos los ensamblados son buenos y la calidad va a depender del coverage obtenido para cada base. Cuánto más coverage, mayor es el número de veces que se ha secuenciado esa base y más fácil será obtener un mejor ensamblado. Los criterios que se usan para estimar la calidad de un ensamblado derivan del tamaño y del número de los contigs obtenidos y de su relación con el tamaño del genoma o metagenoma que se está secuenciando. En un ensamblado de buena calidad los contigs mayores suelen ser al menos un 5% del tamaño del genoma y no necesitan más de 50-100 contigs para contener el 50% del tamaño de ese genoma (N50).

Cuando dentro de la muestra tenemos varios genomas entonces hay que separar los contigs pertenecientes a cada genoma en diferentes grupos (proceso llamado “binning”). Esta separación que es crucial, se basa en la premisa de que los contigs que provienen de un mismo genoma van a tener una composición en nucleótidos similar, y salvando el problema de las secuencias repetidas, todos los contigs deben tener la misma abundancia. De aquí que el “binning” se haga en función de la composición en nucleótidos de las secuencias (por ejemplo la frecuencia de tetranucleótidos o contenido en GC) y/o del coverage (profundidad de lectura) (Sedlar et al., 2017).

Por lo general, en una muestra nos encontramos con especies más abundantes y otras menos, con un coverage tan bajo que sus secuencias pueden acabar siendo clasificadas erróneamente usando únicamente esta

manera de binning. En estos casos, un binning híbrido basado en la composición de secuencias y su abundancia (coverage) como usa por ejemplo el programa Maxbin (Wu et al., 2016) es la manera más robusta para definir cada bin. De nuevo en condiciones ideales cada bin debería contener un conjunto de contigs que representan el genoma de uno de los microorganismos componentes de la mezcla analizada. A estos borradores de genoma obtenidos por este procedimiento los llamamos MAGs (Metagenome-Assembled Genomes).

Existe el peligro real de sacar conclusiones finales de estos borradores de genomas o MAGs debido a la frecuente asignación incorrecta de fragmentos del genoma de un organismo a otro diferente y otros errores similares. Por ello, es necesario un proceso de refinado individual, supervisado por el investigador, de cada MAG antes de proceder a su estudio. El curado de los MAGs consiste en estimar el nivel de contaminación y completitud presente para cada MAG. Una alta contaminación en el MAG puede producir conclusiones evolutivas y ecológicas engañosas, ya que se trataría de un genoma artificial. Para este proceso se utilizan programas como CheckM (Parks et al., 2015) o Anvi'o (Eren et al., 2015) que utilizan genes universales de copia única (SCGs) esperados en el genoma. Estos programas tienen seleccionado un conjunto de SCGs, que normalmente suelen tratarse de proteínas bien conservadas, basándose en su presencia en todos los genomas bacterianos o al menos en todos los genomas dentro de un grupo taxonómico. En un genoma sin contaminación, deberían estar presentes sin redundancia. Por otro lado, la gran dependencia de los SCGs para estimar la calidad de los MAGs no es suficiente para caracterizar el alcance de la contaminación. Contigs contaminantes de otros organismos que no tengan SCGs pueden pasar desapercibidos en la evaluación de la calidad del MAG (Chen et al., 2019).

Una vez que disponemos de MAGs fiables podemos proceder a asignar una taxonomía para cada uno de los MAGs. Cuando tratamos de genomas sin referencias, el enfoque más directo es buscar en los MAGs y extraer las secuencias de 16S rRNA y compararlas con las bases de datos específicas de 16S rRNA. Este enfoque es rápido pero presenta limitaciones, ya que las

secuencias de 16S rRNA son conocidas por ensamblar mal, a menudo en más de un contig y además no se separan en bins de manera eficaz por la similitud que existe entre los diferentes 16S rRNAs. Otros marcadores alternativos que también se utilizan para asignar la taxonomía incluyen los genes que codifican para la subunidad  $\beta$  de la RNA polimerasa (*rpoB*), la proteína RecA o la DNA girasa (*gyrB*).

Si disponemos de MAGs de alta calidad, la asignación taxonómica es posible por comparación del genoma completo con genomas relacionados (Figura 28). De la comparación de dos genomas se obtiene un valor ANI (Average Nucleotide Identity) que representa la identidad media de nucleótidos de todos los genes ortólogos compartidos entre dos genomas. Valores de ANI del 95-96% entre dos genomas indican que pertenecen a la misma especie (Konstantinidis and Tiedje, 2007; Richter and Rosselló-Móra, 2009). Otra estrategia que se puede implementar es usar la información a nivel de proteína. Por comparación de proteínas de dos genomas se obtiene el valor AAI (Average Aminoacid Identity) que representa lo mismo que el valor ANI, pero a nivel de proteína. También encontramos programas como CAT/BAT (von Meijenfeldt et al., 2019) que predicen los ORFs que codifican para proteínas dentro de los contigs binned en MAGs o aún sin binear, y los compara contra la base de datos de proteínas non-redundant del NCBI. El algoritmo del programa saca finalmente la clasificación más conservadora al nivel en el que puede asegurar la taxonomía más correcta en base a la puntuación que se obtiene para cada ORF. Estos valores basados en comparaciones del genoma complementados con la filogenia obtenida por el 16S rRNA son las estrategias más completas para obtener una taxonomía mucho más robusta en genomas de bacterias sin representantes en las bases de datos. Usando estos enfoques hay servidores web como “Microbial Genomes Atlas” (MiGA, Rodriguez-R et al., 2018b) que permite la clasificación de genomas aún no caracterizados, completos o parciales, contra los genomas presentes en las bases de datos (como por ejemplo NCBI\_prok) o colecciones de genomas de proyectos de genómica ambiental (como por ejemplo TARA Oceans).

El último paso en la caracterización de los metagenomas consistiría en una anotación funcional, que puede realizarse individualmente a cada MAG o en conjunto a toda la comunidad, proporcionando información sobre el potencial metabólico. Este paso es desafiante, ya que a veces las proteínas anotadas pueden tener una baja similitud contra las proteínas de referencia validadas experimentalmente. Cuando se trata de anotar un genoma lo más acertado es una anotación detallada gen por gen contra bases de datos que contienen proteínas manualmente curadas como Swiss-Prot (Uniprot) o sólo anotadas computacionalmente como TrEMBL (Uniprot) o RefSeq (NCBI) o una combinación de ambas. Para grandes sets de secuencias, hay anotadores rápidos como MG-RAST (Meyer et al., 2008), eggNOG-mapper (Huerta-Cepas et al., 2017) o MetaErg (Dong and Strous, 2019) que producen a parte de la función y nombre de la proteína otros identificadores como categorías funcionales de grupos de ortólogos (COGs), dominios de proteína, KEGG Orthology (KO), Gene Ontology (GO) o Enzyme Commission (EC) entre otros. Estos identificadores se pueden cargar en el servidor de KEGG (Kanehisa and Goto, 2000) para explorar y visualizar las rutas metabólicas que están presentes en estos genomas.

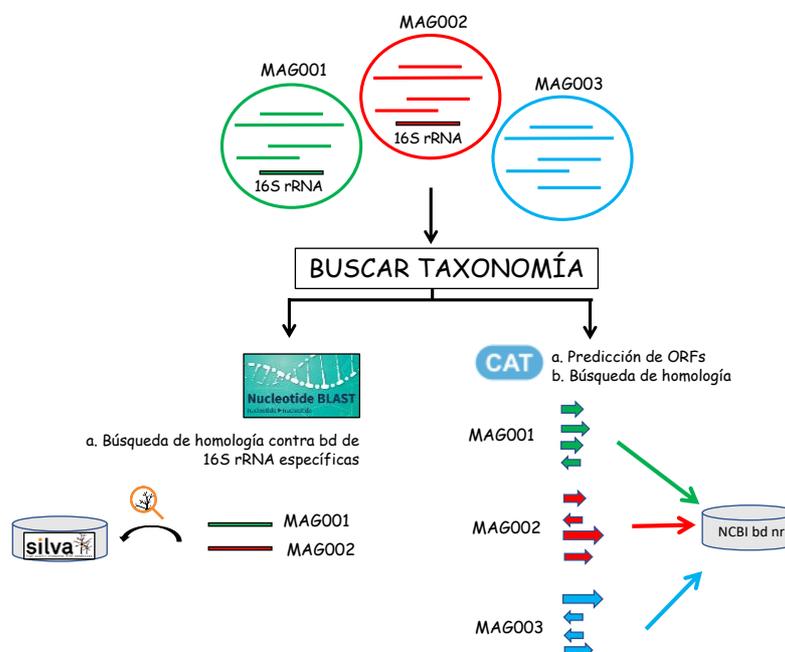


Figura 28. Enfoques empleados en la asignación taxonómica de los borradores de genomas (MAGs) obtenidos en este estudio metagenómico.

### 2.3.3 Amplificación de genomas de células únicas.

La tecnología de célula única o “single-cell” ha representado un avance frente a las anteriores técnicas debido a la capacidad de poder secuenciar genomas de una célula individual separada de una población. El análisis de una población exigiría el análisis individual de todos sus componentes. Aunque puede tener limitaciones en análisis de poblaciones complejas, es suficiente para el análisis de genomas individuales. Al borrador de genoma obtenido por esta técnica lo llamamos SAG (Single cell Assembled Genome), va a ser superior a un MAG en cuanto que en principio elimina el riesgo de contaminación con material de otros genomas. Además, esta técnica permite analizar diversidad intraespecífica. Los métodos de secuenciación tradicionales solo pueden obtener el promedio de muchas células para la recuperación de los genomas, siendo incapaces de detectar heterogeneidad entre células individuales.

El análisis genómico por la técnica de single-cell requiere en primer lugar separar las células presentes en la muestra por métodos de separación como por ejemplo usando un citómetro y un separador de células (fluorescent-activated cell sorting, FACS) o por dilución de las muestras y tratamiento microfluídico. Las células se lisan y se amplifica el DNA, ya que el rendimiento de DNA de una única célula es muy bajo (pocos picogramos o nanogramos). Este paso de amplificación se hace por PCR o por amplificación por desplazamiento múltiple (MDA) usando polimerasas altamente procesivas como la del bacteriófago  $\phi$ 29.

De los métodos que usan PCR el más usado es conocido como “Degenerate Oligonucleotide–Primed Polymerase Chain Reaction” (DOP-PCR) en el que se utilizan primers degenerados que presentan una secuencia de seis bases aleatoria en el extremo 3' y una secuencia fija en el extremo 5'. En la primera amplificación se utilizan condiciones poco rigurosas de baja temperatura en los pasos de hibridación y extensión, que permiten la unión de los primers por el extremo 3' en varios sitios en todo el genoma. En la siguiente amplificación las condiciones son más rigurosas y se eleva la temperatura permitiendo una unión más específica de los primers por el extremo 5' a los amplicones. Una

limitación de la técnica es que la polimerasa impone al tamaño de los fragmentos amplificados que generalmente no puede ir más allá de los 5 kb.

La otra técnica más utilizada se conoce como “Isothermal Multiple Displacement Amplification” (IMDA) y se trata de una amplificación por desplazamiento de hebras múltiples. Se utiliza la DNA polimerasa del bacteriófago  $\phi$ 29 de alta fidelidad y procesividad que amplifica el DNA en una reacción isotérmica con la ayuda de primers hexámeros de bases aleatorias que van a hibridar en diversas zonas del genoma. Cuando la polimerasa encuentra una hebra sintetizada previamente, desplaza la hebra para continuar la polimerización y deja un nuevo template de unión de primers para ser amplificado.

Las técnicas de secuenciación de célula única tampoco están exentas de problemas y los artefactos son frecuentes por ser métodos basados en amplificación. Estos artefactos pueden venir por ejemplo por errores de amplificación cuando el DNA que se amplifica tiene alto contenido en GC o se encuentra en bajas concentraciones o se trata de fragmentos pequeños o degradados, y también cuando hay pérdida del coverage del genoma, mutaciones y/o quimeras producidos (Gawad et al., 2016). Estos errores por tanto pueden acabar comprometiendo el análisis cuantitativo de los metagenomas.



### 3. Hipótesis y objetivos



### 3.1 Hipótesis de trabajo

El arte rupestre representa una parte de nuestra historia, su valor histórico es un bien que prima conservar. La mayoría de estas representaciones que, generalmente se encuentran en el interior de cuevas, continúa en perfecto estado gracias a la protección del exterior que ofrecen estos ambientes. Sin embargo, la presencia humana, así como procesos naturales pueden traducirse en un impacto en la preservación de las pinturas.

En la cueva de Altamira, el 47% de las pinturas que encontramos en su interior han llegado hasta el día de hoy en buen estado. La causa principal de pérdida de las pinturas en sus 30.000 años de existencia se identifica con sucesos hidrogeológicos, derrumbes y arrastre de soporte y pigmentos. Los micro (y macro) organismos que han poblado siempre las cuevas no parecen haber causado un gran daño a las mismas. Esto puede ser debido a la estabilidad de las condiciones ambientales y nutricionales que presenta en su interior a lo largo del tiempo. Aun así, desde su apertura al público se han observado indicios de biodeterioro que representan un riesgo para el estado y conservación de las pinturas. Algunos de estos indicios han correspondido con apariciones de microorganismos fototróficos a consecuencia del uso de luz artificial, que en el momento que se retiraron los puntos de luz, desaparecieron. Sin embargo, el riesgo biológico más alarmante que encontramos en la cueva a día de hoy es el desarrollo de colonias visibles que con los años progresan hacia el interior. Estas colonias representan una mezcla de diferentes componentes procarióticos cuyo origen es de cueva, ya que se detectan en otras cuevas del mundo. Además, no se detectaron antes de 1999 y actualmente la entrada (zona Vestíbulo) y la sala contigua (Cruce) son las salas que se encuentran más afectadas, encontrándose paredes y techos muy colonizados. Ejemplos en cuevas turísticas que han sufrido previamente colonizaciones de microorganismos, no han tenido un final deseado como ocurrió en la cueva de Lascaux.

La experiencia de Lascaux nos dice con claridad que cualquier intervención en una cueva debe ser cuidadosamente analizada antes de su ejecución para

evitar efectos colaterales indeseados. Este análisis no es posible sin un conocimiento detallado de la diversidad microbiana que interviene y de sus condiciones nutricionales.

Tras los primeros indicios de deterioro en la cueva de Altamira, se planteó un programa de investigación para la conservación y régimen de acceso a la cueva de Altamira cuyo objetivo principal fue establecer e identificar indicadores de deterioro para la cueva. De este estudio se confirmó que el riesgo biológico más inminente para las pinturas es el desarrollo y progreso de las colonias visibles. Como consecuencia se planteó la realización de esta tesis enfocada en el conocimiento de estos componentes que dan lugar a las colonias visibles.

Desde el punto de vista de su diseño, esta tesis constituye un trabajo esencialmente descriptivo que se suma a otros realizados anteriormente con el mismo objetivo. Lo que justifica este nuevo trabajo es que se va a aplicar una metodología, la de la microbiología sin cultivo o metagenómica que nunca antes han sido aplicadas y que puede dar una imagen mucho más completa y detallada de la diversidad microbiana de la cueva, llegando a proporcionar borradores de alta calidad de los genomas de los componentes más abundantes con posibles aportaciones taxonómicas de relevancia.

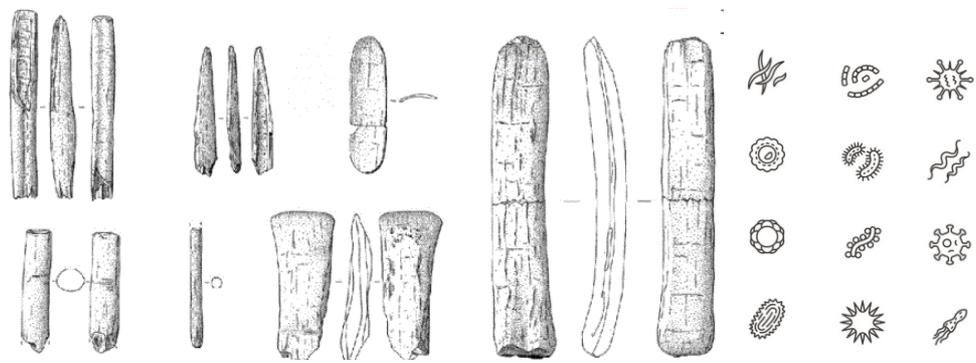
La nueva imagen que esta tesis va a proporcionar y a su vez plantear objetivos correlacionales o ecológicos que podrían identificar aquellos factores de los que depende la proliferación microbiana y que serían las causas directas del biodeterioro. Dicho de otra manera, nos puede proporcionar herramientas en el control del biodeterioro.

Otra avenida que abre esta tesis es la revisión de los métodos vigentes que utilizamos para medir el riesgo de biodeterioro. Al conocer la secuencia de los genomas de los microorganismos clave, podemos diseñar técnicas que permitirán su detección rápida y temprana, lo que supone una mejora sustancial sobre los métodos que ahora son habituales.

### **3.2 Objetivos del estudio de trabajo**

- Descripción detallada de los componentes de las colonias visibles mediante la complementación de diferentes técnicas independientes de cultivo (técnica de amplicones y shotgun) no empleadas hasta ahora en Altamira.
- Desarrollo de nuevos métodos de control enfocados en la detección temprana de los componentes mayoritarios de las colonias antes de su futuro desarrollo en las paredes y techos de la cueva de Altamira.
- Intento de cultivo de los componentes de las colonias detectados por las técnicas moleculares independientes de cultivo empleadas en esta tesis en medios de crecimiento ricos y pobres.





## 4. Materiales y métodos



## 4.1 Toma de muestras

El acceso a la cueva de Altamira y la toma de muestras necesaria para este trabajo se ha hecho con el permiso de la Comisión de seguimiento del PCP y en el contexto del contrato “SERVICIO DE CONTROL Y SEGUIMIENTO DE LAS CONDICIONES AMBIENTALES Y DEL BIODETERIORO DE LA CUEVA DE ALTAMIRA” suscrito entre la Universidad de Cantabria y el Ministerio de Cultura y Deporte. El trabajo no ha supuesto entradas a la cueva adicionales a las previstas para el trabajo de control, y las muestras, que en ningún caso han afectado pinturas, se han hecho dejando el mínimo rastro visible posible siempre bajo la supervisión de personal del Museo y Centro de Investigación Altamira.

El total de las muestras utilizadas en este estudio han sido 55 muestras de colonias: 8 beis, 18 grises, 14 blancas, 15 amarillas, y 2 muestras de pared sin colonia visible aparente que se han utilizado como control (Anexo-Tabla S1). Todas las muestras han sido analizadas por la técnica de amplicones y entre ellas se han elegido 6 muestras para secuenciar por la técnica de DNA total para obtener al menos una muestra para cada subtipo de colonia siendo Y6M1 (YC), W6M1 (W1), W8M2 (W2), G10M1 (G1), G10M2 (G2) y B11M1 (BC) las colonias elegidas, a las cuales se han referido con el tipo de colonia que se representa entre paréntesis en el tratamiento de los resultados de DNA total. Las muestras de colonias se han tomado en 3 zonas de la cueva: Vestíbulo, Cruce y Pasillo de Polícromos (Figura 3).

También se han recogido muestras de suelo y soporte para poder comparar las comunidades presentes en las colonias frente a las presentes en distintos puntos de suelo y soporte de la cueva (Anexo-Tabla S2). Estas muestras han sido un total de 86 muestras y se han tomado en las siguientes zonas: Vestíbulo, Polícromos, Hoya Bisonte y Hoya Cabra (Figura 3). Las muestras de soporte, a diferencia de las muestras de control que también son de soporte, se recogen de manera diferente y de zonas distintas de dónde se encuentran desarrollándose las colonias visibles.

La toma de las muestras de colonias y control se ha realizado poniendo especial cuidado en arrastrar la menor cantidad de soporte posible del techo o de la pared. Para ello con un bisturí estéril se ha raspado la colonia y transferido a un tubo estéril. La toma de las muestras de suelo y soporte ha sido diferente, estas se han recogido con espátulas estériles en tubos de 50 mL en distintos puntos de suelo y pared. Las muestras recogidas para su transporte al laboratorio se han almacenado en un enfriador portátil dentro de una bolsa térmica. El tiempo de vuelta al laboratorio fue alrededor de 2 horas y una vez en el laboratorio se almacenaron a  $-80^{\circ}\text{C}$  hasta su uso.

### **4.2 Amplificación por PCR y secuenciación del gen 16S rRNA**

El DNA se extrajo con el kit PowerSoil DNA isolation kit, Quiagen (Gaithersburg, DC) de acuerdo con las instrucciones del proveedor. La integridad, el tamaño y la concentración de los DNAs siempre se han evaluado con las ScreenTape D1000 en una TapeStation 4200 (Agilent). Las concentraciones de DNA obtenidas para las muestras de colonias oscilaron entre 1,5-40 ng/ $\mu\text{l}$  y para las muestras de control 0,05 y 0,132 ng/ $\mu\text{l}$ . Las concentraciones obtenidas de las muestras de suelo y soporte siempre han producido rendimientos del DNA más bajos 0,02-8 ng/ $\mu\text{l}$ .

Los pasos de elaboración de librerías de amplicones se realizaron acorde al protocolo 16S Metagenomic Sequencing Library Preparation (Illumina). El protocolo consiste en la realización de 2 reacciones de PCR. La PCR1 (25  $\mu\text{l}$  de vol final) se llevó a cabo con:

- 12,5  $\mu\text{l}$  Kappa2G Robust Hot Start Ready Mix (2X) (Kappa Biosystems, Boston, MA)
- 5  $\mu\text{l}$  de cada primer forward y reverse (1  $\mu\text{M}$  cada uno)
- 2,5  $\mu\text{l}$  de muestra de DNA (5 ng/ $\mu\text{l}$ )

Los primers utilizados en esta PCR1 fueron primers universales que permiten la amplificación de procariotas (bacterias y arqueas) incluyendo alguna base

degenerada para asegurar la mejor variabilidad posible. El producto amplificado corresponde con la región V3-V4 del gen 16S rRNA. Los primers utilizados han sido Pro341F-5'CCTACGGGNBGCASCAG3' y Pro805R-5'GACTACNVGGGTATCTAATCC3'. A parte, estos primers tienen una secuencia adicional linker para que sea reconocida por los primers de secuenciación de Illumina en la PCR2 (secuencia subrayada). Los primers de secuenciación de Illumina, además de contener estas secuencias linker, contienen índices duales para diferenciar muestras diferentes y unas secuencias adaptadoras que hibridarán a la plataforma de secuenciación de Illumina, y que serán añadidos al producto obtenido en la PCR1 en la segunda PCR. Las secuencias linker también servirán para la unión de los primers en el proceso de generación de los clusters durante la reacción de secuenciación.

- Pro341F- 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[V3-V4 region]
- Pro805R- 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[V3-V4 region]

La reacción de la PCR1 se hizo en un termociclador 2720 Thermal Cycler (Applied Biosystems, California, Estados Unidos) y las condiciones fueron:

- un ciclo de 95°C durante 3 min
- 25 ciclos de
  - 95°C durante 30 seg
  - 55°C durante 30 seg
  - 72°C durante 30 seg
- un ciclo final de extensión de 72°C durante 5 min.

Se realizó una reacción por muestra y se obtuvo un único amplicon de aproximadamente 450 pb. Tras la reacción se usaron bolas AMPure XP (Beckman Coulter, Brea CA) para purificar los productos de PCR de primers libres y de dímeros de primers.

La PCR2 adjuntó índices duales y adaptadores de secuenciación de la plataforma de Illumina a cada amplicon usando el Kit de índices Nextera XT (Illumina San Diego, CA). La reacción de PCR (50 µl vol final) se realizó con:

- 25  $\mu$ l Kappa2G Robust Hot Start Ready Mix (2X) (KappaBiosystems, Boston, MA)
- 5  $\mu$ l de cada Nextera XT Index: Primer 1 (N7xx) y Primer 2 (S5xx)
- 5  $\mu$ l de de cada amplicon purificado
- 10  $\mu$ l de agua MiliQ

La reacción de la PCR2 se hizo en un termociclador 2720 Thermal Cycler (Applied Biosystems, California, Estados Unidos) y las condiciones fueron:

- un ciclo de 95°C durante 3 min
- 8 ciclos de
  - 95°C durante 30 seg
  - 55°C durante 30 seg
  - 72°C durante 30 seg
- un ciclo final de extensión de 72°C durante 5 min.

Se realizó de nuevo un segundo paso de limpieza adicional de primers libres y/o dímeros de primer con las bolas AMPure XP (Beckman Coulter, Brea CA).

La validación de las librerías se evaluó de nuevo en ScreenTape D1000 en una TapeStation 4200 (Agilent). El tamaño medio de amplicones de la librería final fue ~630 bp. La cuantificación de los amplicones finales fue de 1 a 60 ng/ $\mu$ l dependiendo de las muestras, obteniendo los rendimientos más bajos para los amplicones de suelo y soporte. Todos los amplicones se diluyeron utilizando Illumina Resuspension Buffer (RSB) y se combinaron en concentraciones equimolares en el mismo grupo para secuenciar. Las muestras se enviaron al Centro Nacional de Análisis Genómico (CNAG-CRG) donde se secuenciaron con la tecnología MiSeq Illumina (PE 300X2 de cada molécula).

### **4.3 Procesamiento bioinformático de los datos de amplicones**

Las lecturas de amplicones se procesaron de acorde al pipeline de QIIME2 v2018.4 (Bolyen et al., 2018). La primera parte consistió en la depuración de las lecturas para su asignación en clusters. Las secuencias que pasaron los filtros de calidad fueron trimadas las primeras 9 bases y 5 las últimas del read1,

y las primeras 9 bases y 30 últimas del read2. La detección de quimeras se hizo con el programa DADA2 (Callahan et al., 2016) con el comando *denoise-paired*. Se hizo un segundo paso de detección de quimeras usando el algoritmo de UCHIME (Edgar et al., 2011) con el comando de *uchime-denovo*.

En la segunda parte, las secuencias se agruparon en clusters con el comando *cluster-features-de-novo* usando el algoritmo de VSEARCH (Rognes et al., 2016). El grado de identidad entre las secuencias para que formen parte del mismo cluster se estableció al 97%. Cada uno de estos clusters constituye una OTU (Operational Taxonomic Unit). La asignación taxonómica de las OTUs se realizó con el comando *feature-classifier* usando como referencia la versión 132 de la base de datos SILVA (<https://www.arb-silva.de/documentation/release-132/>).

Por último, los análisis de biodiversidad se obtuvieron con los pasos finales del pipeline de QIIME2. Las curvas de rarefacción y los índices de diversidad alfa se obtuvieron con el comando *diversity alpha*. La diversidad beta en las muestras se sacó con la función *vegdist* de la librería *vegan* v2.4-2 usando el método de disimilitud Bray-Curtis en RStudio.

#### **4.4 Preparación de librerías y secuenciación del DNA total**

El DNA total de las muestras de metagenomas se extrajo con el kit PowerSoil DNA isolation kit, Quiagen (Gaithersburg, DC). La integridad, el tamaño y la concentración de los DNAs se han evaluado con las ScreenTape D1000 en una TapeStation 4200 (Agilent). Las concentraciones de DNA obtenidas para las muestras seleccionadas de metagenomas fueron de 20-40 ng/µl. Los DNAs (2 µg) se enviaron para hacer las librerías a MacroGen (Madrid, España). La elaboración de librerías se realizó acorde al protocolo TruSeq® Nano DNA Library Prep (Illumina).

Resumidamente, se fragmenta el DNA genómico en trozos de 550 bp en un sonicador. Se purifican los fragmentos sonicados con bolas AMPure XP

(Beckman Coulter, Brea CA). Se reparan los extremos resultantes de la fragmentación para obtener extremos romos con la enzima polimerasa ERP (End Repair Mix). Para seleccionar el tamaño deseado de la librería (550 pb), se remueven los fragmentos de mayor tamaño (>550pb), y después de menor tamaño (<550 pb) con una serie de pasos y tiempos de lavado con las bolas AMPure XP (Beckman Coulter, Brea CA). Después, los extremos 3' de los fragmentos seleccionados resultantes se adenilan con la enzima ATL (A-Tailing Mix) para evitar que se religuen durante la reacción de ligación de los adaptadores de secuenciación de Illumina. En el siguiente paso se ligan los índices adaptadores de Illumina, los cuales presentan un nucleótido T en sus extremos 3' complementario a la base A incorporada en los fragmentos en el paso anterior. Finalmente, se purifican los productos ligados con las bolas AMPure XP (Beckman Coulter, Brea CA) y se realiza una última PCR para enriquecer los fragmentos ligados purificados con los reactivos PPC (PCR Primer Cocktail) y EPM (Enhanced PCR Mix). Los fragmentos que tengan ligados adaptadores en ambos extremos se van a amplificar. Aquellos fragmentos que tenga uno o ninguno son ineficiencias de la reacción de ligación y no van a generar clusters posteriormente. Finalmente, se purifican los fragmentos enriquecidos con las bolas AMPure XP (Beckman Coulter, Brea CA).

Las muestras de metagenomas en este estudio se han secuenciado en la plataforma HiSeq de Illumina. Los primeros metagenomas secuenciados de las muestras Y6M1 (YC), W6M1 (W1) y W8M1 (W2) en una carrera PE 2x100nt de cada molécula mientras que para los metagenomas de las muestras G10M1 (G1), G10M2 (G2) y B11M1 (BC) se secuenciaron PE 2x150nt de cada molécula.

#### **4.5 Trimado, ensamblado y bineado de las lecturas de los metagenomas; y análisis de la diversidad**

La calidad de las lecturas crudas de los metagenomas se evaluó usando el software FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

En los primeros metagenomas Y6M1 (YC), W6M1 (W1) y W8M1 (W2) que fueron obtenidos con lecturas de 100 bp, de acuerdo con los indicios de calidad, se trimaron las 10 últimas bases en la muestra Y6M1 (YC) y W6M1 (W1), mientras que para los metagenomas G10M1 (G1), G10M2 (G2) y B11M1 (BC), con reads de 150 pb, se trimó únicamente la muestra G10M2 (G2) en las 3 últimas bases. El programa utilizado para trimar fue FASTX-Toolkit 0.0.14 con el comando *fastx\_trimmer* ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Las secuencias trimadas fueron ensambladas después con el programa meta-SPAdes v3.11.1 (Nurk et al., 2017). Este programa incluye un paso previo de corrección de errores para eliminar lecturas con k-mers raros. A continuación, se realiza el paso de ensamblado usando para ello valores óptimos de *k* que fueron determinados empíricamente previamente para cada tipo de muestra. Se utilizó *k*= 33, 57 y 97 para lecturas de 2 x 150 bases, y *k*= 33, 57 y 81 para lecturas de 2 x 100 bases.

Los contigs obtenidos menores de 1000 pb fueron descartados, y el resto se utilizó por el software Maxbin v2.2.4 para la obtención de MAGs usando las opciones por defecto (Wu et al., 2016). La contaminación, la integridad y la heterogeneidad de cada MAG se evaluaron con el programa checkM v1.0.7 (Parks et al., 2015). Para editar y/o curar los borradores de genomas obtenidos de la contaminación presente se utilizó el programa Anvi'o v6.1 (Eren et al., 2015).

Las curvas de rarefacción en las muestras de metagenomas se hicieron con el programa Nonpareil v3.303 (Rodríguez-R et al., 2018b). Los análisis de  $\beta$ -biodiversidad se hicieron con el programa MetaFast v0.1.4 (Ulyantsev et al., 2016). El plot MDS de la diversidad beta en las muestras de metagenomas representó las distancias obtenidas con el programa MetaFast con la función *vegdist* de la librería *vegan* v2.4-2 usando el método de disimilitud Bray-Curtis en RStudio.

## 4.6 Asignación taxonómica de los MAGs y comparaciones de secuencia de DNA y proteína de los componentes de las colonias

La asignación taxonómica de los MAGs se realizó por una aproximación múltiple basada en el análisis de las comparaciones de las secuencias parciales/enteras del 16S rRNA contra bases de datos específicas, comparaciones de la composición del DNA (valor ANI) contra genomas cercanos y comparaciones de las secuencias codificantes contra las bases de datos de proteínas.

Las secuencias de 16S rRNA presentes en los metagenomas se extrajeron con el programa checkM v1.0.7 utilizando el comando *ssu\_finder* (Parks et al., 2015). Estas secuencias parciales/enteras del gen 16S rRNA se compararon contra la base de datos SILVA\_132 con el programa BLAST v2.10.0 con el comando *blastn* (Altschul et al., 1990). A parte, para cada MAG se anotaron los ORFs presentes en los contigs con el programa CAT/BAT (von Meijenfeldt et al., 2019). Este programa depende de Prodigal v2.6.3 (Hyatt et al., 2010) para predecir los ORFs y de Diamond v0.9.24.125 (Buchfink et al., 2015) para compararlos contra la base de datos non-redundant de proteínas del NCBI.

Las comparaciones a nivel de nucleótido entre los componentes y los genomas de referencia para la obtención de clusters presentes en los MAGs se hicieron con el programa OrthoANI Tool v0.93.1 (Lee et al., 2016). Se estableció un valor de ANI >77% para agrupar dentro del cluster y, por tanto, pertenecer al mismo género. El número de acceso a GenBank de los genomas de referencia utilizados en estos análisis se recoge en el Anexo-Tabla S7. Las comparaciones a nivel de proteína entre los componentes mayoritarios se obtuvieron con el program compareM con el comando *aai\_wf* (<https://github.com/dparks1134/CompareM>).

Para las comparaciones de genomas a nivel de nucleótido de los componentes mayoritarios contra los genomas de referencia más cercanos se utilizó otra estrategia. En el caso de los componentes del cluster III clasificados en la

familia *Pseudonocardiaceae* se seleccionó al menos un genoma de referencia para cada género descrito y lo mismo en el caso de los componentes del cluster IV clasificados en la clase *Nitriliruptoria*. Después las comparaciones de estos genomas seleccionados contra los de los componentes mayoritarios se hicieron calculando las distancias basadas en estimaciones probabilísticas con el programa Mash v1.1 (Ondov et al., 2016) y se visualizaron en un heatmap con la librería “gplots” en RStudio. El número de acceso a GenBank de los genomas de referencia utilizados en este último análisis se recoge en el Anexo-Tabla S8.

#### **4.7 Abundancia relativa de los MAGs.**

La abundancia relativa de cada MAG en los metagenomas se estimó en base a la proporción de lecturas que alineaban con cada uno de ellos, teniendo en cuenta la proporción de lecturas mapeadas en contigs >1000 que no están bineados en MAGs.

Las lecturas corregidas por el ensamblador se alinearon con el programa Bowtie v2.2.6 usando el modo *end-to-end* que contempla el alineamiento del read completo (Langmead y Salzberg, 2012). Se alinearon estas lecturas contra los fasta de los MAGs a las que se añadieron los contigs >1000 bp no bineados en MAGs. El número de lecturas alineadas se obtuvo con el programa Samtools v1.9 (Li et al., 2009). Los archivos BAM se pasaron a formato SAM con el comando *view*, se ordenaron con el comando *sort* y se indexaron con el comando *index*. Por último, se obtuvieron los reads alineados de cada archivo BAM con el comando *idxstats* para cada MAG correspondiente.

Finalmente, la abundancia relativa de los MAGs en cada metagenoma se calculó de la siguiente manera, considerando como lecturas mapeadas totales la suma de las mapeadas en los MAGs y en contigs >1000 no bineados:

$$A. R_{MAG} = \frac{\text{Lecturas mapeadas}_{MAG}}{\text{Lecturas mapeadas totales}} \times \text{Tamaño genoma}_{Mb} \times 100$$

#### **4.8 Construcción de árboles filogenéticos basados en secuencias de 16S rRNA de los componentes mayoritarios**

Los alineamientos y los árboles filogenéticos sólo se construyeron para los 16S rRNA de componentes mayoritarios de las colonias con el software ARB v6.0.2 (Ludwig et al., 2004) junto con las secuencias de referencia más cercanas a estos componentes.

Las secuencias de 16S rRNA completas de cada componente mayoritario fueron alineadas con el programa SINA v1.6.0 (Pruesse et al., 2012) y añadidas a la base de datos de referencia de SILVA\_132. A continuación, se seleccionaron secuencias de especies representantes de cada género descrito en el caso de los componentes mayoritarios que agruparon en el cluster III (un total de 340 secuencias) y lo mismo en el caso de los componentes agrupados en el cluster IV (un total de 140 secuencias) para la reconstrucción *de novo* de los árboles de los componentes mayoritarios de cada cluster. Estos árboles se obtuvieron con el método de máxima verosimilitud usando el algoritmo de PHYML versión 20130708 (Guindon and Gascuel, 2003) y los soportes de ramificación fueron determinados por estimación bayesiana. La barra de escala utilizada indica 0,1 cambios por nucleótido.

#### **4.9 Diseño de ensayos para la detección de componentes mayoritarios de las colonias mediante el diseño de primers específicos**

Los targets seleccionados para el diseño de primers específicos de los componentes mayoritarios cumplieron la condición que sólo estuvieran presentes en los genomas de los componentes mayoritarios contra los que están diseñados.

Las proteínas de cada componente mayoritario se anotaron con Prodigal v2.6.3. A continuación, se hicieron bases de datos de las proteínas de cada componente con el programa Diamond v0.9.24.125 usando el comando *makeadb*, y después con el comando *blastp*, cada base de datos fue comparada con las proteínas de los restantes para buscar aquellos genes únicos no compartidos en cada componente. Este proceso se ha repetido en todos los componentes y se ha llegado a la selección de un gen en particular único para cada componente (Tabla 5). La selección final de targets únicos fue aleatoria en todos los componentes, a excepción de la región utilizada en el componente mayoritario de la colonia W2, que fue diseñada en intrones que presenta esta bacteria en su gen del 23S rRNA (Figura 64).

A continuación, los primers de los targets únicos seleccionados se diseñaron con el software Primer3 (<https://bioinfo.ut.ee/primer3-0.4.0/>) y se comprobaron con el programa BLAST v2.10.0 a través del comando *blastn*, que los primers finales sólo dieran hit con contigs de los componentes mayoritarios contra los que estaban diseñados. Para ello además se hizo una base de datos con todos los contigs que se obtuvieron del ensamblado individual de los reads de cada metagenoma con el programa BLAST v2.10.0 usando el comando *makeblastdb*. Contra esta base de datos se compararon estos posibles primers y se seleccionaron finalmente un par de primers para cada componente.

Los tamaños de los productos de amplificación de cada par de primers se recogen en la Tabla 5. La secuencia de los primers se recoge en el Anexo-Tabla S9.

#### **4.10 Análisis funcional de los metagenomas**

Para el análisis funcional de las colonias se predijo en primer lugar la capacidad para codificar proteínas de los metagenomas y se anotó su función. Dentro de cada metagenoma para el catálogo de genes obtenido se consideraron tanto los ORFs detectados en contigs bineados en MAGs como contigs >1000 que no pudieron ser bineados.

Todos los ORFs predichos inicialmente con el programa Prodigal v2.6.3 fueron anotados funcionalmente con eggNOG-mapper v2.0.0 (Jensen et al., 2008). Las secuencias predichas de las proteínas se compararon contra la base de datos eggNOG 5.0 (Huerta-Cepas et al., 2019). Para los ORFs que se asignaron en grupos de ortólogos, el programa proporciona la clase funcional de cada ORF que son las que se han utilizado para la realización de la gráfica final (Figura 66) en categorías funcionales y propone una función particular para cada ORF en función al grupo de ortología al que se asigna.

En la asignación funcional completa de los metagenomas se investigó la presencia de genes descritos como marcadores de diferentes rutas de metabolismo autotrófico (Tabla7) como fijación de CO<sub>2</sub> (*RuBisCO*) o de nitrógeno (*nifDKH*), o enzimas con función oxidadora de moléculas inorgánicas como amonio (*pamoA/amoA*), monóxido de carbono (*coxLMS*), formato (*fdoGHI*) e hidrógeno (*hyaAB*).

#### **4.11 Condiciones de crecimiento de los aislados de colonias visibles**

Los intentos de cultivo se realizaron en medios ricos y pobres, y en condiciones estándar o similares a las ambientales de la cueva de Altamira.

Los medios ricos empleados fueron dos: uno específico, medio de aislamiento de actinomicetos (AIA), y otro general, Tryptic Soy Agar Agar (TSA). Como medios pobres se emplearon TSA diluido 10 veces para evitar el crecimiento rápido de bacterias fáciles de crecer y este mismo TSA diluido suplementado con 5mM NaHCO<sub>3</sub>.

Las muestras de colonias se plaquearon directamente en estos medios en el mismo momento de la toma de muestras con asas estériles desechables. Las placas hasta su transporte al laboratorio permanecieron dentro de una bolsa térmica, transcurriendo unas 2 horas hasta su incubación. Las condiciones de crecimiento ensayadas fueron: TSA a 28°C y 16°C, TSA 0,1X a 28°C y 16°C

con o sin  $\text{NaHCO}_3^-$  (5mM), AIA a 28°C y 16°C y AIA a 16°C con 80% de humedad.

Colonias bien aisladas obtenidas en las placas primarias se resembraron en placas nuevas del mismo medio hasta obtener cultivos aparentemente puros. A partir de una colonia pura se extrajo el DNA total para la reacción de amplificación. Para cada muestra, la colonia seleccionada se resuspendió en 1 mL de agua MiliQ, se centrifugó a 12.000 rpm durante 1 min y se retiró el sobrenadante. Después a cada tubo de muestra, se añadió 100 µl de reactivo InstaGene™ Matrix (Bio-Rad, Hercules, CA), los tubos se resuspendieron con el vortex durante 10 a 30 segundos y se calentaron a 65°C durante 30 min. Posteriormente, los tubos se agitaron en vórtex de nuevo durante 10 a 30 segundos y se hirvieron a 100°C durante 8 min. Las muestras finalmente se agitaron durante 10 a 30 segundos y se centrifugaron a 12000 rpm durante 2 min.

Se utilizaron 2 µL del extracto de ADN como molde para la amplificación con primers específicos. Los primers utilizados en la reacción de amplificación fueron primers universales que permiten la amplificación de cualquier bacteria. El producto amplificado corresponde con las regiones V1-V9 del gen 16S rRNA, lo que constituye la molécula prácticamente completa. Los primers utilizados fueron PA8F (5'AGAGTTTGATCCTGGCTCAG3') y 1492R (5'GGTTACCTTGTTACGACTT3').

La reacción de PCR (50 µl vol final) se realizó con:

- 25 µl NZYTaq II Green Master Mix (2X) (NZYTech, Lisboa, Portugal)
- 2,5 µl de cada primer forward y reverse (0,5 µM cada uno)
- 2 µl de muestra de DNA (5 ng/µl)

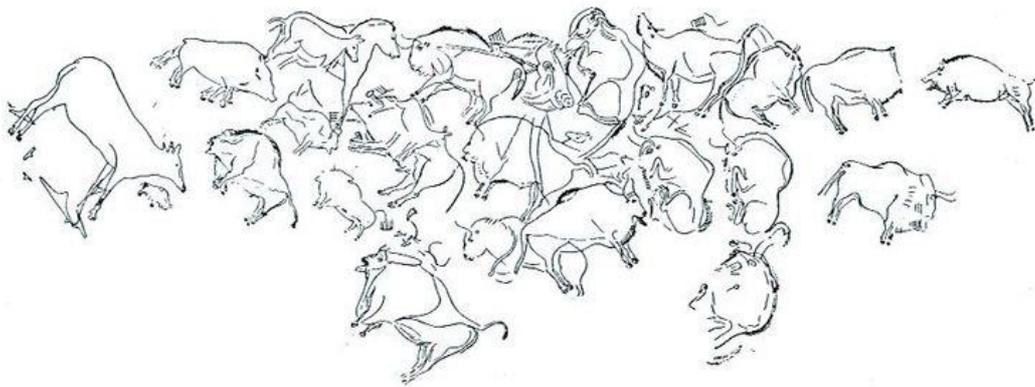
La reacción de la PCR se hizo en un termociclador 2720 Thermal Cycler (Applied Biosystems, California, Estados Unidos) y las condiciones fueron:

- un ciclo de 95°C durante 3 min
- 25 ciclos de
  - 94°C durante 30 seg

- 54°C durante 30 seg
- 72°C durante 1 min
- un ciclo final de extensión de 72°C durante 7 min.

Los productos de amplificación se purificaron con el kit GeneJET PCR Purification (Thermo Fisher Scientific, Waltham, MA). Se realizaron 3 reacciones de secuenciación por muestra y se secuenciaron por la técnica de Sanger. Los primers utilizados para la secuenciación fueron PA8F (5'AGAGTTTGATCCTGGCTCAG3'), 533F (5'GTGCCAGCAGCCGCGGTAA3') y 1100F (5'CAACGAGCGCAACCCT3'). Las secuencias obtenidas de 16S rRNA fueron visualizadas y editadas manualmente con el programa Chromas v2.6.6 (<http://technelysium.com.au/wp/chromas/>).

Las secuencias obtenidas de los aislados cultivables de las colonias (136 secuencias) y de las bacterias de referencias más cercanas (60 secuencias) se alinearon con ClustalW (Sievers et al., 2011). El árbol de los aislados cultivables de las colonias se realizó por el método de máxima verosimilitud basado en el modelo Tamura-Nei con el software de MEGA v7.0.26 (Kumar et al., 1994). Las secuencias utilizadas para este análisis fueron 196, que como se menciona arriba se añadieron las referencias más cercanas a los cultivables para confirmar su taxonomía. El número de acceso en GenBank de las secuencias de referencia se recoge en el Anexo-Tabla S11.



## 5. Resultados



## 5.1 Resultados técnica de amplicones

### 5.1.1 Métrica general

Un total de 57 muestras, de las cuales 8 fueron colonias beis, 18 colonias grises, 14 blancas y 15 amarillas, fueron analizadas por la técnica de amplicones. Sabiendo que las colonias pueden estar contaminadas con componentes propios de la pared, se analizaron 2 muestras de pared sin colonia aparente para tenerlo como control frente a las colonias. Los detalles de la toma de muestras se encuentran en el apartado de materiales y métodos. La estrategia de muestreo en este estudio fue recoger la mayor variabilidad posible dentro de los tipos de colonias distinguibles por su color. Las colonias fueron recogidas entre 2016 y 2019 de las zonas de la cueva en las que son más abundantes: Vestíbulo, Cruce y Pasillo de Polícromos (Anexo-Tabla S1). De las muestras de colonias se obtuvo suficiente cantidad de DNA para la obtención de amplicones y su secuenciación, y fue posible obtener lecturas suficientes como tener resultados significativos. La media de las lecturas obtenidas para la mayoría de las muestras fue cercana a los 100.000.

Los datos obtenidos, desde el procesamiento de las lecturas crudas hasta la obtención de OTUs se recogen en el Anexo-Tabla S3. El total de las lecturas crudas fueron 14.940.806 (Tabla 1). Las muestras control y beis produjeron menos lecturas con 220.883 y 840.587 respectivamente. En el caso del control sin colonia esto se debe a la menor cantidad de DNA obtenida, debido a la menor cantidad de microorganismos presentes en estas muestras sin colonias. En el caso de las colonias beis sospechamos que el menor rendimiento es debido a la presencia de inhibidores de la amplificación asociados a la localización en la que este tipo de colonias es más abundante. Por el contrario, para el resto de colonias se obtuvieron muchas más lecturas 5.923.687 grises, 4.511.624 amarillas y 3.444.025 blancas.

Una vez procesadas las lecturas, las secuencias dereplicadas (diferentes) obtenidas fueron 5.697.876 (un 38% de las lecturas crudas). La selección de las OTUs se hizo *de novo* agrupando las lecturas en clusters, sin ser

comparadas contra ninguna base de datos de secuencias de referencia externa. Se utilizó una homología del 97% para asignar lecturas a un mismo cluster. El resultado final fue la agrupación de secuencias en 4.288 OTUs.

<b>Métrica</b>	<b>Muestra</b>
<b>Número de muestras</b>	57
<b>Lecturas sin procesar</b>	14.940.806
<b>Lecturas totales procesadas</b>	5.697.876
<b>Número de OTUs</b>	4.288

Tabla1. Datos generales obtenidos tras el procesamiento de las secuencias a través de la plataforma de QIIME2.

## **5.1.2 Análisis de la biodiversidad**

### **5.1.2 .1 $\alpha$ -Diversidad**

Para tener una imagen de la riqueza o diversidad de componentes presentes en cada tipo de colonia y por otro lado determinar la homogeneidad y establecer relaciones entre los diferentes tipos de colonias, se llevaron a cabo los análisis de biodiversidad basados en la tabla de abundancias de cada OTU en cada una de las muestras. Para una cuantificación de la biodiversidad específica en las muestras, se utilizó la “diversidad alfa”, esto es, la diversidad dentro de cada muestra, lo que es en parte equivalente a la riqueza y la distribución de la misma en especies.

Antes de proceder con los análisis de  $\alpha$ -diversidad, se comprobó por el método de rarefacción que la cantidad de lecturas representaba la diversidad total. Este análisis demostró que con la cantidad de lecturas utilizadas se alcanzaba en todas las muestras el nivel de saturación, lo que nos indica que los resultados de diversidad no estarán influenciados por el tamaño de cada muestra, a pesar de contener números de lecturas muy diferentes entre ellas. Estas curvas de acumulación de especies crecen a medida que se van recolectando más secuencias diferentes hasta que llega un punto en el que nuevas secuencias no representan nuevas especies, y por tanto se aplanan la curva llegando a una

asíntota. Es lógico por tanto que las curvas tengan un comienzo exponencial elevado al comenzar y, que vaya normalizándose a medida que se completan las secuencias raras. Este momento nos indica que hemos caracterizado la totalidad de las especies diferentes presentes en cada muestra.

En este estudio, el esfuerzo de secuenciación aplicado en las muestras fue suficiente para la caracterización total de la diversidad, y en muchos casos solo con una fracción de las lecturas analizadas (Figura 29). La cantidad de secuencias mínima observada en las muestras fue para B12M6 con 6.027 secuencias y la máxima en G11M3 con 389.515 secuencias. La media de la profundidad de secuenciación de las muestras en este estudio fue de 99.962 secuencias. Para este valor podemos observar en la figura 29 que muchas de las muestras ya han alcanzado su valor máximo de meseta, lo que significa que se ha detectado el máximo de diversidad. El número de secuencias analizado en todas las muestras fue suficiente para determinar la diversidad existente.

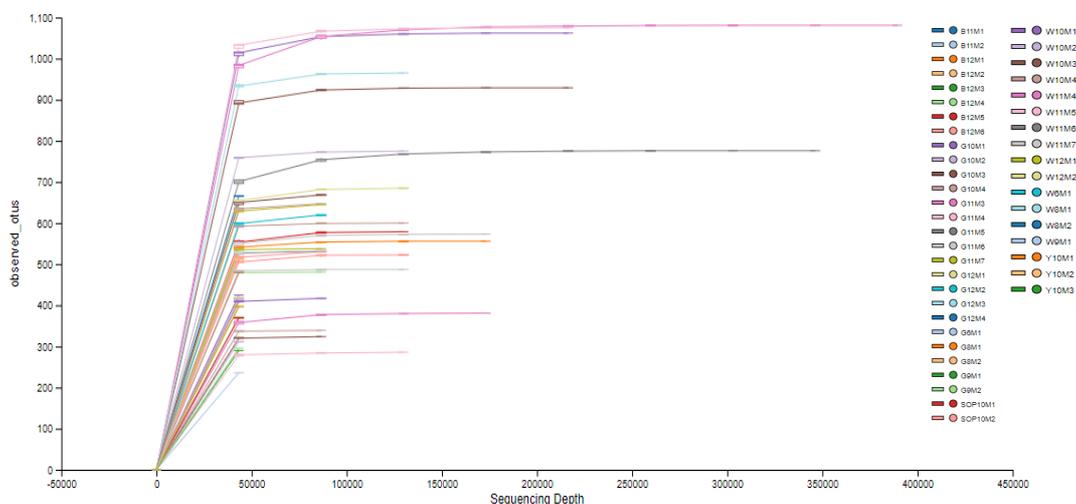


Figura 29. Curvas de rarefacción para las muestras de colonia visibles y control analizados por la técnica de amplicones.

Para cuantificar la biodiversidad específica en las muestras, se tuvo en cuenta el número de OTUs observado y además se calcularon los índices de diversidad de Shannon y el índice de Dominancia. Estos índices nos permiten cuantificar el número de especies y su representatividad en términos de abundancia proporcional de cada especie. El número de OTUs es un estimador

de la riqueza en la muestra, únicamente tiene en consideración el número de especies y no la abundancia de cada una. El índice de Shannon es uno de los índices más empleados en ecología y es un índice de equidad. Este índice mide la cantidad de especies que existen en la muestra (riqueza) y la cantidad relativa de individuos que hay para cada una de las especies (uniformidad). A medida que la riqueza y la equidad de las especies aumentan, la diversidad aumenta. Por tanto, más diversidad significa mayor estabilidad. Por otra parte, el índice de dominancia es un índice inverso al concepto de uniformidad o equidad de la comunidad, y representa la medida de abundancia relativa de una o dos taxones predominantes en la comunidad.

En el análisis de las muestras observamos que las más diversas fueron las muestras control. Estas muestras, a pesar de contener menos biomasa de microorganismos, tienen un gran número de especies (riqueza) que se distribuyen de manera bastante uniforme (Figura 30). Podemos encontrar que muchas de las muestras de colonias grises tienen una riqueza superior (618 media de OTUs en colonias grises y 497 en control), pero, por el contrario, no son tan diversas en comparación al control (5,92 media del índice de Shannon en las colonias y 7,09 en el control).

Si nos fijamos tan sólo en las muestras de las colonias, las colonias grises fueron con diferencia las más diversas y ricas. Esta riqueza disminuye en las colonias amarillas, blancas y beis siendo muy parecido entre ellas. Por el contrario, la diversidad fue menor para las colonias amarillas y beis, y especialmente en blancas. En concreto en algunas muestras de colonias blancas se observa una alta dominancia de un taxón. Debemos tener en cuenta que en la toma de muestras se obtiene material de la pared que se analiza junto con el de la colonia y esto puede sobredimensionar los resultados de diversidad. Los resultados en general indican que las colonias presentan un grado medio de diversidad en comparación con otros microbiomas naturales.

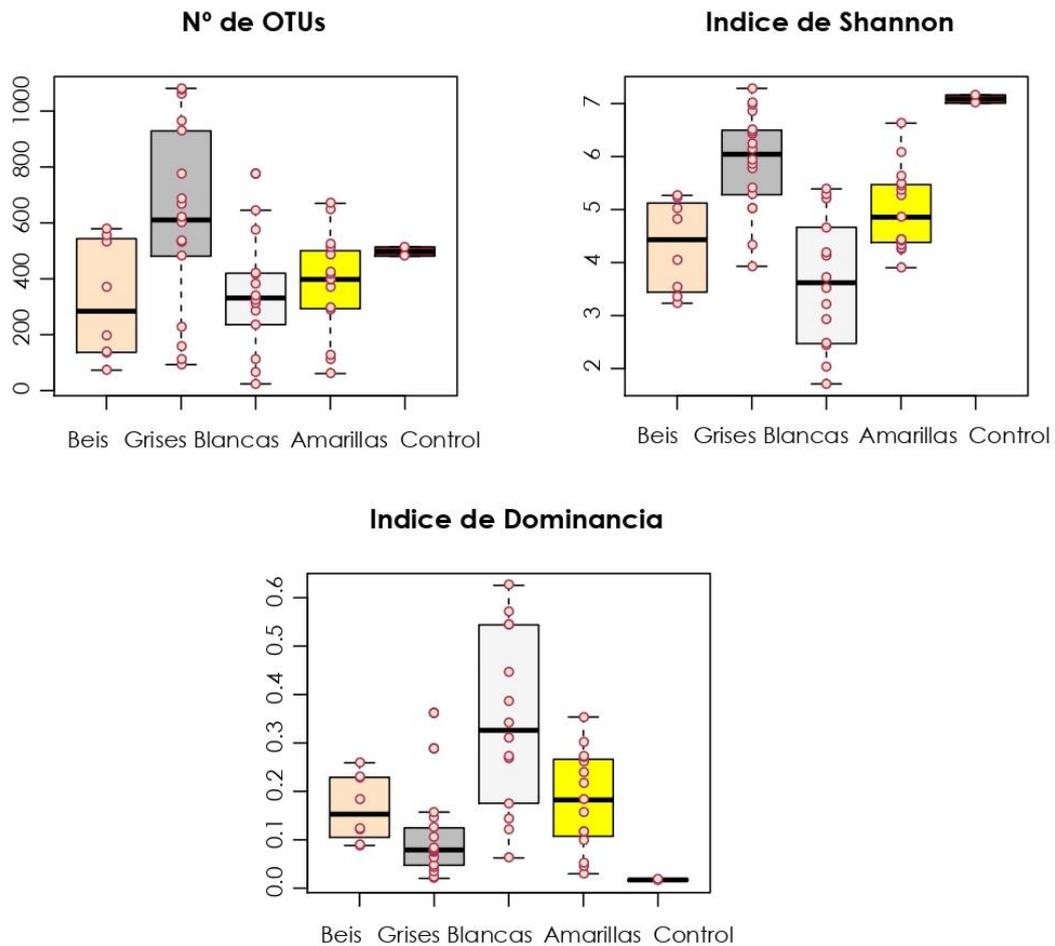


Figura 30. Representación de los valores de  $\alpha$ -diversidad en las muestras de colonia y control analizado por la técnica de amplicones. Para cada muestra se representa el n° de OTUs y los índices de Shannon y de Dominancia.

### 5.1.2 .2 $\beta$ -Diversidad

El análisis de la beta diversidad mide el grado de diversidad entre comunidades basados en diferencias de diferentes parámetros. Dichas diferencias pueden evaluarse con criterios basados en distancias o divergencia (indicadores que miden el grado de semejanza) o en similitudes (indicadores que miden la similitud) entre las muestras a partir de datos cualitativos (índices de incidencia como presencia/ausencia de especies con el índice de Jaccard) o cuantitativos (abundancia proporcional de cada especie como el índice de Bray-Curtis). Para comparar gráficamente en estos términos están los métodos de ordenación, los

cuales organizan a partir de matrices la semejanza en la composición (riqueza) y estructura (abundancia) entre las muestras en base a estos coeficientes.

El análisis multidimensional (MDS) aplicado a las muestras de colonias para este estudio separa las muestras en función a la disimilitud o desemejanza en la composición y estructura basado en el índice de Bray-Curtis, el cual se considera como una medida de la diferencia entre las abundancias de cada especie presente. Como se puede observar en la figura 31, con este método encontramos que las muestras se agruparon en 4 clusters bien definidos. El cluster1 correspondió con la mayor parte de las muestras blancas (12/14) y el cluster 2 con todas las muestras beis (8/8). El tercer cluster, agrupó alrededor de la mitad de las muestras de colonias grises (10/18). Estos clusters se encuentran muy desplazados del resto indicando una composición bastante diferente del resto. El último cluster, el 4, es una mezcla de todas las muestras amarillas (15/15), blancas (2/14), restantes muestras grises (8/18) y muestras control (2/2) en dónde observamos una distribución no tan homogénea y compacta como en los anteriores. Se observa que las muestras G9M2, G11M3, G12M3 y las muestras control se encuentran desplazadas del resto de las muestras dentro de este último cluster. En estas muestras la diversidad es más alta (con índices de Shannon alrededor de 7) y eso hace que su composición difiera del resto de muestras del cluster. La presencia de dos clusters dentro de colonias grises y blancas sospecha la posibilidad de detección de muestras con distinta composición para estas colonias a pesar de presentar el mismo color.

Las muestras de colonias por tanto se asociaron en grupos diferentes, las beis formaron el grupo más compacto y homogéneo (cluster 2), por el contrario, las colonias blancas y grises formaron grupos compactos (cluster 1 y 3) pero que no contuvieron todas las muestras. Parte de las muestras blancas y grises agruparon en un cuarto cluster, el 4, que es el más disperso de todos.



En el caso de las muestras control, el filo más abundante fue *Proteobacteria* con 37,59% de las secuencias seguido de *Acidobacteria* con 18,74% y a continuación, sin mucha diferencia entre ellos, *Actinobacteria* con 9,04% y *Planctomycetes* con 9,00%.

Al comparar ambas muestras se observan los mismos filios, aunque con proporciones muy diferentes. Entre las diferencias destaca un filo *Gemmatimonadetes* con 4,99%, mientras que, en las muestras de colonias representa un 0,50% del total. Por el contrario, el filo *Rokubacteria* detectado en las muestras de colonias no es un filo abundante en las muestras control con 0,60% de las secuencias. Los filios minoritarios cuya abundancia fue inferior al 1%, fueron un total de 32 filios en ambas muestras, contenían el 4,72% de las secuencias en las colonias y 4,10% en el control, mientras que las secuencias no clasificadas fueron un porcentaje mínimo, con 0,10% y 0,23% del total, respectivamente.

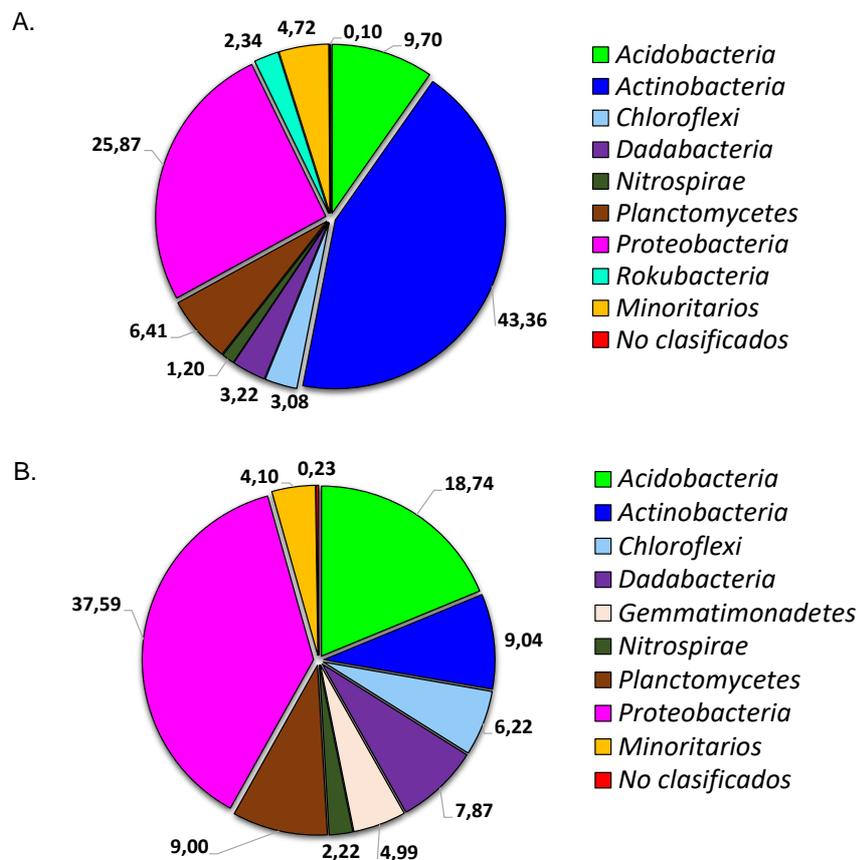


Figura 32. Representación de los filios más abundantes (con una abundancia >1%) en las muestras de las colonias (A) y control (B).

Si nos fijamos en las 10 OTUs con mayor abundancia de las 4.288 OTUs resultantes (Figura 33), observamos algunas OTUs, en concreto OTUs 1, 2 y 4, muy predominantes para cada tipo de colonia y que pertenecen al filo *Actinobacteria*. Esta abundancia tan fuerte las diferencia claramente de las demás OTUs y nos permite denominarlas como Componentes Mayoritarios (CMs) de las colonias. La taxonomía asignada a estas 10 OTUs más abundantes se recoge en el Anexo-Tabla S4.

En base a esta distinción, encontramos que en las colonias amarillas el CM era la OTU2, con una frecuencia que variaba del 12-60% entre las diferentes muestras. Esta OTU2 fue clasificada por QIIME2 en la familia *Pseudonocardiaceae* y género *Crossiella*.

En las colonias blancas observamos una situación diferente, la mayoría de las muestras presentaron la OTU1 como CM (abundancia individual 33-79%) mientras que, en 2 muestras el CM fue la OTU2 (abundancia individual 21,41 y 33,08%). La OTU1 correspondió con la familia *Euzebyaceae*. En base a la diferencia observada en sus CMs, proponemos separar las colonias blancas en dos subtipos: el primero con la OTU2 como CM que hemos llamado subtipo1 (W1), y el que tiene la OTU1 como CM al que hemos llamado subtipo2 (W2).

En las colonias grises también observamos el mismo escenario que en las colonias blancas. Encontramos un grupo de colonias grises que presentaron como CM la OTU4 (abundancia individual 7-39%) y que hemos denominado subtipo1 (G1), y otro cuyo CM fue la OTU2 (abundancia 6-60%) al que hemos llamado subtipo2 (G2). La OTU4 que representa el CM del subtipo G1 también fue clasificado en la familia *Pseudonocardiaceae* y género *Crossiella*.

Las colonias beis no presentaron un CM único. En ellas observamos dos OTUs (OTUs 1 y 2) con una abundancia muy destacada por encima del resto de componentes. Estas dos OTUs ya habían sido identificadas como CM en otros tipos de colonia. Por tanto, ambos fueron considerados como CMs en estas colonias siendo la abundancia de la OTU1 entre el 20-42% y del 14-30% la de la OTU2.

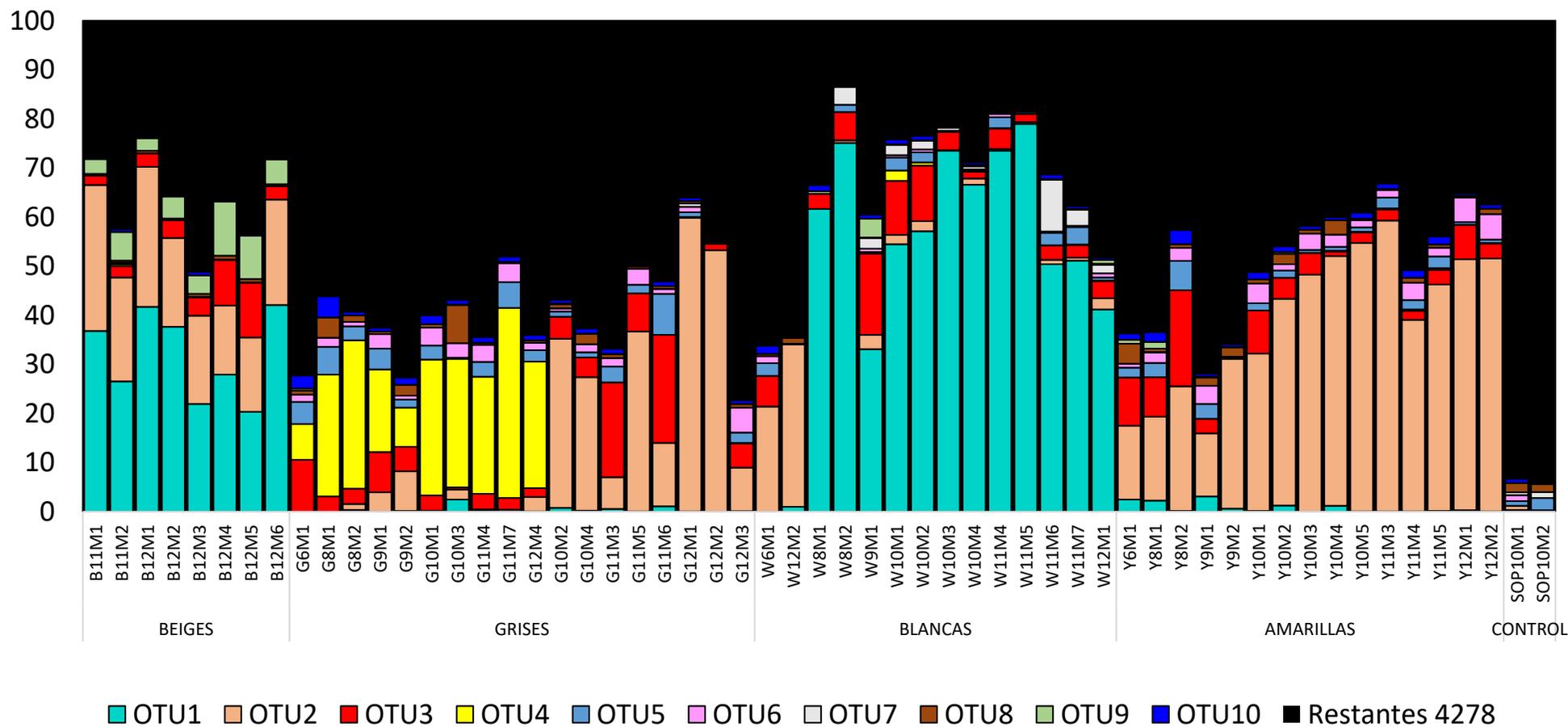


Figura 33. Lista de las 10 OTUs con mayor frecuencia en las muestras de colonias visibles y en el control analizadas por la técnica de amplicones.

Los análisis de  $\beta$ -Diversidad también han separado estos subtipos en diferentes clusters, indicando por tanto una composición diferente entre ellos (Figura 31). En el caso de las colonias blancas, las muestras del subtipo W2 se agruparon con el cluster 1, mientras que las muestras del subtipo W1 lo hicieron con el cluster 4. Lo mismo ocurrió en el caso de las colonias grises, las muestras subtipo G1 agruparon con el cluster 3, a excepción de la muestra G9M2. La muestra G9M2 es un caso particular al presentar misma proporción de la OTUs 2 y 4 (CMs en ambos tipos de colonias grises), posiblemente esta muestra contenga colonias grises de ambos tipos que se han cogido conjuntamente en la toma de muestras. Las muestras de colonias grises del subtipo G2, sin embargo, agruparon todas en el cluster 4 con las muestras blancas subtipo W1 y amarillas.

La presencia de subtipos de colonias nos indica que es posible cierto grado de heterogeneidad en muestras de colonias del mismo color, que a simple vista es inapreciable. Dicho de otro modo, nos previene ante la opción de que colonias con una apariencia similar puedan presentar diferente composición, sobre todo diferentes CMs. Una descripción de las colonias visibles teniendo en cuenta estos nuevos subtipos reflejará mucho mejor la realidad microbiológica de las colonias.

De las casi 5.000 OTUs detectadas, hemos querido seleccionar un número más manejable, y posiblemente también más realista, de OTUs que nos sirvan para definir la diversidad de las colonias visibles de la cueva de Altamira. Para ello nos fijamos en que las 20 OTUs más abundantes (aquellas con mayor frecuencia) representaban alrededor del 60% de las secuencias procesadas. Al mismo tiempo se detecta un número muy alto de OTUs muy minoritarias que podrían ser contaminantes del soporte. Por tanto, esta selección se hizo filtrando aquellas OTUs que mostraron una abundancia media  $>1\%$  para cada subtipo y que al menos estuvieran presentes en 2 muestras.

De esta manera, las OTU seleccionadas conformarían lo que consideramos un “core” para las muestras en la cueva de Altamira. Este core que contiene los

componentes más abundantes en las muestras analizadas de las colonias visibles de Altamira, contiene las 36 OTUs que se muestran en la Figura 34. Dentro de este core seleccionado, las OTUs correspondientes a los CMs fueron las más abundantes y características para cada subtipo clasificando con el filo *Actinobacteria*, familias *Euzebyaceae* y *Pseudonocardiaceae*. El filo *Proteobacteria* fue el segundo filo más predominante entre los componentes secundarios. En concreto, se identificó la OTU3 correspondiente con la clase *Gammaproteobacteria*, orden *Nitrosococcales*, familia *Nitrosococcaceae* y género *wb1-P19* como la OTU más predominante en todas las muestras, a excepción de 5 (G12M1, W12M2, Y9M2, SOP10M1 y SOP10M2).

Otras 3 OTUs también fueron frecuentes en muchas de las muestras de las colonias concretamente la OTU5 (filo *Acidobacteria*, clase *Blastocatellia* Subgrupo 4 y orden 11-24), la OTU6 (filo *Dadabacteria*, clase *Dadabacteriia* y orden *Dadabacteriales*) y la OTU8 (filo *Proteobacteria*, clase *Alphaproteobacteria*, orden *Rhizobiales* y familia *Hyphomicrobiaceae*). Se detectaron otros filos entre los componentes secundarios que correspondieron con *Chloroflexi*, *Nitrospirae* y *Rokubacteria*. La taxonomía de las OTUs del core se recogen en el Anexo-Tabla S4.

A continuación, hacemos un análisis más detallado de las OTUs presentes para cada subtipo de colonia. Estas OTUs son el resultado de un análisis de varias muestras del mismo tipo y corresponden por tanto con componentes que se suelen detectar con frecuencia y que esperaríamos detectar en la mayoría de las muestras de cada subtipo que se fueran a analizar en el futuro.

### Colonias beis

Las colonias beis fueron las menos diversas y las que presentaron una mayor homogeneidad presentando no más de 9 componentes con una abundancia promedio mayor al 1% en las muestras. Lo más característico en estas colonias fue la presencia de 2 actinobacterias como CMs (OTUs 1 y 2) con una abundancia media >20%. Como componentes secundarios más abundantes se detectaron 2 bacterias de la clase *Gammaproteobacteria*, la OTU9 (orden

*PLTA13*) con una abundancia del 5,57% y la OTU3 (*Nitrosococcaceae* del género *wb1-P19*) con 4,70%. Como componentes secundarios minoritarios (<2,5%) se detectaron bacterias correspondientes con los siguientes filios: *Proteobacteria* con la OTU21 (clase *Gammaproteobacteria*, orden *Xanthomonadales*, familia *Xanthomonadaceae* y género *Lysobacter*) y la OTU66 cuya asignación se queda a nivel de la clase *Deltaproteobacteria*, el filo *Actinobacteria* con la OTU57 (clase *Thermoleophilia*, orden *Solirubrobacterales* y familia *67-14*), el filo *Dadabacteria* con la OTU37 (clase *Dadabacteriia* y orden *Dadabacteriales*), y, finalmente, el filo *Chloroflexi* con la OTU36 (clase *Anaerolineae*, orden *Caldilineales*, familia *Caldilineaceae* y género *Litorilinea*). Estas OTUs secundarias, a excepción de la OTU3, se pueden considerar características de las colonias beis.

### Colonias grises

Las colonias grises fueron las más diversas de todas y como ya se ha mencionado antes las hemos categorizado en 2 subtipos según el CM que presentaban. Las del subtipo G1 presentaron la OTU4 (género *Crossiella*) y las del subtipo G2 la OTU2 (género *Crossiella*). En ambos casos se detectaron con una abundancia media >20% en las muestras. En ambos subtipos se observó una alta presencia de componentes secundarios clasificados con el filo *Proteobacteria*.

El análisis en las colonias grises subtipo G1 nos sugiere unos 15 componentes con una abundancia >1%. Encontramos la OTU4 como CM para este subtipo y una serie de OTUs secundarias, algunas mencionadas anteriormente como la OTU3 género *wb1-P19* (4,08%), OTU5 *Acidobacteria* orden *11-24* (3,25%), OTU6 orden *Dadabacteriales* (2,34%), OTU18 orden *Betaproteobacteriales* (2,24%), OTU2 género *Crossiella* (1,91%) y OTU8 familia *Hyphomicrobiaceae* (1,85%). El resto de OTUs (8) con una abundancia <1,6% correspondieron con componentes pertenecientes la gran mayoría al filo *Proteobacteria* entre las que encontramos la clase *Alphaproteobacteria* con la OTU13 (orden *Rhizobiales*, familia *Hyphomicrobiaceae* y género *Hypomicrobium*) y la clase *Gammaproteobacteria* con los órdenes *Betaproteobacteriales* en OTUs 10, 14 y

34 y *PLTA13* en la OTU12, y la OTU26 cuya asignación se queda a nivel de clase. Otros filos encontrados fueron el filo *Rokubacteria* en la OTU15 (clase *NC10* y orden *Rokubacteriales*) y *Nitrospirae* en la OTU16 (clase *Nitrospira*, orden *Nitrospirales*, familia *Nitrospiraceae* y género *Nitrospira*).

En las colonias grises subtipo G2 el número de componentes con una abundancia >1% fue de 11. Aparte de su CM, se detectaron entre los componentes secundarios la OTU3 como la más abundante (8%). El resto de los componentes (9) presentaron una abundancia menor (<2,5%), destacando a parte de las OTUs 5 y 6 ya mencionadas compartidas con el subtipo G1, el filo *Proteobacteria* de nuevo como el más abundante con la OTU19 (clase *Deltaproteobacteria*, orden *Myxococcales*, familia *Haliangiaceae* y género *Haliangium*), OTU30 (clase *Gammaproteobacteria* y orden *CCD24*), OTU25 (clase *Alphaproteobacteria* y orden *Rhizobiales*), OTU47 (clase *Alphaproteobacteria*, orden *Azospirillales*, familia *Inquilinaceae* y género *Inquilinus*), y finalmente, OTU50 (clase *Gammaproteobacteria*, orden *Xanthomonadales*, familia *Xanthomonadaceae* y género *Arenimonas*). También se detectaron los filos *Chloroflexi* en la OTU20 (clase *Anaerolineae* y orden *Ardenticatenales*) y *Actinobacteria* en la OTU58 (clase *Actinobacteria*, orden *Pseudonocardiales*, familia *Pseudonocardiaceae* y género *Pseudonocardia*). Estas OTUs secundarias, a excepción de las OTUs 3, 5 y 6, se pueden considerar características en este subtipo de colonia.

### Colonias blancas

Las colonias blancas mostraron un contexto muy parecido al de las colonias grises. También se distinguieron 2 subtipos en ellas: W1 que presentó la OTU2 (género *Crossiella*) como CM con una abundancia media del 27%, y W2 con la OTU1 (familia *Euzebyaceae*) como CM con una abundancia media del 60%.

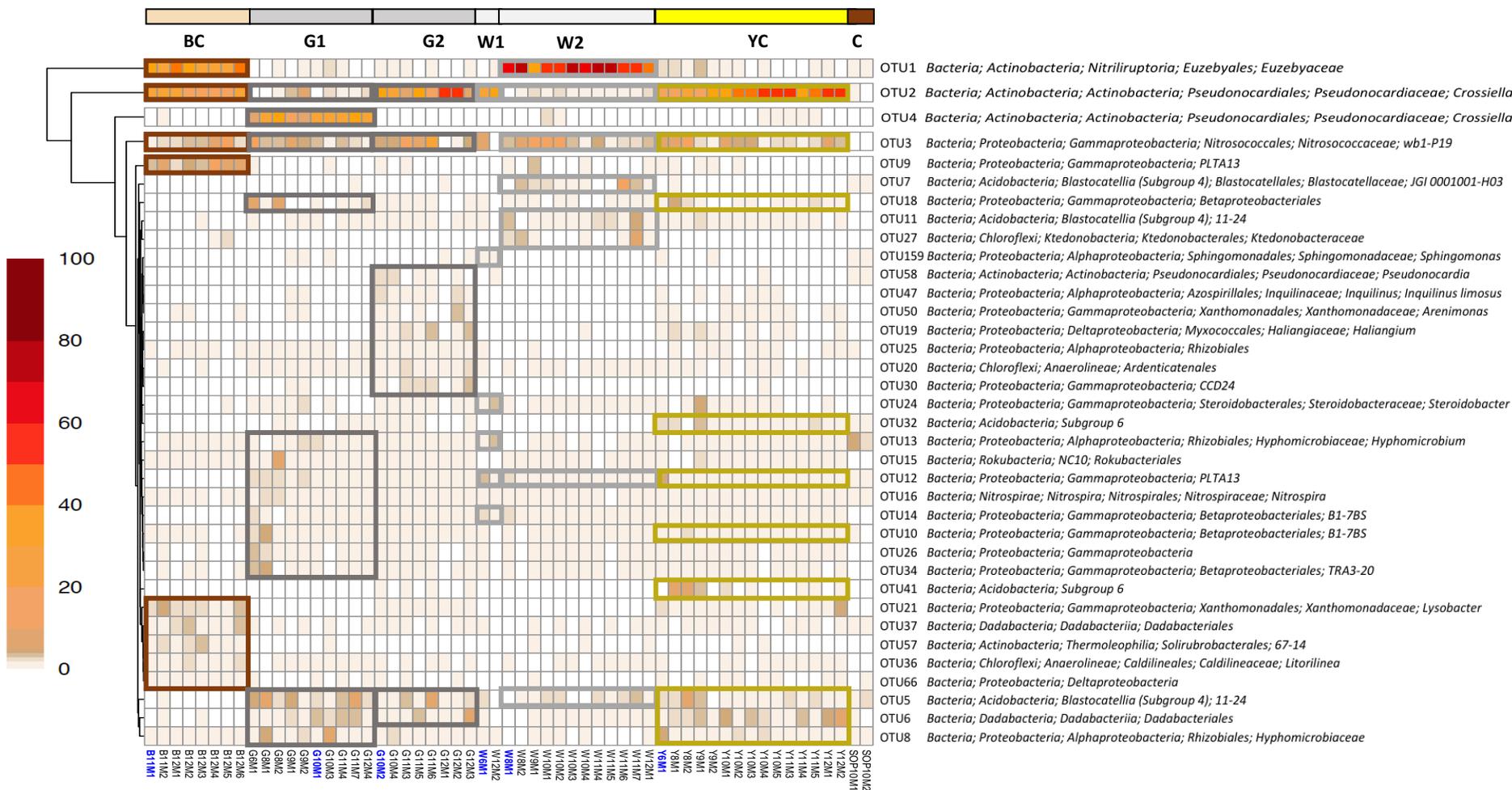


Figura 34. Heatmap de abundancia de las OTUs que conforman el core en las muestras de colonias visibles y control de la cueva de Altamira analizadas por la técnica de amplicones. La selección del core para cada subtipo representa aquellas OTUs con un promedio >1% y presentes en al menos 2 muestras. Las muestras en color azul fueron seleccionadas para ser secuenciadas además por la técnica de shotgun. Las OTUs recogidas en cuadros de color representan el core seleccionado para cada subtipo en particular.

Las colonias blancas subtipo W1 fueron detectadas con menos frecuencia, de hecho, sólo 2 muestras de las 14 colonias blancas analizadas correspondieron a este subtipo. Con un tamaño muestral tan pequeño, la composición final de este subtipo podría variar al analizar más muestras. De 17 OTUs con una abundancia mayor al 1%, sólo 6 fueron compartidas entre estas dos muestras que fueron las que se seleccionaron para su core. Como ocurre en todas las colonias, la OTU más característica en estas muestras fue claramente su CM (OTU2). Las otras cinco OTUs con menor abundancia (<2,5%) fueron clasificadas dentro del filo *Proteobacteria*, tres de ellas en la clase *Gammaproteobacteria* con la OTU24 (orden *Steroidobacterales*, familia *Steroidobacteraceae* y género *Steroidobacter*), OTU12 (orden *PLTA13*) y OTU14 (orden *Betaproteobacteriales* y género *B1-7BS*); y las otras dos en la clase *Alphaproteobacteria* con la OTU13 (orden *Rhizobiales*, familia *Hyphomicrobiaceae* y género *Hyphomicrobium*) y la OTU159 (orden *Sphingomonadales*, familia *Sphingomonadaceae* y género *Sphingomonas*).

Las colonias blancas subtipo W2 fueron las más frecuentes entre las colonias blancas, perteneciendo a él 12 muestras de un total de 14 analizadas. El análisis de composición simplificada nos sugiere que ocho OTUs representan la mayoría de su diversidad. A parte de su CM, OTU1, encontramos también la OTU3 (orden *Wb1-P19* de las *Nitrosococcaceae*) formando parte del grupo de los componentes secundarios más abundantes (6%). Entre los otros seis componentes secundarios con menor abundancia destaca el filo *Acidobacteria* con la OTU7 (clase *Blastocatellia Subgrupo 4*, orden *Blastocatellales*, familia *Blastocatellaceae* y género *JGI 0001001-H03*) y las OTUs 11 y 5 (clase *Blastocatellia Subgrupo 4* y orden *11-24*). El resto de los filios fueron *Chloroflexi* con la OTU27 (familia *Ktedonobacteraceae*), el filo *Actinobacteria* con la OTU2 (género *Crossiella*) y el filo *Proteobacteria* con la OTU12 (clase *Gammaproteobacteria* y orden *PLTA13*). De esta selección de core, tan sólo las OTUs 7, 11 y 27 se puede considerar características únicamente en este subtipo.

Colonias amarillas

Las colonias amarillas pueden considerarse de 10 OTUs como sus componentes esenciales. La OTU principal para este tipo de colonias fue su CM correspondiente con la OTU2 (género *Crossiella*) con una abundancia del 38%. En los componentes secundarios destacó de nuevo la OTU3 como secundario más abundante con una abundancia del 5%. El resto de OTUs (8) constituyen componentes secundarios con una menor abundancia perteneciendo a 3 filos: *Dadabacteria* con la OTU6 (clase *Dadabacteriia* y orden *Dadabacteriales*), *Acidobacteria* con la OTU5 (clase *Blastocatellia Subgroup 4* y orden *11-24*) y las OTUs 41 y 32 (ambas correspondieron a la clase *subgrupo 6*), y finalmente el filo *Proteobacteria* con la OTU8 (clase *Alphaproteobacteria*, orden *Rhizobiales* y familia *Hyphomicrobiaceae*), OTU10 (clase *Gammaproteobacteria*, orden *Betaproteobacteriales* y familia *B1-7BS*), OTU18 (clase *Gammaproteobacteria* y orden *Betaproteobacteriales*) y OTU12 (clase *Gammaproteobacteria* y orden *PLTA13*). De todas estas OTUs, sólo las OTU41 y OTU32 se pueden considerar características en colonias amarillas.

En esta selección de componentes centrales encontramos claramente una serie de OTUs más abundantes presentes en muchas de las muestras de colonias, en especial, las OTUs correspondientes a los componentes mayoritarios (OTUs 1, 2 y 4). Pero también entre los componentes secundarios encontramos OTUs con una abundancia menor a los CMs, pero compartidas entre los subtipos de colonias como las OTUs 3, 5, 6 y 8. De igual manera también se detectaron OTUs más específicas para cada subtipo de colonia.

Algunas de estas OTUs compartidas entre colonias, fueron también detectadas en las muestras control, pero en menor abundancia. Esta abundancia media en las muestras control fue para la OTU3 0,05%, OTU5 1,72%, OTU6 0,61% y OTU8 1,75%. Hay 4 OTUs más que también se detectaron en las muestras control con la misma o mayor abundancia que en las colonias sospechando que pudieran ser contaminantes de soporte que fueron: la OTU13 (3,04%), OTU16 (0,96%) y la OTU37 (0,95%). Se considera, por tanto, que salvo las 3 últimas OTUs (OTUs 13, 16 y 37) que presentaron una mayor o similar

abundancia también en las muestras control, el resto corresponderían con componentes reales de las colonias.

Para ampliar la comparación entre los componentes de las colonias y la microbiota propia del suelo y del soporte, comparamos las muestras de colonias con más muestras de suelo y soporte, obtenidas de diferentes zonas de la cueva. Se analizaron un total de 86 muestras de suelo y soporte en las siguientes zonas de la cueva: Vestíbulo, Cruce, Polícromos, Hoya Bisonte y Hoya Cabra.

De un total de 15.783.202 lecturas crudas, se filtraron en 4.451.858 que se agruparon en 7.933 OTUs. En este caso el filo más abundante para estas muestras fue *Proteobacteria* con 28,07% de las secuencias seguido de *Acidobacteria* con un 19,48%, *Actinobacteria* 10,83%, y *Planctomycetes* 10,21% (Anexo-Figura S1). Otros filios detectados correspondieron con *Chloroflexi* con un 6,53% de las secuencias, *Firmicutes* 6,27%, *Rokubacteria* 4,71%, *Gemmatimonadetes* con 2,75%, *Nitrospirae* con 1,88%, *Latescibacteria* con 1,85%, *Bacteroidetes* con 1,72% y *Verrucomicrobia* con 1,69%. Los otros filios restantes (34), los cuales individualmente obtuvieron una abundancia <1%, representaron un 3,77% de las secuencias entre los que destacaron los filios *GAL15* o *Dadabacteria* entre otros. Las secuencias no clasificadas correspondieron con un porcentaje mínimo 0,26%.

Cuando comparamos los componentes de muestras de colonias con los de estas muestras de suelo y soporte, destaca la abundancia diferencial del filo *Actinobacteria* en las muestras de colonias, reiterando el hecho de que este filotipo es de los más importantes en este tipo de muestras (Figura 35). Por el contrario, en las muestras de suelo y soporte fueron más abundantes los filios *Proteobacteria* y *Acidobacteria*, al igual que otros filios no detectados en las colonias como *Firmicutes* especialmente en muestras de la Hoya Cabra, *Latescibacteria* en muestras de Polícromos o *GAL15* y *Gemmatimonadetes* en casi todas las localizaciones.

En cuanto a las 10 OTUs detectadas con mayor frecuencia en estos 2 tipos de muestras, constatamos en primer lugar que la abundancia que representaron estas OTUs en las muestras de suelo y soporte (4-50%) fue mucho más baja que en las colonias (27-87%) reflejando la existencia de una mayor diversidad en las muestras de suelo y soporte (Figura 36). La diversidad en estas muestras fue con diferencia mucho más alta que en las muestras colonias con valores de Shannon en la mayoría de 5 a 8,5 (Anexo-Figura S2). Además, en estas muestras no hay una dominancia de una o unas cuantas OTUs como sucede en las muestras de colonias (índice de dominancia  $< 0,2$ ). Sin embargo, sí que se observan algunas OTUs con una abundancia mayor en algunas muestras como por ejemplo en muestras de Polícromos, Vestíbulo y Hoya Bisonte destaca una *Acidobacteria Subgroup 6* y una *Gammaproteobacteria* del género *wb1-P19*, y en la zona de la Hoya Cabra 2 *Firmicutes* asignadas a los géneros *Bacillus* y *Sporosarcina* (Figura 36).

La única OTU de esta selección que está compartida entre las colonias y las muestras de suelo y soporte fue la *Nitrosococcaceae* perteneciente al género *wb1-P19* tratándose del mismo amplicón con un 99,54% identidad (OTU3 en colonias y OTU2 en muestras de suelo y soporte). Esta OTU que fue muy poco abundante en las muestras control utilizadas para comparar contra las muestras de colonias, sin embargo, aparece como segunda OTU más abundante en general en las muestras de suelo y soporte, aunque no alcanza el nivel observado en las colonias. No obstante, la OTU no fue detectable en todas las muestras, solamente en 45 de las 86 muestras analizadas. En el Vestíbulo (donde se concentra la mayor abundancia de colonias) fue detectada solamente en el soporte en 2 muestras de 14 con una abundancia que osciló entre 1 y 3%. Por el contrario, en el suelo se encontró en 11 de 12 muestras con abundancias que oscilaron entre 1 y 37%.

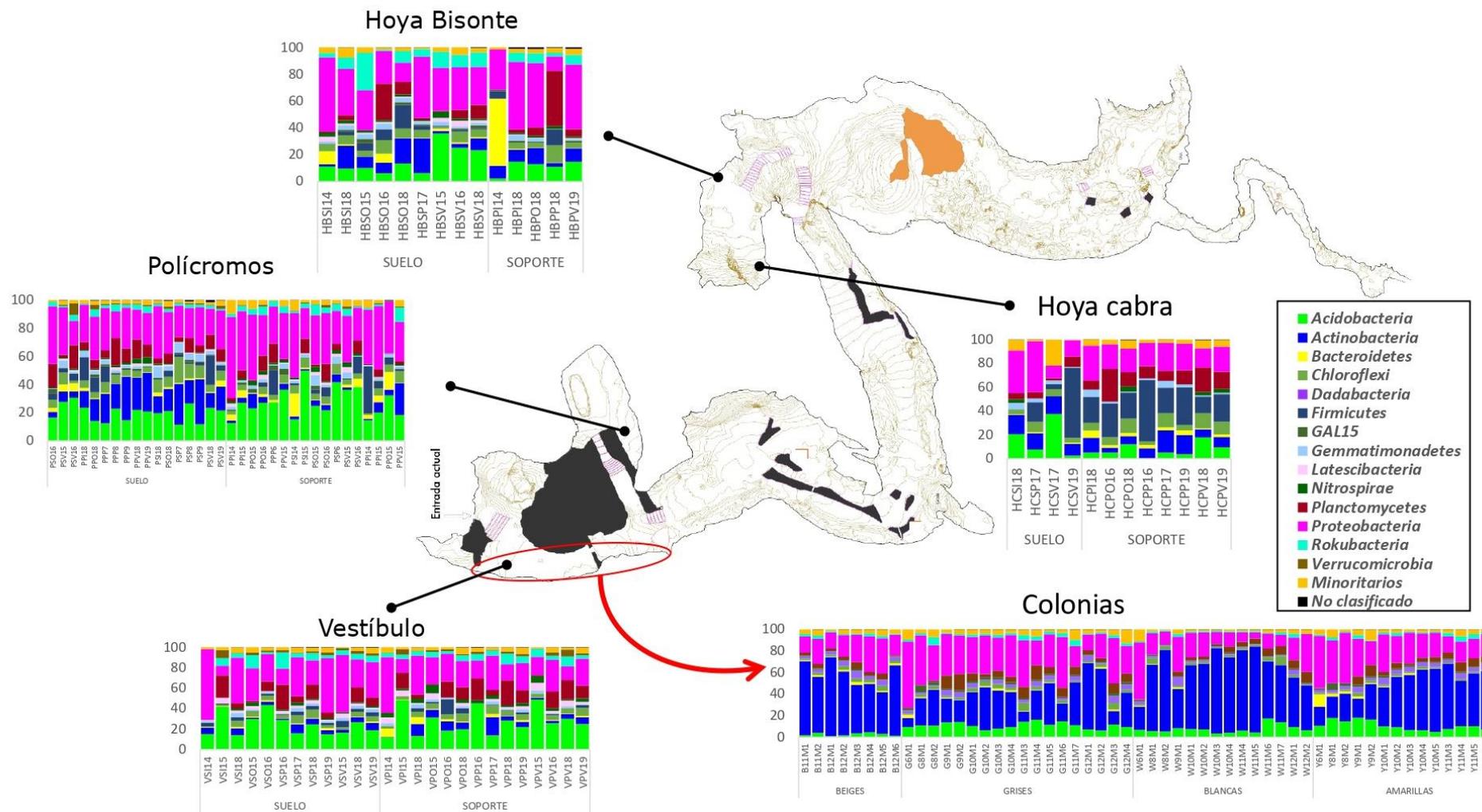


Figura 35. Comparación de los filos más representativos para las muestras de suelo y soporte y muestras de colonia en distintos puntos de la Cueva de Altamira analizadas por la técnica de amplicones.

En general los CMs de las colonias están ausentes o son muy poco abundantes en muestras de suelo y soporte. Así la OTU2 de las colonias, que constituye un componente mayoritario en colonias amarillas, blancas subtipo W1, grises subtipo G2 y beis, fue detectada con una abundancia cercana al 1% también en estas muestras, en concreto en 7 muestras de Polícromos (6 de soporte y 1 de suelo) y en 2 muestras de Vestíbulo soporte. Sin embargo, los otros componentes mayoritarios en concreto las OTUs 1 y 4 sólo fueron detectadas en una muestra de suelo del Vestíbulo.

En resumen, podemos concluir que las OTUs descritas como componentes más abundantes de las colonias son claramente características de las colonias y son poco abundantes en muestras de suelo y soporte. Sin embargo, su presencia en muestras de soporte no descarta que las colonias puedan formarse a partir de componentes del suelo o el soporte combinados en las proporciones adecuadas que pueden proliferar cuando las condiciones sean favorables para ello.

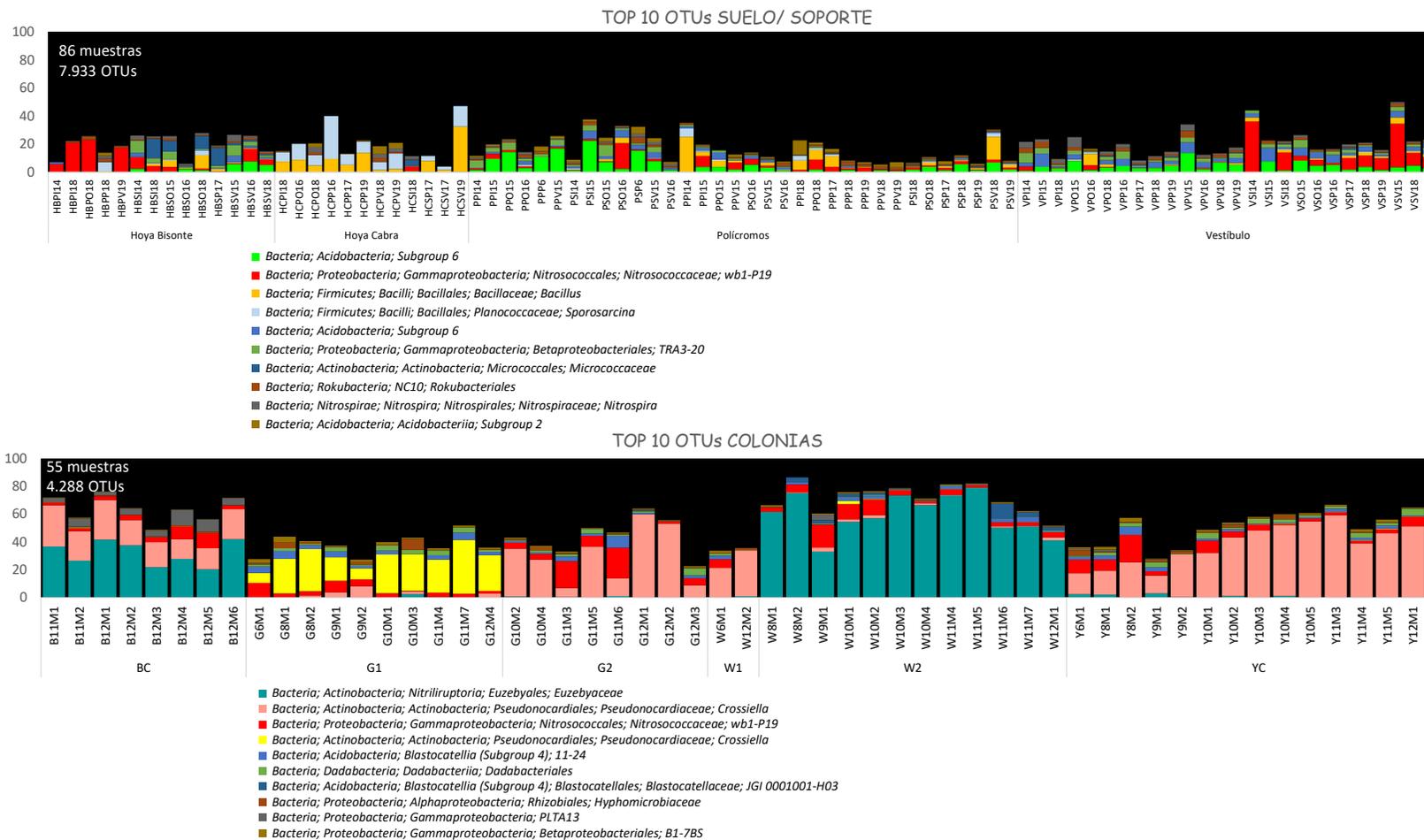


Figura 36. Lista de las 10 OTUs con mayor frecuencia en las muestras de suelo y soporte en comparación con las muestras de colonias visibles en la cueva de Altamira analizadas por la técnica de amplicones.

Colonia subtipo	Muestra	Lectura	Lecturas trimadas	Nº Total contigs	Total (Mb)	Nº contigs >1 kb	Total >1 Kb (Mb)	Nº MAGs	Contigs MAGS	MAGS (Mb)
G1	G10M1	150	54.066.040	1.068.638	582,10	63.430	188,20	34	31.364	112,03
G2	G10M2	150	53.795.588	818.733	484,40	59.974	183,30	27	27.040	108,77
W1	W6M1	100	62.611.472	159.816	108,70	16.444	48,70	9	9.152	31,86
W2	W8M1	100	63.607.924	82.108	62,10	8.767	34,80	6	5.091	24,30
YC	Y6M1	100	58.775.026	177.582	117,70	17.419	48,90	9	9.072	30,36
BC	B11M1	150	24.908.766	149.332	82,80	5.465	25,30	5	2.817	20,60

Tabla 2. Métrica general obtenida para cada muestra tras los pasos de ensamblado y bineado de las lecturas de los metagenomas.

## 5.2 Resultados técnica de shotgun

### 5.2.1 Métrica general

Por la técnica de shotgun se secuenciaron un total de 6 muestras, una muestra por cada subtipo de colonia identificado por la secuenciación de amplicones. Las primeras muestras (YC, W1 y W2) se secuenciaron con lecturas pareadas de 100 pb, y las siguientes (G1, G2 y BC) con lecturas de 150 pb. El total de lecturas obtenidas en los metagenomas fueron alrededor de 50 millones, exceptuando el metagenoma de la muestra de colonias beis que produjo solamente unos 25 millones.

Las lecturas obtenidas fueron filtradas y ensambladas *de novo* con metaSpades como se describe en materiales y métodos. En todos los metagenomas los contigs menores de 1000 pb no se tuvieron en cuenta, y el resto fueron utilizados para la obtención de MAGs (genomas derivados de metagenomas) tras su clasificación con Maxbin y posterior curado con Anvi'o. El software Anvi'o además de estimar la contaminación y la completitud con unos sets de genes marcadores propios que deberían estar presentes sin redundancia en cada taxón, genera clusters de contigs de un MAG en base a la frecuencia de tetranucleótidos y los representa gráficamente generando un árbol, de manera que los contigs que se consideran contaminantes se visualizan fácilmente y es posible desecharlos del MAG.

Respecto al tamaño de los metagenomas obtenidos (Tabla 2), hay que mencionar que un porcentaje muy importante de las secuencias ensambladas están en contigs muy pequeños y con una "cobertura según metaSpades" muy baja, lo que refleja que provienen de moléculas de DNA poco abundantes. Por eso hemos desestimado contigs menores de 1 kb, lo que ha supuesto eliminar entre un 50% y un 70% de secuencia según los metagenomas.

Adicionalmente, no todos los contigs mayores de 1 kb fueron correctamente clasificados por Maxbin. De un 15 a un 30% de secuencia en contigs mayores de 1Kb no ha sido clasificada en "bins" y por lo tanto no está representada en los MAGs. En resumen, solo un 20-40% de la secuencia ensamblada se

encuentra representada en los MAGs, aunque esto incluye todas las secuencias de buena calidad y presumiblemente, la mayoría de los genes codificantes que o bien están contenidos en los MAGs o al menos en los contigs no clasificados mayores de 1 kb.

Si estimamos que el tamaño medio de un genoma es de 5 Mb, el número de genomas recuperados como MAGs en cada metagenoma oscila entre 5 y 34, dato que coincide bien con el número de componentes significativos de las colonias obtenido a partir del análisis de amplicones.

Para analizar la distribución de las lecturas originales (reads) en los MAGs finales, hemos mapeado las lecturas en los contigs de los MAGs. De este modo tendremos una idea del porcentaje absoluto de información que se estaría perdiendo en las secuencias no clasificadas, y al mismo tiempo podremos conocer la abundancia relativa de cada MAG dentro de su metagenoma.

Por tanto, se alinearon las lecturas contra los MAGs utilizando el software Bowtie2 y los reads correctamente alineados se contaron con SAMtools, descrito en el apartado de materiales y métodos (Figura 37). El porcentaje de reads mapeados en los metagenomas YC, G1 y G2 fue alrededor de 55%, mientras que fue superior en los metagenomas de las colonias W1 y W2 con 62,79% y 71,50% respectivamente. El mayor porcentaje de mapeo se obtuvo en el metagenoma BC con un alineamiento del 84,53% de las lecturas con los contigs contenidos en MAGs. Esta diferencia puede ser debida a que el grado de diversidad que encontramos en las colonias beis, así como en las colonias blancas no es tan alta como en las otras colonias, y permite que la recuperación de los genomas de los componentes sea más completa y se pierda menos información.

No todos los contigs >1000 se binaron en MAGs y puesto que éstos también tienen información sobre los genes presentes en los metagenomas, estos contigs se agruparon en un mismo archivo (al que llamamos MAG0XX siendo XX el subtipo de colonia) al cuál se mapearon también las lecturas. A excepción del metagenoma G2 en el cual las lecturas mapeadas aumentaron

un 10% al incluir en el metagenoma los contigs >1000 no bineados, en el resto de metagenomas no se observó un incremento significativo, indicando que la mayor parte de información genómica de los componentes de las colonias está siendo recogida en los (bins) MAGs. A parte, siempre hay lecturas que no mapean y que corresponden a contigs <1000 o lecturas con errores que haya descartado el ensamblador. La proporción de estas lecturas no mapeadas en los metagenomas más diversos fue alta (YC:42,3%; G2:34,5%; G1:37,5% y W1:35,5%), mientras que este porcentaje disminuye en los metagenomas menos diversos (BC:23,8% y W2:14,2%), indicando igualmente el menor grado de contaminación de componentes menos esenciales de las colonias para las muestras menos diversas.

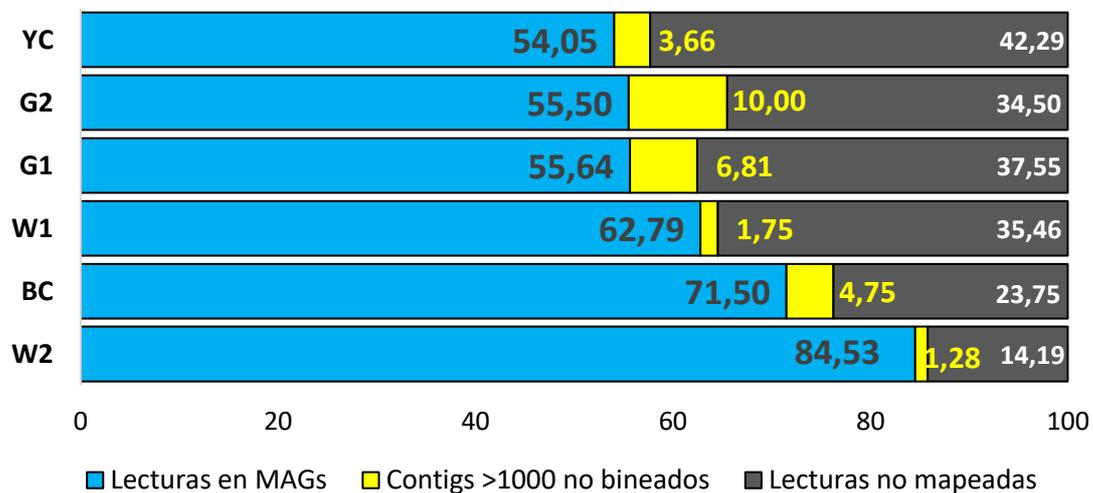


Figura 37. Mapabilidad de las lecturas para cada metagenoma. En color azul el porcentaje de las lecturas totales que han mapeado contra los MAGs obtenidos y en amarillo contra los contigs >1000 no bineados. En negro el porcentaje de las lecturas no mapeadas.

A partir de los datos de alineamiento, también es posible contar el número de lecturas que mapean en cada uno de los MAGs de forma individual. Los valores de esta medida se representan en la figura 38. En cada metagenoma se han repartido las lecturas que mapearon contra todos los contigs mayores de 1 kb (contenidos en MAGs o no) y calculado el porcentaje que mapea en cada MAG o no asignados a MAGs (en amarillo). Como puede observarse, en cada metagenoma existe un MAG en el que mapearon la mayoría de las lecturas (desde el 26 al 95% de las lecturas totales), indicando una abundancia

de este genoma en la comunidad mucho mayor que la del resto. Esta observación confirma la idea, propuesta a partir del análisis de amplicones, de la existencia de ciertos componentes más abundantes y que consideramos como componentes mayoritarios en las colonias. En el caso del metagenoma de las colonias beis observamos la presencia de dos MAGs con una abundancia relativa superior al resto, lo que también se desprende del análisis de amplicones.

También se observó que la abundancia de los CMs en los metagenomas de G1 y G2 no fue tan abundante en comparación al resto de metagenomas, aunque, por el contrario, sí lo fue si lo comparamos frente al resto de los MAGs obtenidos en esos metagenomas. Esta diferencia entre los metagenomas de las colonias grises y el resto puede deberse a una mayor complejidad en la composición de estos metagenomas. Una consecuencia directa de la mayor complejidad es una menor eficiencia en el proceso de clasificación de los contigs en MAGs y a su vez una menor proporción de lecturas que mapean con los componentes mayoritarios.

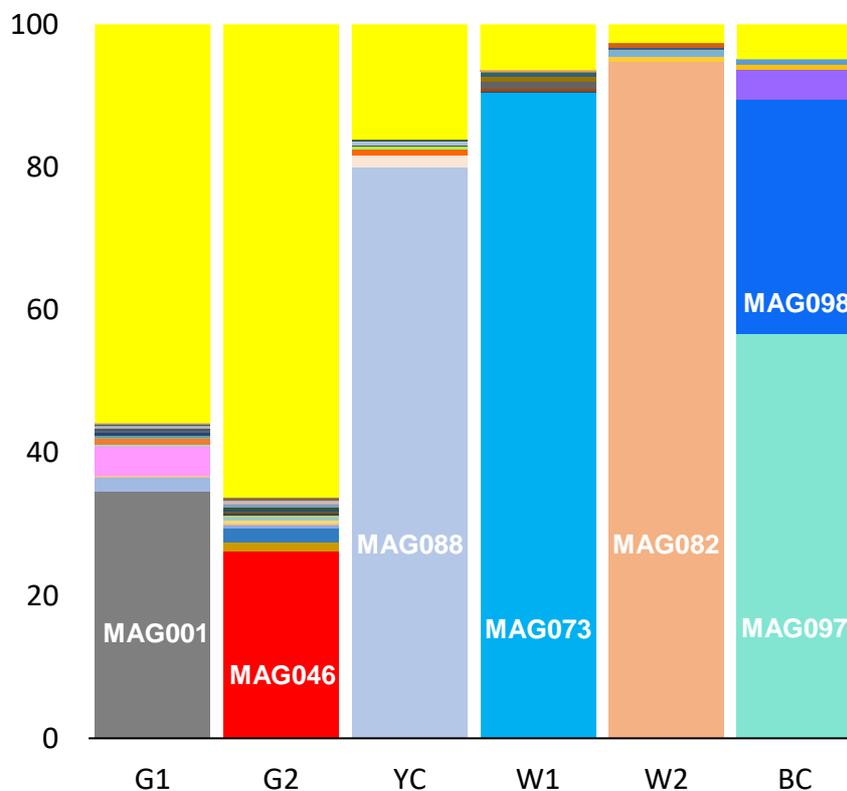


Figura 38. Abundancia relativa de los MAGs en cada metagenoma. El color amarillo representa la abundancia relativa (%) del archivo con los contigs >1000 no bineados.

### Análisis de la calidad de los MAGs

De los seis metagenomas, se obtuvieron un total de 90 MAGs que se depuraron para eliminar la contaminación resultante. Para medir la calidad de los MAGs obtenidos, es necesario tener una idea de la contaminación presente provenientes de otras bacterias y cómo de completo se ha recuperado su genoma porque se podría dar el caso de la obtención de artefactos como genomas quiméricos, etc... que no reflejan la realidad de la muestra. Este control de calidad se llevó a cabo con el programa CheckM. Hay que considerar también que estos programas sólo estiman la contaminación en base a la presencia de modelos de genes que deberían ser únicos en cada taxón, por tanto, los contigs contaminantes sin estos genes no se detectarían. Según los valores de contaminación e integridad (completitud) en los MAGs, los hemos clasificado en tres grupos: MAGs de calidad alta (>90% completitud y <5% contaminación), MAGs de calidad media (90-50% completitud y 5-15% contaminación) y MAGs de baja calidad (<50 completitud y 5-15% contaminación). De acuerdo con este criterio, de los 90 MAGs obtenidos, 16 fueron de alta calidad, 48 de media calidad y 26 de baja calidad (Figura 39).

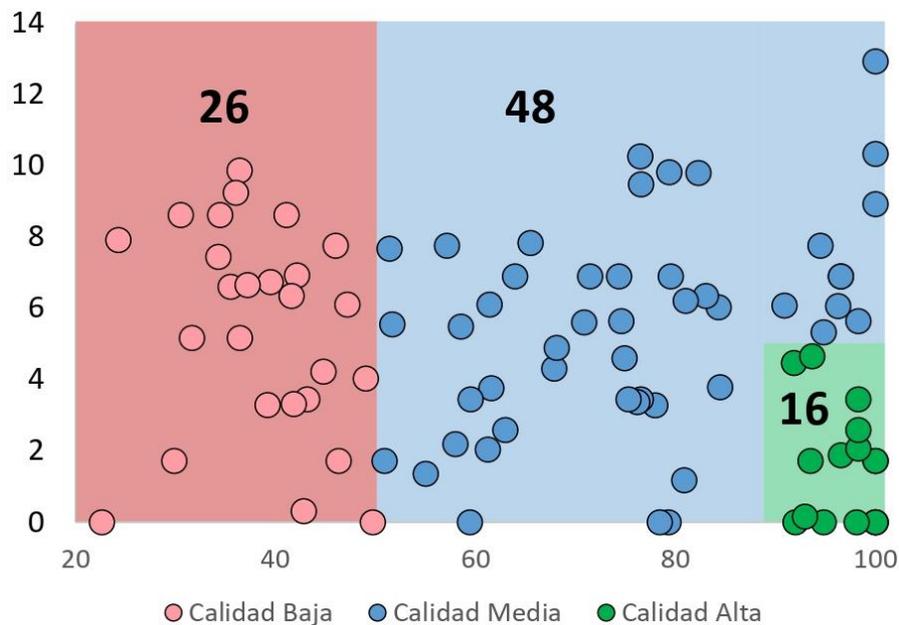


Figura 39. Representación de la calidad en los MAGs. Alta calidad (>90 completitud y <5% contaminación), calidad media (90-50 completitud y 5-15% contaminación) y baja calidad (<50 completitud y >15% contaminación).

En cuanto a su posición taxonómica, 2 MAGs correspondieron con genomas de arqueas (MAG010 y MAG074), mientras que el resto fueron bacterianos. Los genomas de arqueas presentaron un tamaño de genoma de  $\leq 2$  Mb y un contenido en GC bajo de  $< 0,45$  (Anexo-Tabla S5). Por lo general, los borradores de genomas obtenidos tienen un número alto de contigs ( $> 500$ ). El tamaño del genoma de los MAGs varía de 0,9 a 11 megabases, situándose la mayoría entre 2 a 5 Mb. Tienen un contenido alto de GC (0,5-0,7), a excepción de los genomas de arqueas.

El número de genomas que se han obtenido en los metagenomas no ha sido uniforme, ya que de las colonias grises (G1 y G2) hemos recuperado 34 y 27 MAGs, un número de genomas significativamente mayor que del resto. El número de MAGs recuperados del resto de colonias han sido 9 para las colonias amarillas y las blancas subtipo W1, 6 para las blancas subtipo W2 y 5 para las beis.

### **5.2.2 Análisis de la biodiversidad**

Para estimar la cantidad de diversidad que estos metagenomas están captando de las muestras originales, llevamos a cabo análisis equivalentes a las estimaciones de  $\alpha$  y  $\beta$  diversidad de manera similar a como se realizaron para los análisis de amplicones. Las curvas de rarefacción y medidas de diversidad  $\alpha$  se calcularon con el programa Nonpareil 3.

A diferencia de las curvas de rarefacción obtenidas para las muestras de amplicones, las curvas obtenidas de las muestras de metagenomas indicaron que en ninguna de estas muestras se llegó a caracterizar la diversidad total en ellas (Figura 40). Con las secuencias obtenidas en las muestras de metagenomas, se alcanzó más del 78% de la diversidad en todas las muestras con valor máximo del 91% para las colonias beis. Se estima que una representación satisfactoria de una muestra se consigue con valores superiores al 95%, para lo que se requeriría un esfuerzo de secuenciación adicional (Tabla 3).

No obstante, no podemos olvidar que las muestras secuenciadas no contienen “colonias puras”, sino que inevitablemente están contaminadas con una cantidad variable de soporte que aporta una cantidad de DNA no deseado a la muestra. Si consideramos, como ya hemos hecho antes, que estos contaminantes son poco abundantes en comparación con los componentes reales de las colonias podemos aceptar que el esfuerzo de secuenciación aplicado debe ser suficiente para describir la diversidad de las colonias.

El índice  $N_d$  que se obtiene en los metagenomas es un índice de diversidad de secuencias y se correlaciona positivamente con el índice de diversidad de Shannon obtenido en estas mismas muestras a partir de los análisis de amplicones (Anexo-Figura S3). Según los documentos del programa un entorno dominado en gran parte por unas pocas especies bacterianas como lo es el drenaje ácido de una mina, produce con Nonpareil valores de  $N_d$  alrededor de 15 a 16, suelos agrícolas, fondos de agua dulce y arenosos alrededor de 20 a 22, y suelos marinos y otros alrededor de 21 a 25. Las muestras en este estudio tienen un índice entre 16-19, indicando que se trata de muestras con una diversidad no muy compleja.

Muestras	Cobertura	Pearson's R	Esfuerzo de secuenciación del estudio	Esfuerzo calculado para una cobertura casi completa	Indice $N_d$
<b>G1</b>	0,6834861	0,9916706	5,41E+07	1,94E+11	18,97797
<b>G2</b>	0,8192145	0,992743	5,38E+07	1,05E+11	18,64075
<b>W1</b>	0,7834019	0,9689683	6,26E+07	3,43E+10	17,48001
<b>W2</b>	0,8979823	0,9869165	6,36E+07	4,38E+08	15,88844
<b>YC</b>	0,7346665	0,9738803	5,88E+07	1,36E+11	18,16294
<b>BC</b>	0,915041	0,9941717	2,49E+07	1,47E+09	16,8138

Tabla 3. Resultado de los análisis de  $\alpha$ -Diversidad en los metagenomas de colonias visibles.

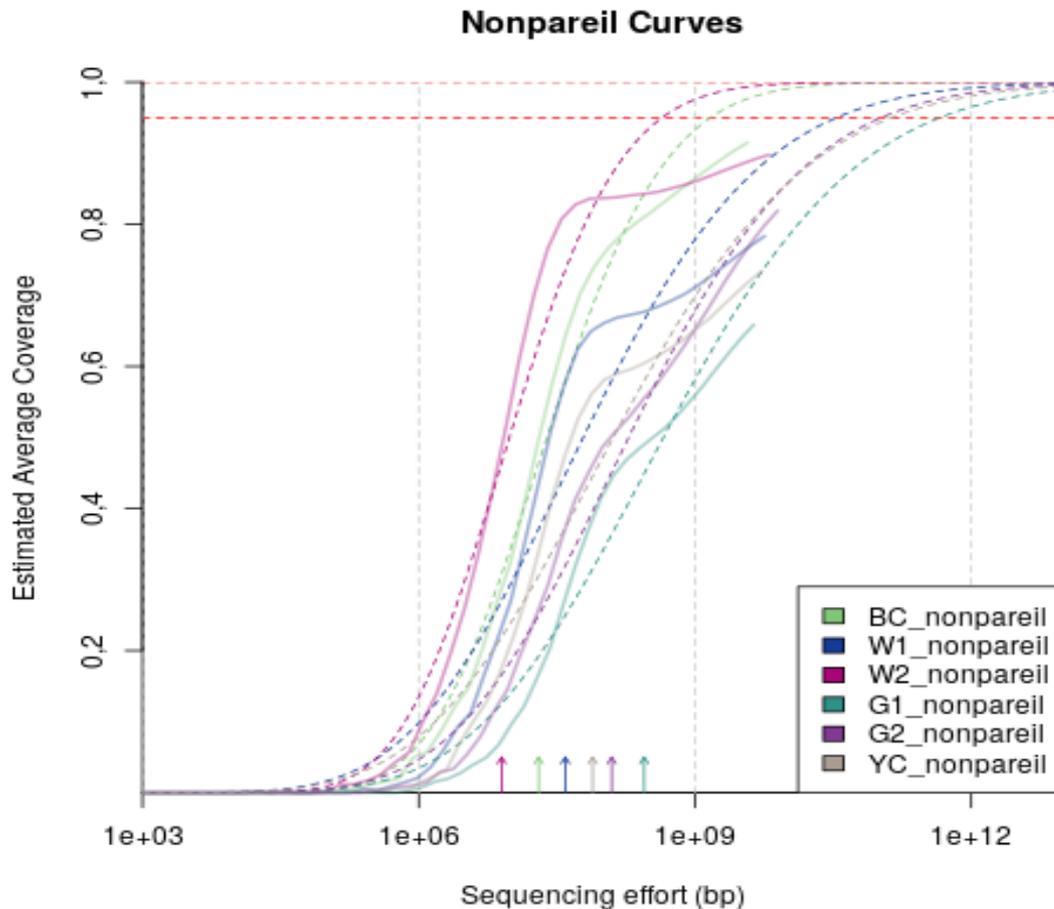


Figura 40. Curvas de rarefacción en los metagenomas de colonias visibles.

Se comparó también el grado de similitud entre los diferentes metagenomas (Figura 41). Para ello se utilizó el programa MetaFast. Como se puede observar, los metagenomas son bastante diferentes en general, en especial, los metagenomas BC, W2 y G1 que se pueden apreciar como los más distantes de todos. La separación en estas muestras ya se pudo apreciar por la técnica de amplicones en los análisis de  $\beta$  diversidad, que también las agrupó en clusters distintos y muy separados (Figura 31). Respecto a las muestras W1, YC y G2, por la técnica de amplicones fueron agrupadas en el mismo cluster, sin embargo, solamente los metagenomas de colonias W1 e YC fueron los únicos que agruparon más cercanos manteniendo una cierta similitud en su composición, mientras que el metagenoma G2 se sitúa equidistante del G1 y del grupo que forma W1 e YC.

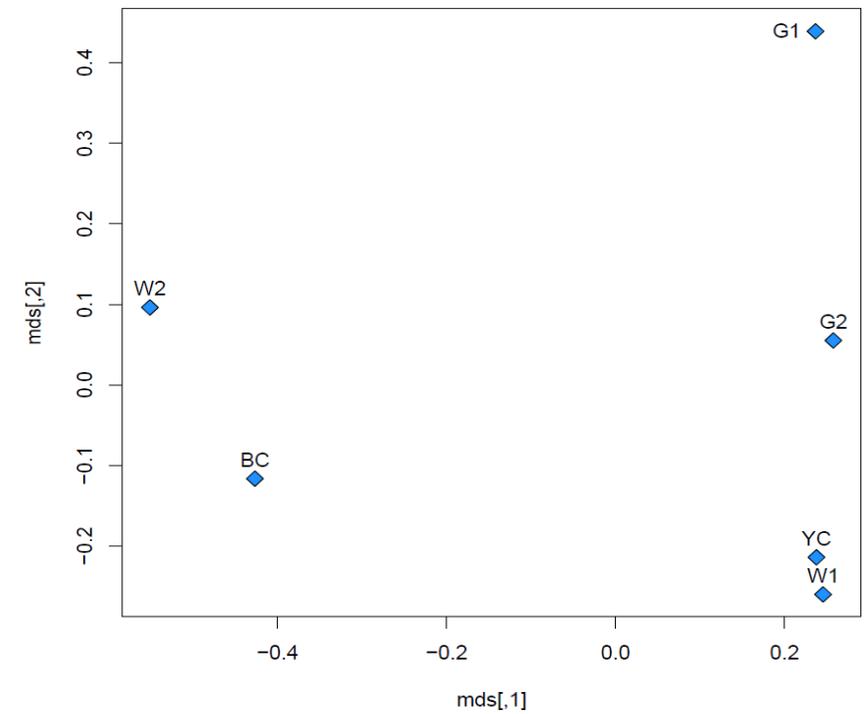
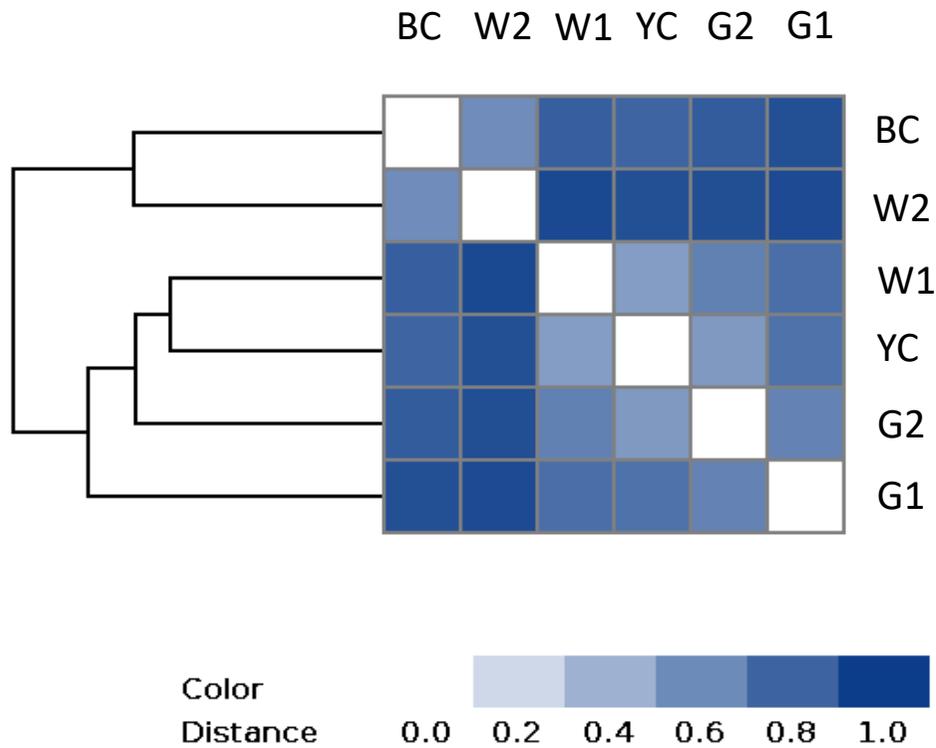


Figura 41.  $\beta$ -Diversidad presente en los metagenomas de colonias visibles. A la izquierda se representan las distancias estimadas entre los metagenomas obtenidas con el programa Metafast en un heatmap y a la derecha se representan estas distancias en un plot MDS con la función vegdist de la librería vegan v2.4-2 usando el método de disimilitud Bray-Curtis.

### 5.2.3 Análisis taxonómico de los MAGs

Los *phyla* en los que se clasificaron los MAGs recuperados en las colonias fueron *Acidobacteria*, *Actinobacteria*, *Candidate division NC10*, *Candidatus Dadabacteria*, *Candidatus Rokubacteria*, *Chloroflexi*, *Parcubacteria*, *Proteobacteria* (*alpha*, *beta*, *gamma* y *delta*), *Nitrospirae* y *Thaumarchaeota*. Estos taxones fueron los mismos detectados en la selección de las muestras de colonia por la técnica de amplicones, a excepción de los filos *Candidate division NC10* y *Thaumarchaeota* que solo fueron detectados en OTUs minoritarias (<1%).

El único taxón no detectado por la técnica de amplicones fue el superfilo *Parcubacteria*, lo que se atribuye a la presencia de intrones en el 16S rRNA de este grupo, impidiendo su amplificación con los primers universales usados en la técnica de amplicones.

Cuando se compararon las secuencias de los MAGs calculando los valores de ANI y AAI entre ellos, encontramos que algunos MAGs, se encuentran compartidos entre los diferentes metagenomas y que podrían pertenecer a la misma especie. Por ello como paso previo a asignar taxonomía a los MAGs, contando con la posibilidad de que algunos estuvieran compartidos entre diferentes metagenomas o que pudieran ser muy similares entre si, hemos calculado la similitud entre ellos comparando los MAGs completos entre si y calculando el grado de identidad de secuencia de nucleótidos (ANI) y de los polipéptidos predichos por Prodigal (AAI). Esta etapa sería equivalente a la de agrupación de las secuencias de amplicones en OTUs.

En estudios anteriores se ha establecido que los valores de ANI requeridos para que dos MAGs pertenezcan a bacterias del mismo género deben ser >77%, y mayor del 93-96% para considerarlos pertenecientes a la misma especie. (Rosselló-Móra and Amann, 2015). Otros estudios establecen el límite para pertenecer al mismo género a partir del 62% de identidad y para la misma especie mayor del 95-96% (Kim et al., 2014).

En cuanto a los valores de AAI, los valores requeridos son parecidos a los de ANI, debiendo ser para pertenecer a la misma especie mayor del 93-95% de identidad y para pertenecer al mismo género mayor del 65-72% (Konstantinidis and Tiedje, 2007).

Siguiendo estos valores establecidos, hemos agrupado los MAGs en clusters utilizando un valor de ANI superior al 77% para considerar a los MAGs que agrupen dentro de estos clusters como miembros del mismo género. Esta agrupación en los MAGs obtenidos ha resultado en 13 clusters (Anexo-Tabla S6). Además, la secuencia de nucleótidos de los MAGs (valor ANI) se ha comparado contra los genomas de referencia disponibles en GenBank para el taxón determinado en ese cluster. Adicionalmente a la comparación de los MAGs completos con genomas obtenidos de las bases de datos, hemos utilizado la comparación de rRNAs ribosomales cuando se han encontrado presentes en los MAGs del cluster y la comparación de los ORFs predichos a partir de la secuencia de nucleótidos. Los genomas de referencia y su número de acceso a GenBank utilizados en estos análisis se encuentran en el Anexo-Tabla S7. Combinando las aproximaciones anteriores proponemos para los grupos de MAGs obtenidos en los 13 clusters, la siguiente clasificación:

Cluster I	<i>Nitrosopumilaceae (Thaumarchaeota)</i>
Cluster II	<i>Blastocatellia (Acidobacteria)</i>
Cluster III	<i>Pseudonocardiaceae (Actinobacteria)</i>
Cluster IV	<i>Nitriliruptoria (Actinobacteria)</i>
Cluster V	<i>Rhizobiales (Aphaproteobacteria)</i>
Cluster VI	<i>Ferrovaceae (Betaproteobacteria)</i>
Cluster VII	<i>Betaproteobacteria</i>
Cluster VIII	<i>Lysobacter (Gammaproteobacteria)</i>
Cluster IX	<i>Chromatiales (Gammaproteobacteria)</i>
Cluster X	<i>Nitrospira (Nitrospirae)</i>
Cluster XI	Candidate division NC10
Cluster XII	<i>Parcubacteria</i>
Cluster XIII	<i>Candidatus Dadabacteria</i>

A continuación, pasamos a describir los clusters mencionados mediante el análisis recogido en base a las 3 aproximaciones descritas previamente:

### **Cluster I**

En este cluster agruparon los únicos genomas de arqueas obtenidos: MAG010 (G1) y MAG074 (W1). Según los valores de ANI de 96,51% y de AAI 95,64%, estas arqueas, obtenidas de dos metagenomas diferentes, corresponderían a dos individuos de la misma especie. En uno de ellos (MAG074) hemos encontrado dos fragmentos conteniendo secuencias parciales del 16S rRNA (728/692 bp). Por comparación de este 16S rRNA podemos determinar que los MAGs representan a dos arqueas del filo *Thaumarchaeota*, clase *Nitrososphaeria*, orden *Nitrosopumilales* y familia *Nitrosopumilaceae*. Según la clasificación de sus ORFs con CAT/BAT, clasificaron en el filo *Thaumarchaeota* y la arquea más cercana a estos MAGs fue *Thaumarchaeota archaeon CSP1-1* (GenBank accession: GCA\_001443365.1), que se encuentra como una especie no clasificada dentro del filo *Thaumarchaeota*, probablemente porque el borrador de genoma no contiene ningún rRNA ribosómico.

Hay en GenBank 502 ensamblados bajo el filo *Thaumarchaeota*, 150 ensamblados bajo el orden *Nitrosopumilales* y 130 ensamblados bajo la familia *Nitrosopumilaceae*. Para afinar la asignación taxonómica del cluster se descargaron algunos genomas de referencia de los 3 órdenes descritos en la familia *Nitrosopumilaceae* (*Candidatus Nitrosomarinus*, *Nitrosarchaeum* y *Nitrosopumilus*), además del genoma de *Thaumarchaeota archaeon CSP1-1*. El valor de ANI del cluster contra estos genomas seleccionados de los diferentes géneros fue sólo alto contra el genoma *Thaumarchaeota archaeon CSP1-1* (>94%), por el contrario, fue bajo respecto al resto (71%), lo que sugiere que los MAGs en este cluster, junto con *Thaumarchaeota archaeon CSP1-1*, estarían en el límite de poder pertenecer cuando menos a un género no descrito, diferente del resto de la familia *Nitrosopumilaceae* presente en las bases de datos (Figura 42).

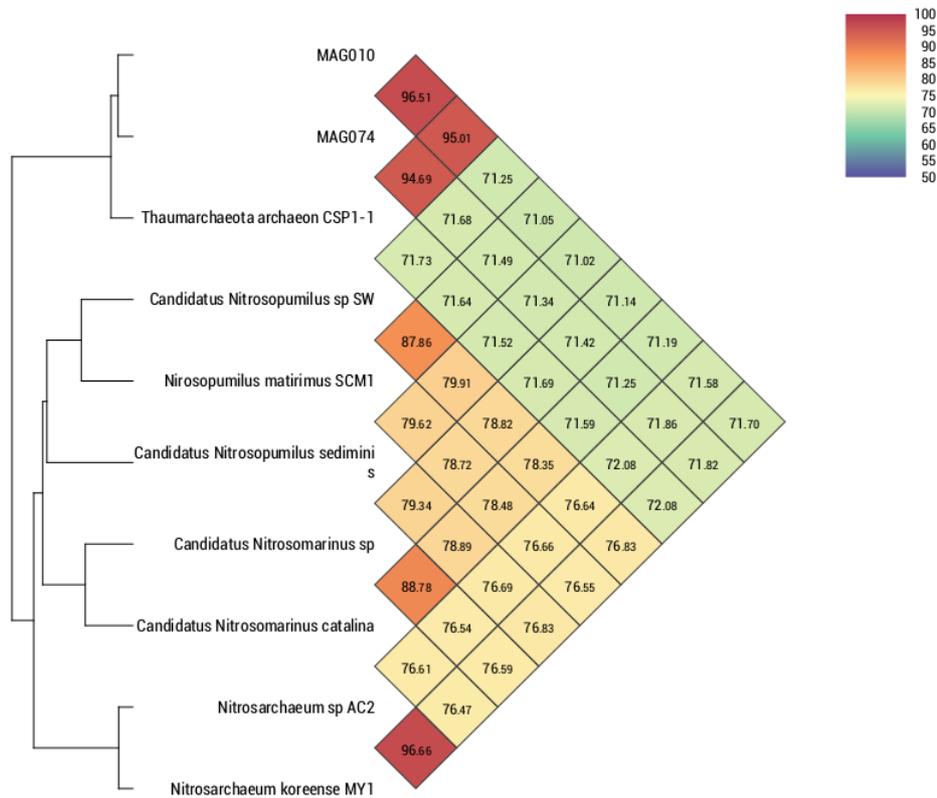


Figura 42. Valores ANI para los MAGs que agruparon en el cluster I contra los genomas disponibles en la familia *Nitrosopumilaceae*.

Al comparar las secuencias de 16S rRNA encontradas en el MAG074 contra la base de datos nr/nt (nucleotide collection) de GenBank (26/12/2020), los hits más cercanos fueron dos clones de rRNA (GenBank: AM055706.1 y KP308700.1) que mostraban porcentajes de identidad del 99,7% y 98,5% con MAG074. Estos 16S rRNA fueron clones obtenidos en un análisis de comunidades microbianas en manantiales sulfurosos y sedimentos de acuíferos y representan arqueas no cultivadas. (Castelle et al., 2015; Koch et al., 2006).

### **Cluster II**

En este cluster han agrupado 3 genomas de bacterias: MAG081 (W1), MAG004 (G1) y MAG061 (G2) con valores de ANI >94,49% y de AAI >90,09%, sugiriendo que podrían pertenecer a diferentes aislados de una misma especie. En los MAGs 81, 04 y 61 han bineado 3 fragmentos conteniendo una

secuencia parcial (779 pb) o completa del 16S rRNA (1529/1492 pb respectivamente). Basándonos en los alineamientos de estos rRNAs en SILVA, podemos determinar que los MAGs 81, 04 y 61 correspondieron a Bacterias del filo *Acidobacteria*, clase *Blastocatellia* (*Subgroup4*) y orden 11-24. Según la clasificación de sus ORFs con CAT/BAT, los tres MAGs del cluster clasificaron en el filo *Acidobacteria* sin ser posible una mejor asignación.

Hay en GenBank 76 ensamblados genómicos en la clase *Blastocatellia*, con dos órdenes, *Blastocatellales* y *Chloracidobacteriales*. Seis de estos ensamblados genómicos fueron seleccionados tratando de que representaran la mayor diversidad posible dentro de la clase *Blastocatellia*, que no es mucha. Estos genomas se compararon contra los genomas de los tres MAGs de este cluster. El valor de ANI del cluster contra estos genomas de referencia seleccionados fue muy bajo (<68%) lo que sugiere que las bacterias detectadas estarían en el límite de poder pertenecer cuando menos a un género de la clase *Blastocatellia* diferente de los representados hasta la fecha en GenBank (Figura 43).

Al comparar las secuencias de 16S rRNA encontradas en estos MAGS contra la base de datos nr/nt de GenBank (26/12/2020) se encontró un clon de rRNA (GenBank: FJ535090.1) que mostraba porcentajes de identidad de 99,6% con MAG004 y MAG061, y de 99,2% con MAG081. Este 16S rRNA fue obtenido en un análisis de comunidades microbianas que colonizan una cueva kárstica en Eslovenia (Pasic et al FEMS, 2010). Podemos por tanto concluir, al menos de este análisis, que este cluster representa posiblemente un “género nuevo” de *Acidobacteria*, sin miembros cultivados, perteneciente a la clase *Blastocatellia* y especialmente adaptado al crecimiento en cuevas formando parte de consorcios.

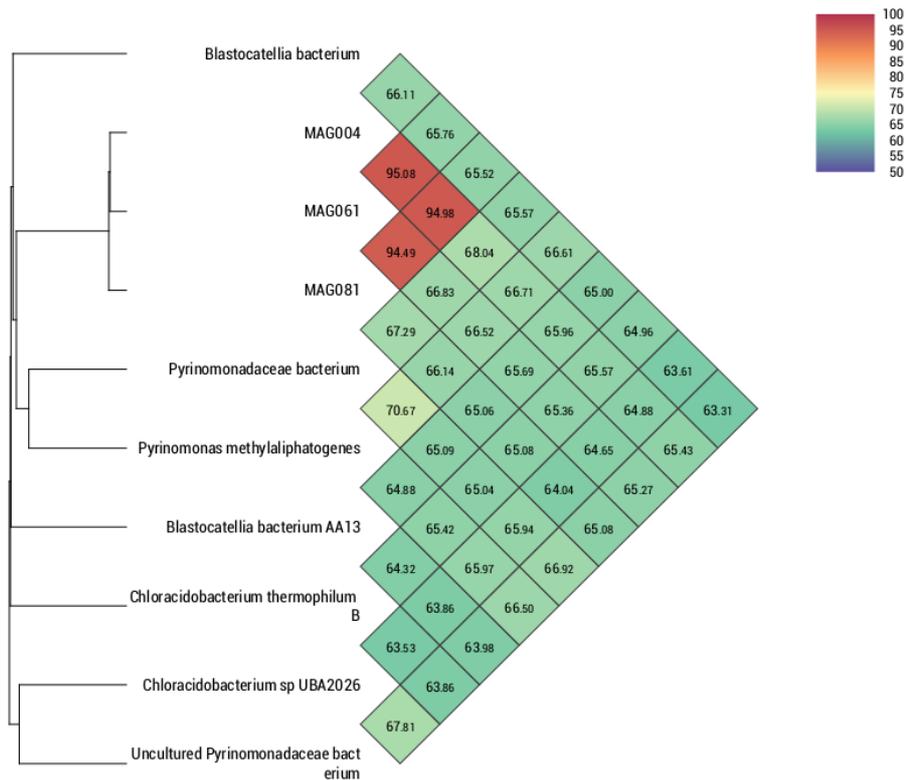


Figura 43. Valores ANI para los MAGs que agruparon en el cluster II contra los genomas disponibles de la clase *Blastocatellia*.

### Cluster III

En este cluster han agrupado la gran mayoría de los componentes mayoritarios de las colonias: MAG088 (YC), MAG046 (G2), MAG098 (BC), MAG001 (G1) y MAG073 (W1), y un componente secundario MAG060 (G2) con valores de ANI >77,10% y AAI >71,66%. MAG060, que aparece como componente secundario en las colonias grises subtipo G2, puede ser idéntico a MAG098 (ANI: 99,69% y AAI: 99,49%), por lo tanto, este cluster estaría formado prácticamente por componentes mayoritarios, añadiendo la posibilidad de que un componente puede ser mayoritario en un tipo de colonia y secundario en otro.

Según los valores de ANI (exceptuando MAG060) estarían en el intervalo 77,81-92,77% y de AAI 71,66-89,78%, por tanto, estas bacterias corresponderían a distintas especies, probablemente dentro de un mismo género.

En todos ellos, a excepción del MAG060, ha bineado un fragmento conteniendo secuencias enteras del 16S rRNA (1464/1434/1513/1455/1533) lo que nos permitió determinar por comparación con SILVA que estos MAGs correspondieron a Bacterias del filo *Actinobacteria*, clase *Actinobacteria*, orden *Pseudonocardiales* y familia *Pseudonocardiaceae*. Según la clasificación de sus ORFs con CAT/BAT, clasificaron en el orden *Pseudonocardiales*, y las bacterias más cercanas fueron bacterias no clasificadas dentro del orden *Pseudonocardiales* (Figura 44).

En GenBank bajo el orden *Pseudonocardiales* hay 421 ensamblados habiendo sólo una familia descrita, *Pseudonocardiaceae* con 409 ensamblados, y como *Pseudonocardiales* no clasificados 12 ensamblados, dónde se encuentra los hits más cercanos a los componentes agrupados en este cluster. En SILVA la nomenclatura es igual habiendo tan sólo la familia *Pseudonocardiaceae* dentro del orden *Pseudonocardiales*, sin embargo, se encuentra alguna diferencia a nivel de género que se discutirá más adelante. Teniendo en cuenta esto y al tratarse de un cluster formado prácticamente por componentes mayoritarios, una asignación más precisa en estas bacterias se discutirá en el apartado de la filogenia de los componentes mayoritarios.

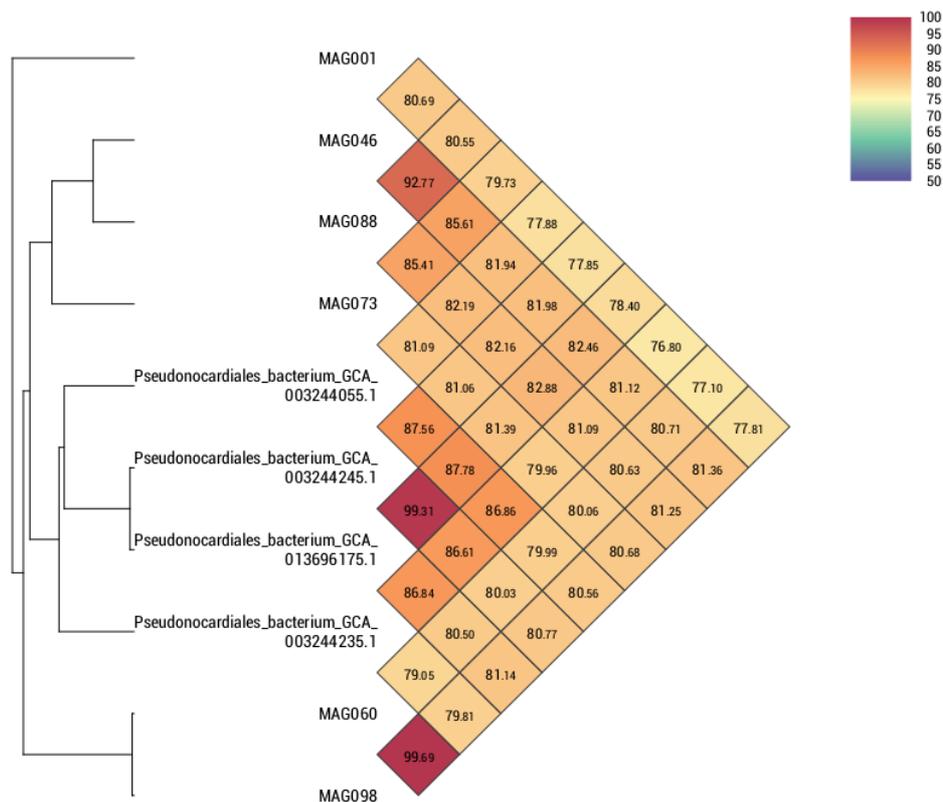


Figura 44. Valores ANI para los MAGs que agruparon en el cluster III contra los genomas más cercanos disponibles dentro del orden *Pseudonocardiales*.

#### **Cluster IV**

En este cluster se clasificaron el resto de componentes mayoritarios MAG097 (BC) y MAG082 (W2), y los secundarios MAG090 (YC) y MAG050 (G2) con valores de ANI >85,85% y de AAI >82,25%. MAG097 puede considerarse idéntico a los secundarios MAG090 en colonias amarillas y MAG050 en colonias grises subtipo G2 en base a los valores de AAI (>99,25%) y ANI (>99,59%). Por los valores de ANI entre MAG082 contra MAG097, 90 y 50 (85,85-87,40%) y de AAI (82,25–83,86%), estas bacterias pertenecerían a distintas especies y probablemente dentro de un mismo género.

En todos los MAGs han bineado fragmentos conteniendo secuencias completas (1458/1458/1437/1458) de 16S rRNA pudiendo a través de ellas determinar que los MAGs representaban bacterias del filo *Actinobacteria*, clase *Nitriliruptoria*, orden *Euzebyales* y familia *Euzebyaceae*. Según la clasificación de sus ORFs con CAT/BAT, clasificaron a nivel de clase *Actinobacteria*, sin ser posible una mejor asignación a mayor nivel e indicando por tanto baja identidad con los genomas disponibles, probablemente por la poca información que hay en la base de datos del NCBI en torno a la clase *Nitriliruptoria*.

En GenBank bajo la clase *Nitriliruptoria* hay 91 ensamblados habiendo 4 órdenes descritos que son *Egibacterales* (3 ensamblados), *Egicoccales* (2 ensamblados), *Euzebyales* (9 ensamblados) y *Nitriliruptorales* (73 ensamblados), y como no clasificados 4 ensamblados. Ninguno de estos ensamblados mostró suficiente similitud con los MAGs en base al valor ANI. Teniendo en cuenta esto y al tratarse de un cluster formado prácticamente por componentes mayoritarios, una asignación más precisa en estas bacterias se discutirá en el apartado de la filogenia de los componentes mayoritarios.

### Cluster V

En este cluster han agrupado el MAG094 (YC) y el MAG067 (G2) con un valor de ANI del 95,79% y de AAI del 93,21%. Acorde a estos valores estarían en el rango de poder considerarse miembros de la misma especie. Ninguno de estos MAGs presentó bineado algún fragmento conteniendo secuencias de 16S rRNA, y según la clasificación de sus ORFs con CAT/BAT clasificaron con el filo *Proteobacteria*, clase *Alphaproteobacteria* y orden *Rhizobiales*. Además, estos MAGs presentaron una calidad baja teniendo una completitud inferior al 50% y por ello resulta difícil una asignación taxonómica más completa.

### Cluster VI

En este cluster han agrupado el MAG049 (G2) y el MAG003 (G1) con un valor de ANI de 99,84% y un AAI de 99,81%. Según los valores de AAI y ANI, estas bacterias pueden considerarse como individuos casi idénticos de la misma especie. Habiendo bineado en los dos MAGs fragmentos conteniendo secuencias parciales (1251 pb) o completas (1546 pb) de 16S rRNA, podemos determinar con SILVA que los MAGs correspondieron a bacterias del filo *Proteobacteria*, la clase *Gammaproteobacteria*, orden *Betaproteobacteriales*, familia *Ferrovaceae* y género *Ferrovum*. Según la clasificación de sus ORFs con CAT/BAT, clasificaron con el filo *Proteobacteria*, sin ser posible una mejor asignación dado que estos MAGs presentan baja identidad con los genomas disponibles de la base de datos, probablemente por la poca información que hay en torno al género *Ferrovum*.

Hay en GenBank 87 ensamblados disponibles dentro de la familia *Ferrovaceae* en el que sólo hay descrito un género *Ferrovum*, y dentro de él sólo una especie descrita *Ferrovum myxofaciens* con 3 ensamblados, y como no clasificadas en este género 84 ensamblados. Hemos utilizado algunos de estos genomas para compararlos a nivel de nucleótido con los genomas dentro de este cluster (Figura 45). El valor de ANI del cluster con algunos de los genomas mencionados fue muy bajo (<66%) lo que sugiere que las bacterias detectadas

estarían en el límite de poder pertenecer cuando menos a un género diferente del género descrito *Ferrovum*.

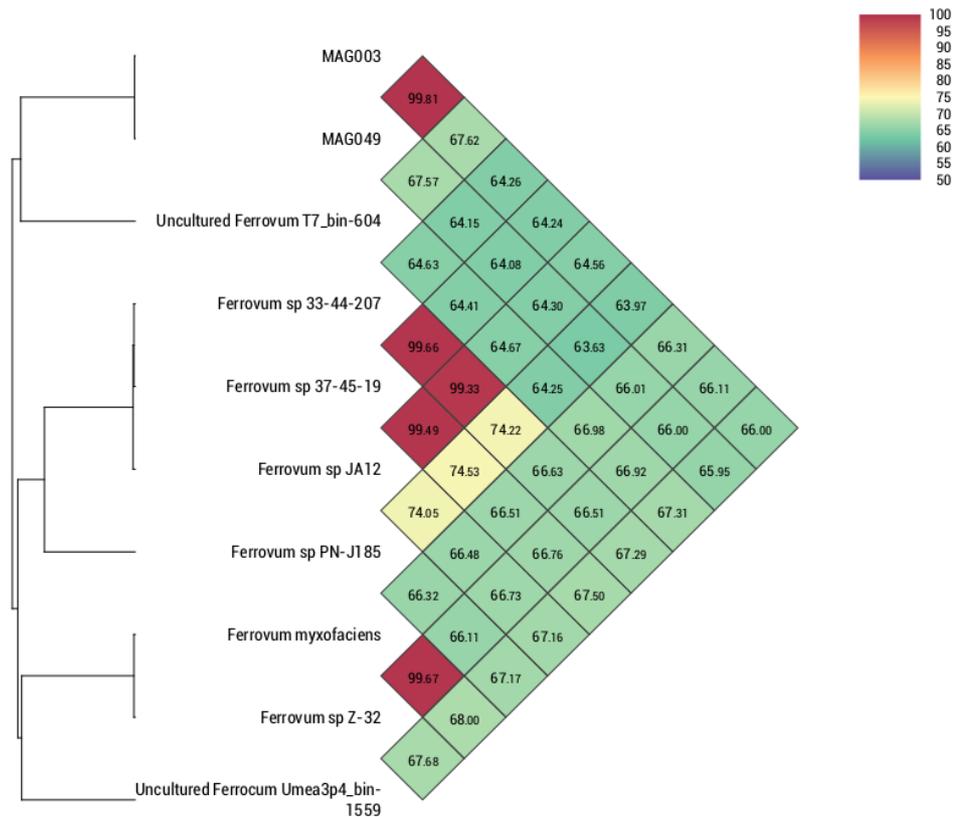


Figura 45. Valores ANI para los MAGs que agruparon en el cluster VI contra los genomas más cercanos disponibles dentro del género *Ferrovum*.

La comparación de las secuencias de 16S rRNA encontradas en estos MAGS contra la base de datos nr/nt de GenBank (26/12/2020) no mejoró su situación taxónomica. No se encontraron bacterias muy cercanas, lo que podría justificar la sospecha de un género diferente al descrito dentro de la familia *Ferrovaceae*. Los hits más cercanos fueron clones de rRNA obtenidos a partir de muestras no cultivadas obtenidos de muestras de sedimentos, roca y suelo (GenBank HQ445760.1 y EF020061.1) con porcentajes de identidad del 93% con MAG003 y MAG049.

### **Cluster VII**

En este cluster han agrupado el MAG086 (W2) y MAG005 (G1) con un valor de ANI del 78,99% y de AAI del 72,33%. Según estos valores de ANI y AAI, estas dos bacterias no serían de la misma especie, aunque posiblemente sean del mismo género. Ninguna de ellas presentó binnedo algún fragmento conteniendo secuencias de 16S rRNA, y según la clasificación de sus ORFs con CAT/BAT únicamente se clasificaron en la clase *Betaproteobacteria*, sin ser posible una mejor asignación.

### **Cluster VIII**

En este cluster han agrupado el MAG101 (BC) y el MAG058 (G2) con un valor de ANI 96,55% y un AAI 95,41%. Según estos valores de ANI y AAI, ambos genomas pertenecerían a bacterias de la misma especie. Ambos MAGs contienen contigs con secuencias enteras del 16S rRNA (1556/1556). Gracias a ello podemos determinar que los MAGs correspondieron a Bacterias del filo *Proteobacteria*, clase *Gammaproteobacteria*, orden *Xanthomonadales*, familia *Xanthomonadaceae*, género *Lysobacter* y especie *Lysobacter koreensis*. Según la clasificación de sus ORFs con CAT/BAT, clasificaron con la familia *Xanthomonadaceae*.

En GenBank hay 84 ensamblados de *Lysobacter*, aunque no hay uno disponible de la especie *Lysobacter koreensis*. El valor de ANI del cluster contra algunos de los genomas en el género *Lysobacter* fue >77%, lo que sugiere que las bacterias detectadas pertenecen a este mismo género (Figura 46).

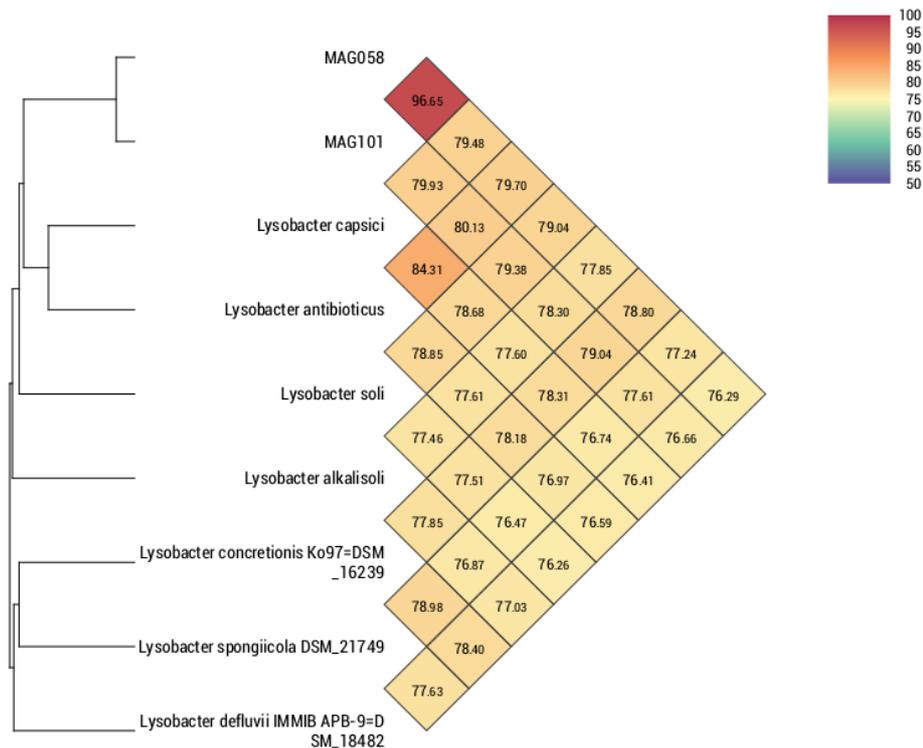


Figura 46. Valores ANI para los MAGs que agruparon en el cluster VIII contra los genomas más cercanos disponibles dentro del género *Lysobacter*.

### Cluster IX

En este cluster encontramos 4 MAGs: MAG047 (G2), MAG089 (YC), MAG083 (W2) y MAG078 (W1) con valores de ANI >91,30% y AAI >86,91%. Los MAGs de este cluster se agruparon a su vez de dos en dos. Así MAG047 (G2) y MAG089 (YC) presentaron entre si valores de ANI 99,61% y AAI 99,29%, y MAG083 (W2) y MAG078 (W1) presentaron ANI del 97,62% y AAI 96,15%. En todos los MAGs salvo en MAG078 han bineado fragmentos conteniendo secuencias enteras (1543/1543 pb) o parciales (1135 pb) del 16S rRNA. Podemos por tanto determinar que los MAGs correspondieron a Bacterias del filo *Proteobacteria*, clase *Gammaproteobacteria*, orden *Nitrosococcales*, familia *Nitrosococcaceae* y género *wb1-P19*. Aunque no se detecte 16S rRNA en el MAG078, encontramos un 16S rRNA idéntico a las secuencias de 16S rRNA anteriores en un contig del metagenoma W1 que no está bineado, por lo que atendiendo a los valores de AAI podemos concluir que MAG078 también corresponde a un miembro del mismo género. Según la clasificación de sus

ORFs con CAT/BAT, clasificaron con la clase *Gammaproteobacteria* y orden *Chromatiales*, y la bacteria más cercana a los componentes agrupados en este cluster fue *Chromatiales bacterium USCg\_Taylor*.

En GenBank hay 520 ensamblados bajo el orden *Chromatiales*, y el hit más cercano en base a la asignación de los ORFs fue el genoma *Chromatiales bacterium USCg\_Taylor*, GenBank assembly accession: GCA\_002007425.1, que se encuentra como no clasificado dentro de este orden. El valor de ANI del cluster contra el genoma *Chromatiales bacterium USCg\_Taylor* fue >92%, lo que sugiere que todas ellas podrían pertenecer cuando menos a un género diferente en este mismo orden *Chromatiales* (Figura 47).

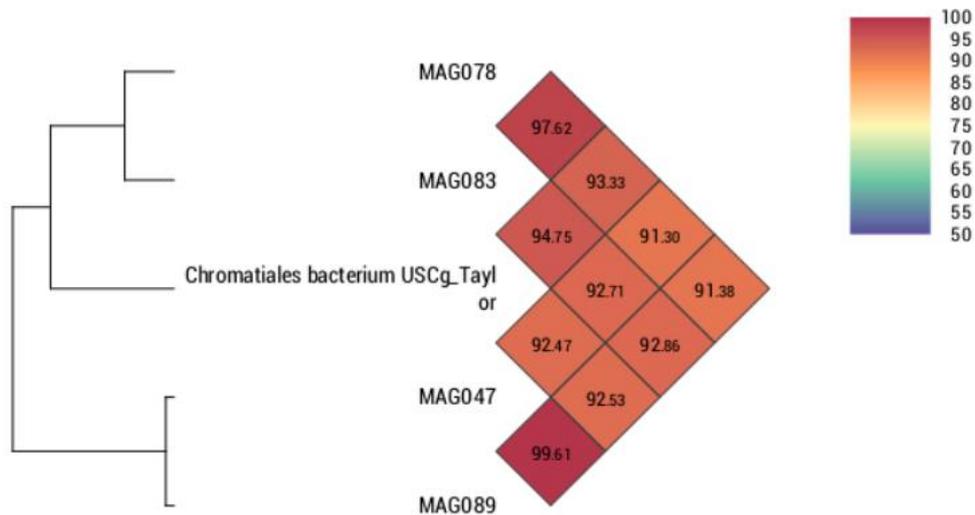


Figura 47. Valores ANI para los MAGs que agruparon en el cluster IX contra el genoma de referencia más cercano (*Chromatiales bacterium USCg\_Taylor*).

### Cluster X

En este cluster encontramos el MAG077 (W1) y el MAG062 (G2) con un valor de ANI del 92,31% y de AAI del 91,60%. Según estos valores, estas bacterias corresponderían a distintas especies de un mismo género. En el MAG077 se encuentra un fragmento conteniendo una secuencia parcial de 16S rRNA (590 pb) que clasificó con el filo *Rokubacteria* y clase *NC10*, y en el MAG062, dos fragmentos conteniendo secuencias parciales (837/675) que correspondieron con el filo *Nitrospirae*, clase *Nitrospira*, orden *Nitrospirales* y familia

*Nitrospiraceae*. Según la clasificación de sus ORFs con CAT/BAT, clasificaron con el filo *Nitrospirae* sin ser posible una mejor asignación, con lo que podemos determinar finalmente que se trata de bacterias que clasificaron con el filo *Nitrospira*, y, por tanto, el 16S rRNA correspondiente al filo *Rokubacteria* se encuentra mal bineado en el MAG077.

En GenBank hay 130 ensamblados bajo el género *Nitrospira* y hemos comparado algunos de estos ensamblados dentro de este género con los componentes de este cluster. El valor de ANI del cluster contra algunos de estos genomas fue 71%, al igual que el valor que se comparte entre las distintas especies descritas, las cuales todas pertenecen al mismo género, lo que sugiere que los MAGs detectados para este cluster pertenecen probablemente a una especie diferente de las descritas utilizadas al menos en este análisis (Figura 48).

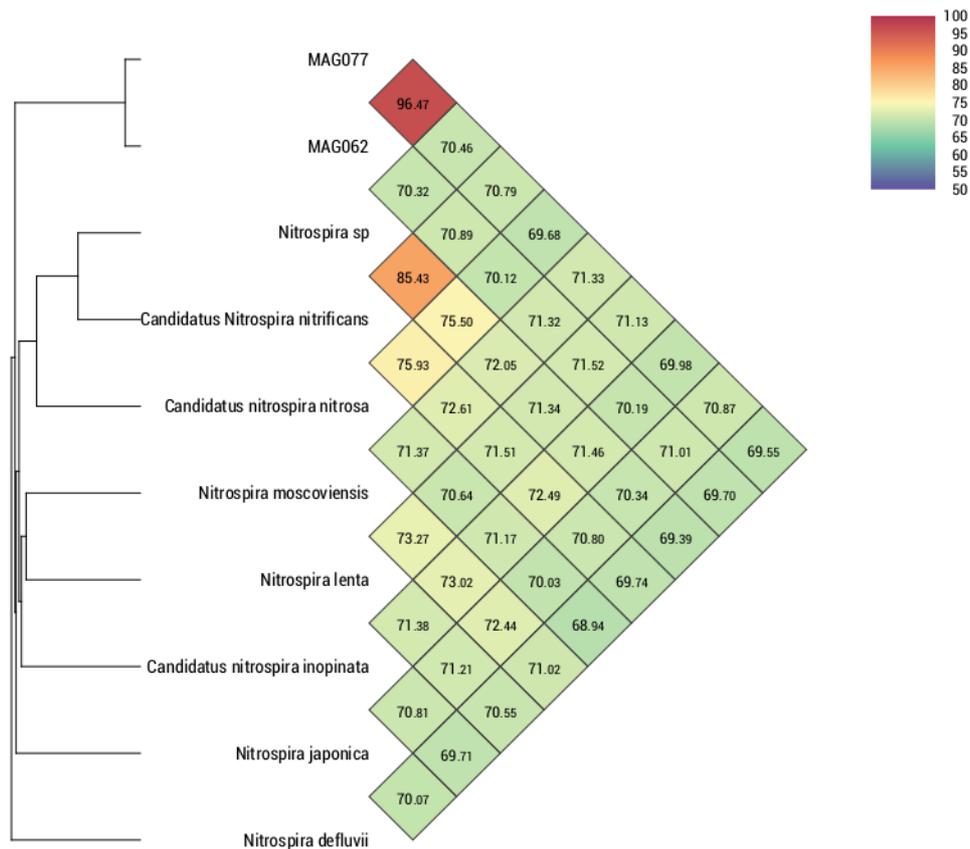


Figura 48. Valores ANI para los MAGs que agruparon en el cluster X contra los genomas más cercanos disponibles dentro del género *Nitrospira*.

## Cluster XI

En este cluster encontramos el MAG017 (G1) y el MAG075 (W1) con un valor de ANI de 98,40% y un AAI de 96,90%. Según los valores de AAI y ANI, estas bacterias pertenecerían probablemente a la misma especie. En uno de estos MAGs se encuentra un contig conteniendo una secuencia parcial del 16S rRNA (208 pb). Con la precaución que impone el tamaño del fragmento, podemos aventurar que los MAGs correspondieron a bacterias del filo candidato *Rokubacteria*, clase NC10, orden *Methylomirabiales*, familia *Methylomirabilaceae*, género *wb1-A12* y la bacteria *Candidate division NC10 bacterium CSP1-5*. Según la clasificación de sus ORFs con CAT/BAT, clasificaron con el filo Candidate division NC10 y de nuevo con la bacteria Candidate division NC10 bacterium CSP1-5.

En GenBank hay 22 ensamblados dentro del filo Candidate division NC10, y sólo un género descrito *Candidatus Methylomirabilis* con 8 ensamblados. Los otros 14 borradores de genomas no están aún clasificados y entre ellos se encuentra el hit más cercano a los componentes agrupados en este cluster en base a la asignación de sus ORFs y su secuencia del 16S rRNA, el genoma NC10 bacterium CSP1-5, GenBank assembly accession: GCA\_001443495.1. En SILVA este filo candidato NC10 se clasifica como una clase dentro del filo *Rokubacteria*. El valor de ANI del cluster contra los genomas seleccionados fue alto con 2 bacterias, en especial con Candidate division NC10 bacterium CSP1-5, lo que sugiere que estas bacterias detectadas cercanas a los componentes agrupados en este cluster junto con ellos estarían en el límite de poder pertenecer cuando menos a un género diferente en el *filo Candidate division NC10* (Figura 49).

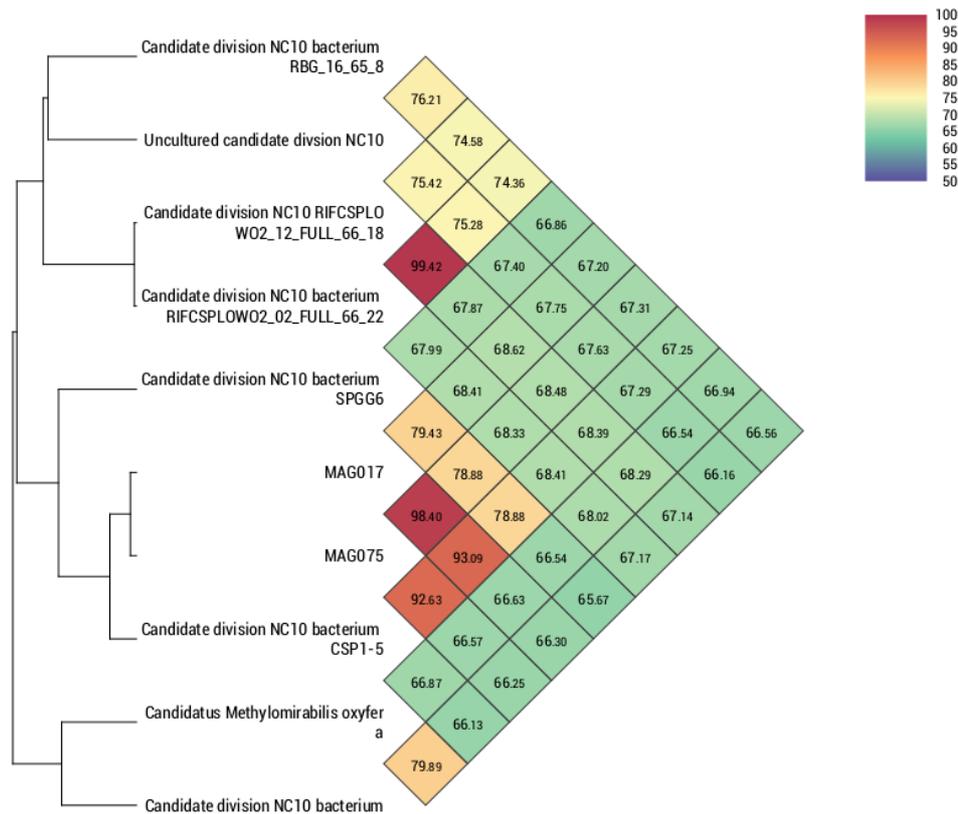


Figura 49. Valores ANI para los MAGs que agruparon en el cluster XI contra los genomas más cercanos disponibles dentro del filo Candidate division NC10.

### Cluster XII

En este cluster agruparon el MAG006 (G1) y el MAG085 (W2) con un valor de ANI del 99,90% y un AAI del 99,50%. Según estos valores de ANI y AAI, es probable que se trate de la misma bacteria. Los dos MAGS contienen secuencias enteras del 16S rRNA idénticas (casi) e interrumpidas por al menos 3 intrones (Figura 50). Una pista para notar esta situación en estas secuencias fue que la búsqueda de homologías con 16S rRNA modelo, en el caso de MAG006 producía varios hits cortos (unos 500 bp) localizados en el centro de un contig grande (NODE\_134\_length\_48745\_cov\_3.49823, -range 45312-41390) y no en el extremo del contig como suele ocurrir cuando se localiza una secuencia parcial de 16S rRNA. Las regiones V3-V4 de la secuencia del 16S rRNA en estos MAGs estarían interrumpidas por la presencia de los intrones 1 y 2 explicando la razón de no detectar este taxón por la técnica de amplicones. Además, los homólogos más cercanos son del grupo de

*Latescibacteria/Parcubacteria*, que forman parte del grupo CPR en los que se ha descrito previamente la presencia de intrones en los rRNA ribosomales.

El alineamiento del contig #134 del MAG006 con 16S rRNA próximos, permite definir con alguna precisión la estructura del gen 16S de este MAG que tendría 4 exones (151/286/989/96 pb) interrumpidos por tres intrones (358/1140/903 pb) con un tamaño total de 3923 bp (Figura 50).



Figura 50. Representación de la secuencia del 16S rRNA identificado en el MAG006 en colonias G1 que clasifica con una *Parcubacteria*. Intrones: cajas amarillas. Exones: cajas blancas.

De acuerdo con esto, el 16S rRNA maduro tendría unos 1522 bp con alguna incertidumbre en la precisión de los límites intron/exon y en los extremos de la molécula. En los intrones 2 y 3 se pueden identificar algunos elementos funcionales relacionados con la maduración del rRNA o con la integración/movilidad de los intrones. Dos CDS codifican para “Homing endonucleases” de la familia LAGLIDAG y en el extremo 5’ del intrón 2 [834..1316] se encuentra un posible RNA autocatalítico del grupo intrón I (Rfam RF00028).

El 16S rRNA presente en el MAG085 (segundo MAG agrupado en este cluster), presenta la misma organización del gen 16S encontrado en el MAG006. En las partes presentes, las secuencias son idénticas entre sí, aunque el segmento 2188..2732 del rRNA crudo que contiene 545 bp del exón 3 no se encuentra en el MAG085. La secuencia completa se encuentra repartida en tres contigs de los que solo dos (#5689 y #4440) están dentro del MAG085. El contig #14209 tiene un tamaño inferior a 1000 pb por lo que no está incluido en MAGs.

	%id	size	16S rDNA		Contig position	
NODE_14209_length_548_cov_0.929336	100.000	488	1	488	61	548
NODE_5689_length_1703_cov_0.776202	100.000	1703	485	2187	1	1703
NODE_4440_length_2236_cov_0.771230	99.916	1191	2733	3923	2236	1046

A pesar de los indicios de elementos móviles en los exones de estos rRNAs, no se detectaron secuencias en el resto del genoma con identidad significativa con estos intrones que pudieran indicar su movilidad reciente.

Los anteriores rRNAs (MAG006 y MAG085) presentan una secuencia muy relacionada en un contig en el metagenoma W1 que contiene un 16S prácticamente idéntico a ellos (NODE\_1920\_length\_3945\_cov\_0.589027). De hecho, este contig es idéntico al MAG006 excepto por una delección de 150 bp (en la posición 3164 del contig W1\_1920). Esta falta es extraña ya que corresponde con una zona “exónica del rRNA”, y aunque no se puede descartar de que se trate de un error de ensamblado, es al menos curioso que coincide con la ausencia mencionada de un segmento en el rRNA del MAG085 respecto al de 006.

La comparación del 16S completo (incluyendo intrones) no da un hit que cubra sus 3,9 kb. Las regiones con similitudes se restringen a los exones (Figura 51). El resumen gráfico de blastn contra la base de datos nr/nt (nucleotide collection) del NCBI refleja la organización del gen, y al mismo tiempo indica que no existen en las bases de datos otros genomas con la misma organización intrón/exón similar a la detectada en estos MAGs.



Figura 51. Salida gráfica de la comparación con Blastn del 16S rRNA (nativo) del MAG006 contra secuencias más cercanas en la base de datos nr/nt del NCBI.

Usando el 16S rRNA entero sin quitar los intrones, el mejor hit contra la base de datos de SILVA\_132 fue contra LN880482, un clon de 16S de una bacteria no cultivada obtenida del drenaje ácido de una mina ártica. Si se limita en la búsqueda en nr/nt del NCBI a los miembros del grupo *Parcubacteria*, entonces el mejor hit para el 16S completo fue con KX123463.1, *Candidatus* Nomurabacteria bacterium GW2011\_GWE1\_32\_28 16S ribosomal RNA gene, con un 80,68% de identidad que se limita a un 36% de la secuencia (solo los exones), el cual representa una especie no cultivada en este filo cuyo genoma ha sido recuperado de estudios metagenómicos de muestras de sedimentos de acuífero (Brown et al., 2015). Esta búsqueda en la base de datos gss (genomic survey sequences) y htgs (high throughput genomic sequences) del NCBI, también restringida a *Parcubacteria*, no dió ningún hit significativo.

Cuando se usa el posible 16S maduro (obtenido por yuxtaposición de los exones) la homología se extiende a la secuencia completa. Esto nos da cierta garantía de que el 16S maduro está bien ensamblado. En este caso, con el 16S rRNA maduro, el mejor hit obtenido contra nr/nt fue *Candidatus* Campbellbacteria bacterium GW2011\_GWC1\_35\_31 16S ribosomal RNA gene (KX123470.1) con un 79,12% de identidad con el 100% de la secuencia. De nuevo ocurre que representa una especie no cultivada en este filo cuyo genoma ha sido recuperado también de estudios metagenómicos de muestras de sedimentos de un acuífero (Brown et al., 2015). La comparación con la base de datos de 16S rRNA de referencia del NCBI (refseq\_RNA) no fue satisfactoria, el mejor hit (73,89% id y 89% cobertura) fue con NR\_040972 que perteneció a una bacteria del filo *Proteobacteria* y especie *Hyphomicrobium denitrificans* (NR\_074189.1).

Siguiendo en el MAG006 (MAG modelo en este análisis), el 23S rRNA se detectó en el contig mayor (#17, tamaño 113.411 bp) en un segmento de 7491 bp. La organización en el 23S fue similar a la del 16S. La posición probable de los intrones se ha deducido por comparación de la secuencia cruda con un modelo de covarianza de 23S rRNA bacteriano (RF002541.cm) usando Infernal cmsearch (Nawrocki and Eddy, 2013) y con algunos 23S cercanos. La falta de homólogos similares puede hacer que el modelo no sea totalmente correcto.

En sus proximidades también se encontró el gen del 5S rRNA, detectado igualmente con su modelo RF00001.cm.

En el gen del 23S rRNA se detectó la existencia de al menos 8 intrones (Figura 52). En los intrones mayores se han identificado genes codificantes para “Homing endonucleases” y al menos un RNA autocatalítico de la familia de los intrones del grupo I. Es de notar que los dos rRNAs mayores del MAG006 no parecen formar una unidad transcripcional, ya que se encuentran separados, de hecho se localizan en dos contigs diferentes. En el mejor de los casos, (que los dos contigs fueran adyacentes y en la orientación más favorable) la distancia entre los dos rRNAs mayores sería de al menos 14 kb. La secuencia de este 23S rRNA no ayuda a mejorar la situación taxonómica de este cluster. SILVA lo clasifica de nuevo en el grupo *Parcubacteria*, pero solamente con un 61% de identidad. Estos valores bajos, aunque pueden estar afectados por la presencia de intrones en el rRNA coinciden en no detectar secuencias clasificadas muy similares y soportarían que estos MAGs representaran una nueva división dentro de *Parcubacteria* con las limitaciones que presentan estos análisis.



Figura 52. Representación de la secuencia del 23S rRNA correspondiente al MAG006 en colonias G1 que clasifica con una *Parcubacteria*. Intrones: cajas amarillas. Exones: cajas blancas.

Por lo tanto, en base a la asignación de los genes ribosomales, podemos determinar que se trata de una bacteria que clasificó con el superfilo *Parcubacteria*, aunque no está claro si con el filo *Candidatus* Campbellbacteria o *Candidatus* Nomurabacteria. Según la clasificación de sus ORFs con CAT/BAT, no clasificaron con ningún filo, pero indicaron que eran bacterias.

En GenBank bajo el filo *Candidatus* Nomurabacteria no hay ninguna clase/orden/familia/género descrito encontrando 187 ensamblados, de los cuales 35 se encuentran bajo el nombre de *Candidatus* Nomurabacteria bacterium, mientras el resto son ensamblados de especies únicas. Una situación similar ocurre con el filo *Candidatus* Campbellbacteria, con 31 ensamblados. El valor de ANI del cluster contra algunos de estos genomas, incluidos los hits más cercanos por comparación de su secuencia del 16S rRNA contra las bases de datos, fue alrededor del 63%, y si sumamos la escasa identidad tanto de la secuencia completa del 16S rRNA como de las comparaciones de sus ORFs contra las referencias de las bases de datos, soportarían la idea de que puedan pertenecer a una especie diferente en alguno de estos filos descritos (Figura 53).

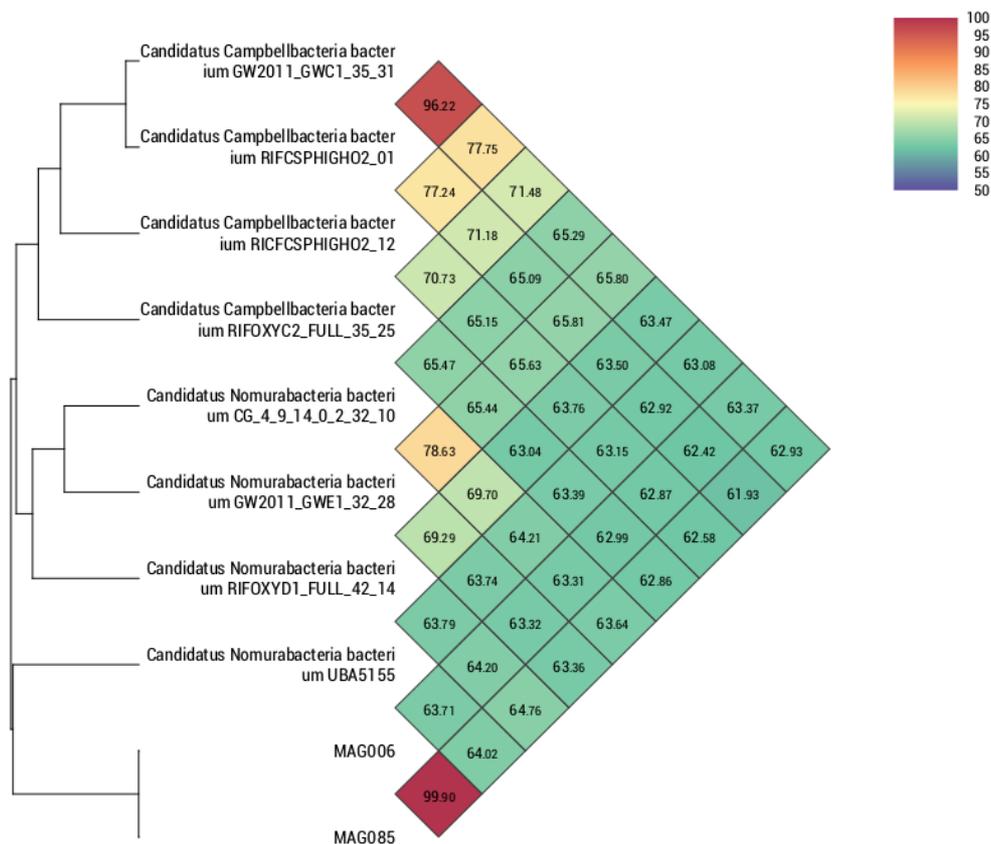


Figura 53. Valores ANI para los MAGs que agruparon en el cluster XII contra los genomas más cercanos disponibles de los filos *Candidatus* Nomurabacteria y *Candidatus* Campbellbacteria.

**Cluster XIII**

En este último cluster agruparon el MAG002 (G1), el MAG079 (W1) y el MAG052 (G2) con un valor de ANI del 99,80% y de AAI >99,70%. Según estos valores de ANI y AAI indicaron que todas son la misma bacteria habiendo bineado a excepción del MAG079 dos fragmentos conteniendo secuencias enteras (1585/1584) del 16S rRNA. Podemos determinar que los MAGs correspondieron a bacterias del filo *Dadabacteria*, clase *Dadabacteriia*, orden *Dadabacteriales* y la bacteria *Candidatus Dadabacteria bacterium* CSP1-2 con una identidad del 91%. Según la clasificación de sus ORFs con CAT/BAT, no clasificaron con ningún filo, pero indicaron que eran bacterias.

En GenBank hay 52 ensamblados bajo este filo candidato y ninguna clase/orden/familia/género descrito, sólo 6 especies descritas y bajo el nombre *Candidatus Dadabacteria bacterium* encontramos 47 ensamblados. El valor de ANI del cluster contra algunos de estos genomas fue 62-67%, lo que sugiere que las bacterias detectadas pertenecen cuando menos a una especie diferente de las analizadas en este análisis en el filo *Candidatus Dadabacteria* (Figura 54).

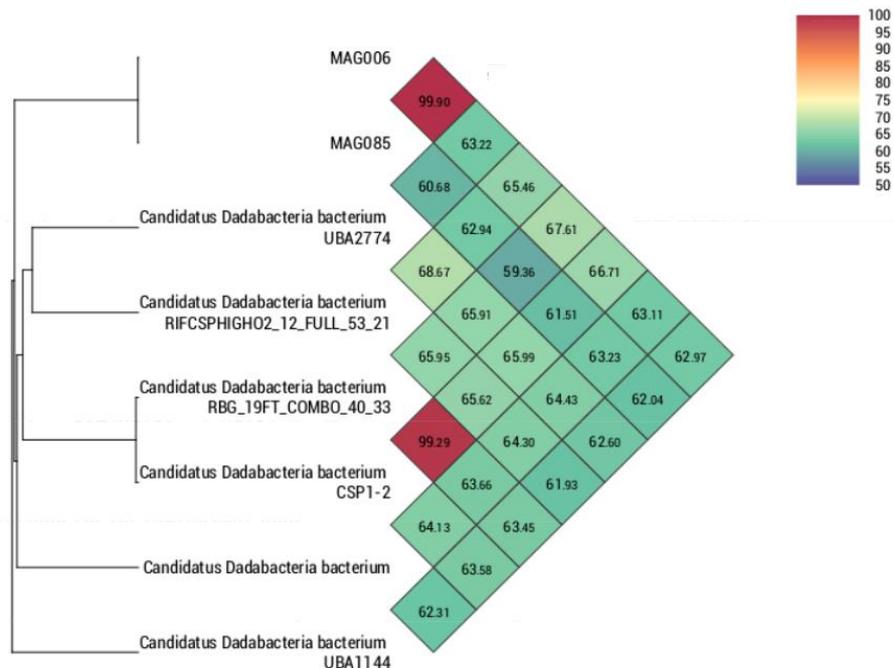


Figura 54. Valores ANI para los MAGs que agruparon en el cluster XIII contra los genomas más cercanos disponibles dentro del filo *Candidatus Dadabacteria*.

Al comparar las secuencias de 16S rRNA encontradas en estos MAGS contra la base de datos nr/nt de GenBank (26/12/2020), el hit más cercano fue un clon no cultivado de rRNA (GenBank: EF494370.1) con una identidad del 91% y una cobertura del 96% obtenido del fondo de un río. La escasa identidad tanto de la secuencia completa del 16S rRNA como de las comparaciones de sus ORFs contra las referencias de las bases de datos, soportan la idea de que puedan pertenecer a una especie diferente en el filo *Candidatus* Dadabacteria.

A parte de los MAGs que han podido ser agrupados en clusters, encontramos aún otros 53 que no agruparon en estos clusters mencionados y que constituirían MAGs únicos dentro de los taxones *Acidobacteria*, *Actinobacteria*, *Candidatus* Rokubacteria, *Chloroflexi*, *Proteobacteria* (*alpha*, *beta*, *gamma* y *delta*) y *Nitrospirae*.

En el filo *Acidobacteria* se asignaron 5 MAGs (MAG009, 26, 32, 87 y 93) en las colonias YC, W2, G1 y G2. Según la clasificación de sus ORFs con CAT/BAT, no clasificaron más que a nivel de filo. Solo tres de ellos contenían fragmentos con secuencias parciales del 16S rRNA (368/384/998/1276) que clasificaron con *Acidobacteria* de los subgrupos 6, 9 y 17.

Del filo *Actinobacteria* se obtuvieron otros 7 MAGs (MAG014, 18, 48, 84, 92, 95 y 100), de los cuales dos, MAG014 (G1) y MAG084 (G2) se clasificaron en la misma familia *Pseudonocardiaceae* que los componentes mayoritarios, y que según la clasificación de sus ORFs con CAT/BAT se clasificaron con el género *Pseudonocardia*. Los otros 5 MAGs del filo *Actinobacteria* fueron el MAG048 (W2) que clasificó con la clase *Thermoleophilia* y orden *Solirubrobacterales*, el MAG092 (BC) clase *Actinobacteria*, orden *Corynebacteriales* y familia *Mycobacteriaceae*, el MAG018 (G1) clase *Actinobacteria*, orden *Micromonosporales* y familia *Micromonosporaceae*, y finalmente MAG095 (YC) y MAG100 (BC) filo *Actinobacteria*, en todos los casos según la asignación obtenida con CAT/BAT. En estos dos últimos MAGs, sólo el MAG100 tenía bineado un fragmento conteniendo secuencia parcial (1155 pb) del 16S rRNA, que fue clasificada con el filo *Actinobacteria*, orden *Thermoleophilia*, orden *Solirubrobacterales* y género 67-14.

En la clase *Alphaproteobacteria*, se recuperaron 9 MAGs (MAG012, 19, 27, 28, 55, 57, 59, 65 y 76), todos detectados en colonias grises subtipo 1 y 2, y tan sólo uno en colonias blancas subtipo W1. La mayoría de ellos (a excepción del MAG055, 65 y 27), se clasificaron en las familias *Hyphomicrobiaceae* y *Rhodospirillaceae* según la asignación obtenida con CAT/BAT.

Nos llama la atención el MAG055, que según la clasificación de sus ORFs con CAT/BAT fue clasificado solamente en el filo *Proteobacteria*, pero, sin embargo, es un MAG bastante completo (94,51%) y no muy contaminado (7,76%) presentando un fragmento conteniendo una secuencia completa del 16S rRNA que clasificó con el orden *Azospirillales*, la familia *Inquilinaceae* y el género *Inquilinus*. Dentro del género *Inquilinus* encontramos también dos aislamientos cultivables idénticos recuperados de muestras de colonias grises y blancas obtenidos en los intentos de cultivo que describiremos más abajo.

En GenBank hay 6 ensamblados de *Inquilinus*, 2 de ellos son aislados cultivables validados como *Inquilinus ginsengisoli* (Jung et al., 2011; aislado de suelo) e *Inquilinus limosus* (Coenye et al., 2002; aislado de secreciones respiratorias de pacientes con fibrosis quística).

Para poder esclarecer la taxonomía en este MAG, hemos comparado la secuencia de DNA del MAG055 con los genomas disponibles y con el de uno de los dos aislamientos crecidos por nosotros perteneciente al género *Inquilinus*. Los genomas de referencia de *I. limosus* e *I. ginsengisoli* fueron los más parecidos al MAG055 con un valor de ANI de 89% similar al que presentaba con el genoma de nuestro aislado, lo que sugiere que todas ellas pertenecen al género *Inquilinus*, representando el MAG055 una especie diferente al igual que nuestro aislamiento cultivado (Figura 55).

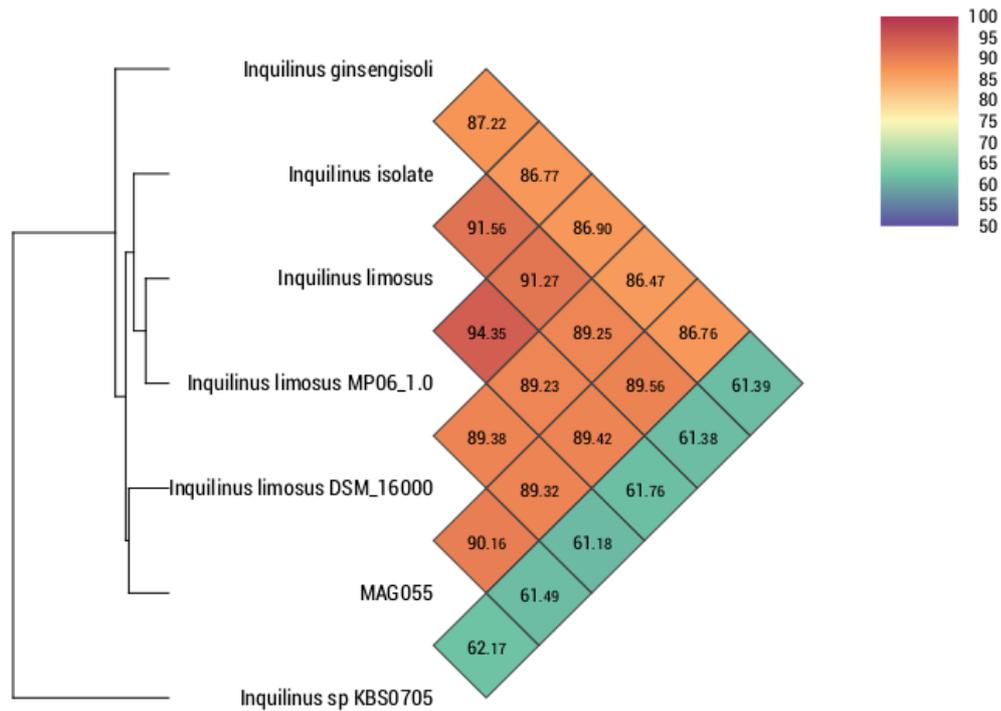


Figura 55. Valores ANI del MAG055 contra los ensamblados de referencia de la especie *Inquilinus* disponibles en la base de datos “genoma” del NCBI y también el genoma obtenido de la secuenciación del genoma del cultivable recuperado del cultivo de colonias.

Dentro de la clase *Betaproteobacteria* se recuperaron otros 5 MAGs (MAG007, 08, 24, 29 y 30) y todos ellos en colonias grises subtipo G1, en los que destacó según la clasificación de sus ORFs con CAT/BAT el orden *Burkholderiales* al que fueron asignados 2 MAGs (MAG029 y 30).

En la clase *Gammaproteobacteria* encontramos 5 MAGs (MAG016, 51, 54, 91 y 99), de los cuales el MAG051 (G2) fue clasificado según la taxonomía de sus ORFs con CAT/BAT con el orden *Xanthomonadales*. Sin embargo, MAG051 no agrupó con el cluster VIII clasificado con el género *Lysobacter*, además no tiene ningún fragmento que contenga alguna secuencia entera o parcial del 16S rRNA bineado, aunque si se encuentra no bineado un fragmento de 1408 pb (contig 6905) clasificando con la familia *Xanthomonadaceae* y género *Arenimonas*. Los otros 4 MAGs restantes detetados en colonias G1, G2, YC y BC que según la clasificación de sus ORFs con CAT/BAT, correspondieron con la clase *Gammaproteobacteria*, presentando fragmentos que representaron

secuencias parciales de 16S rRNA permitiendo clasificarlos en los órdenes *CCD24* y *PLTA13*. Estos órdenes son sólo existentes en SILVA, y eso hace que por la clasificación de sus ORFs usando la taxonomía del NCBI no se clasificasen más que a nivel de filo (*Proteobacteria*).

Dentro de la clase *Deltaproteobacteria* hay 4 MAGs (MAG013, 53, 66 y 99) detectados en colonias G1, G2 y W1 que según la clasificación de sus ORFs con CAT/BAT no se clasificaron a mayor nivel que la clase. Sólo 2 de ellos tenían binned fragmentos que contienen secuencias parciales (840 pb) que clasificaron con el orden *MBNT15*, y entera y parcial (1569/137) del 16S rRNA, con el orden *Myxococcales*.

Finalmente, encontramos 6 MAGs en colonias G1 y G2 que según la clasificación de sus ORFs con CAT/BAT no clasificaron más que a nivel del filo *Proteobacteria*, y no presentaron fragmentos que representaran secuencias parciales o enteras del 16S rRNA, además 4 de ellos tienen una completitud muy baja (<46%) dificultando una mejor clasificación.

En el filo *Nitrospirae* encontramos según la clasificación de sus ORFs con CAT/BAT 2 MAGs (MAG020 y 56), en colonias G1 y G2, que clasificaron con este filo pero que no agruparon dentro del cluster X asignado a este mismo filo. Además, no presentaron ningún fragmento que contuviera secuencias parciales o enteras de 16S rRNA.

Bajo el filo *Candidatus Rokubacteria* clasificaron 2 MAGs (MAG033 y 68) en colonias G1 y G2 que no presentaron ningún 16S rRNA binned, pero que según la clasificación de sus ORFs con CAT/BAT se asignaron bajo este filo candidato. Sus genomas se han recuperado muy poco completos (<30%), y de nuevo dificulta una mejor caracterización. Esta situación sucede con el MAG096, el único que clasifica según la clasificación de sus ORFs con CAT/BAT en el filo *Bacteroidetes*. No presentaba ningún fragmento que representaran secuencias del 16S rRNA y su poca completitud (35,52%) no ayuda en su caracterización.

Bajo el filo *Chloroflexi* encontramos según la clasificación de sus ORFs con CAT/BAT, 4 MAGs (MAG021, 25, 64 y 69) en colonias grises G1 y G2, los cuáles dos de ellos (MAG021/69) presentaron secuencias parciales y enteras (728/1502/1429) del 16S rRNA clasificando con la clase *Anaerolineae* y órdenes *Ardenticatenales* y *SBR1031*.

Por último, encontramos dos MAGs como no clasificados, el MAG072 (G2) y el MAG022 (G1), que según la clasificación de sus ORFs con CAT/BAT no clasificaron más que al dominio Bacteria y no presentaban un 16S rRNA bineado. Tienen una completitud de 39,22% y 58,56%, respectivamente. También encontramos dos MAGs, MAG031 (G1) y el MAG071 (G2), sospechosos de estar sujetos de contaminación, que según la clasificación de sus ORFs con CAT/BAT clasificaron en 2 taxones diferentes. La estimación de la contaminación obtenida en estos últimos MAGs en base a la presencia del modelo de genes que deberían estar sin redundancia por CheckM no es del todo realista, ya que seguramente la contaminación real por contigs provenientes de distintas bacterias es alta en estos MAGS.

#### **5.2.4 Análisis filogenético de los componentes mayoritarios de las colonias**

Los componentes mayoritarios de las colonias representan los genomas más abundantes que encontramos en las colonias. Su abundancia ha sido corroborada tanto por la técnica de shotgun como por la de amplicones, ambos enfoques empleados en esta tesis (Figura 34 y 38). Estos componentes mayoritarios que también se pueden encontrar como componentes secundarios en otro tipo de colonias, son un total de 7 bacterias cuya presencia es característica para cada subtipo de colonia siendo:

- MAG001 en colonias grises subtipo1 (G1)
- MAG046 en colonias grises subtipo2 (G2)
- MAG073 en colonias blancas subtipo1 (W1)
- MAG082 en colonias blancas subtipo2 (W2)
- MAG088 en colonias amarillas (YC)

- MAG097 y MAG098 en colonias beis (BC)

Los genomas de los componentes mayoritarios por lo general se han recuperado muy completos >96% y presentan una contaminación baja <13% (Tabla 4). El tamaño de estos genomas es de 5 a 7 Mb y el contenido en GC es similar. El número de contigs obtenidos en algunos casos es alto como en los CMs de las colonias grises y en el MAG097 de las colonias beis.

MAG	Compl	Cont	Tamaño genoma	Longest contig	Nº contigs	Genes	%GC
MAG001	100	12,93	7.488.565	131.484	438	8.121	0,67
MAG046	98,28	3,45	6.936.749	158.394	480	7.421	0,68
MAG073	100	0	6.458.129	310.637	124	6.583	0,67
MAG082	96,55	6,9	5.463.669	126.841	268	5.477	0,69
MAG088	100	10,34	6.179.751	256.681	172	6.599	0,68
MAG097	96,55	6,9	5.316.480	59.461	689	5.568	0,69
MAG098	98,28	2,59	5.780.955	172.933	267	5.977	0,67

Tabla 4. Parámetros principales de calidad y tamaño de los componentes mayoritarios de las colonias. Compl: Completitud; Cont: Contaminación.

Una característica particular de estos componentes mayoritarios en las colonias fue su taxonomía, todos pertenecieron al filo *Actinobacteria*. Este filo ha sido descrito como el más abundante en las colonias mediante la técnica de amplicones, representando el 43,36% de las secuencias totales, mientras que en los controles (soporte sin colonización visible) representó un 9,04% y en las muestras de suelo y soporte un 10,83%. Específicamente, los componentes mayoritarios apenas fueron detectables en las muestras de suelo y soporte. En base a los enfoques empleados para caracterizar taxonómicamente los borradores de genomas que se han recuperado de los metagenomas y que se recoge en el apartado previo de asignación taxonómica de los MAGs, los componentes mayoritarios se organizarían en 2 clusters diferentes correspondientes a 2 taxones distintos: el cluster III que agrupa los MAGs que clasifican en base a la asignación taxonómica de sus ORFs con el orden

*Pseudonocardiales* y según su 16S rRNA con el género *Crossiella*, y el cluster IV que se clasifica en base a la asignación taxonómica de sus ORFs con la clase *Actinobacteria* y según su 16S rRNA con la familia *Euzebyaceae*.

### **Cluster III**

La mayoría de los componentes mayoritarios clasificaron en esta familia. También se ha detectado un componente secundario en las colonias grises subtipo G2 (MAG060) que agruparía en este cluster. Estos componentes mayoritarios son:

- MAG001 en colonias grises subtipo1 (G1)
- MAG046 en colonias grises subtipo2 (G2)
- MAG073 en colonias blancas subtipo1 (W1)
- MAG088 en colonias amarillas (YC)
- MAG098 en colonias beis (BC)

Como primer paso en su caracterización llevamos a cabo comparaciones entre todos de ellos de sus secuencias de DNA y de las proteínas predichas, para determinar si algunos de estos componentes de las colonias representaban las mismas bacterias. Como se ha mencionado en el apartado de la taxonomía de los clusters tras las comparaciones de secuencia de DNA, los valores de ANI deben superar el límite del 77% de identidad para considerarlos perteneciente al mismo género, y a partir de 93-96% se considerarían miembros de la misma especie (Rosselló-Móra and Amann, 2015). En cuanto a los valores de AAI, para ser la misma especie este debería ser mayor del 93-95% de identidad y para pertenecer al mismo género sería mayor del 65-72% (Konstantinidis and Tiedje, 2007).

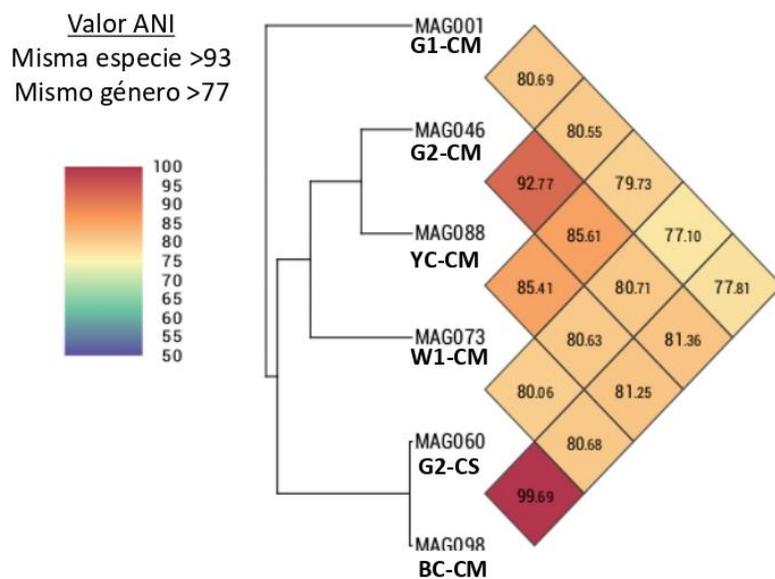
La primera información que se obtuvo de estos cinco componentes en base a la técnica de amplicones y utilizando la región V3-V4 (411 pb) del gen 16S rDNA fueron clasificados en dos OTUs: OTU2 para los CMs de G2, W1, YC y BC y OTU4 únicamente para el CM de G1. Ambas OTUs se asignaron al género *Crossiella*.

Sin embargo, los valores de ANI y de AAI obtenidos en la comparación de los genomas de los componentes mayoritarios indicarían una situación diferente, sugiriendo que a pesar de su agrupación en solo dos OTUS, en realidad contienen 5 bacterias distintas, que posiblemente agrupen dentro del mismo género (Figura 56). Estos valores también indicarían que el componente secundario detectado en este cluster MAG060 (G2) sería casi idéntico al componente mayoritario MAG098 (BC), demostrando que los componentes mayoritarios también pueden aparecer como secundarios en colonias de diferente tipo.

Para saber si estas bacterias son las mismas, también tenemos la opción de comparar las secuencias de los 16S rRNA ya que todos ellos, a excepción del MAG060, presentaron una o dos copias enteras del 16S rRNA. La identidad del 16S rRNA entre miembros de una misma especie siempre se ha visto establecida al 97% (Tindall et al., 2010), aunque debido a la diversidad intraespecífica, esta cifra se ha visto debatida y propuesta a niveles superiores 98,7–99,0% (Kim et al., 2014). Si nos fijamos entonces en la identidad obtenida sólo de las comparaciones de las secuencias del 16S rRNA, y fijando el 97% como criterio de pertenencia a la misma especie, todas estos MAGs lo cumplen, a excepción del MAG001 en colonias G1 (Figura 57). Si consideramos el valor de 98,7-99% para ser la misma especie, sólo en este caso los componentes mayoritarios de G2 (MAG046) e YC (MAG088) serían la misma especie (id de 99,41%). Sin embargo, los valores obtenidos de ANI (92,77%) y AAI (89,78%) más bien sugieren que estos MAGs serían especies diferentes.

Ante esta incertidumbre, una decisión más robusta provendrá del resultado consensuado de diferentes aproximaciones, ya que una nunca sería suficiente. En base a esto, si nos fijamos en las 3 aproximaciones la hipótesis más sólida sería que estas bacterias pertenecen a diferentes especies (Figura 57).

### Comparación de la secuencia de DNA



### Comparación de la secuencia de proteína

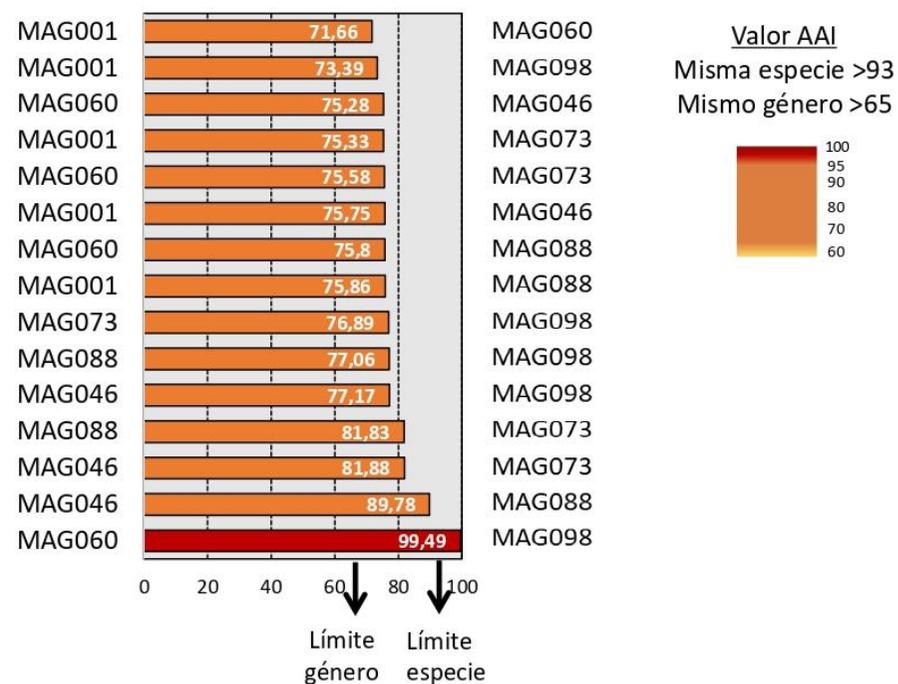


Figura 56. Valores de ANI y AAI de los componentes mayoritarios en las colonias que agruparon en el cluster III. En este cluster ha agrupado también el componente secundario MAG060 perteneciente a la misma especie que el componente mayoritario MAG098. CM: Componente Mayoritario; CS: Componente Secundario.

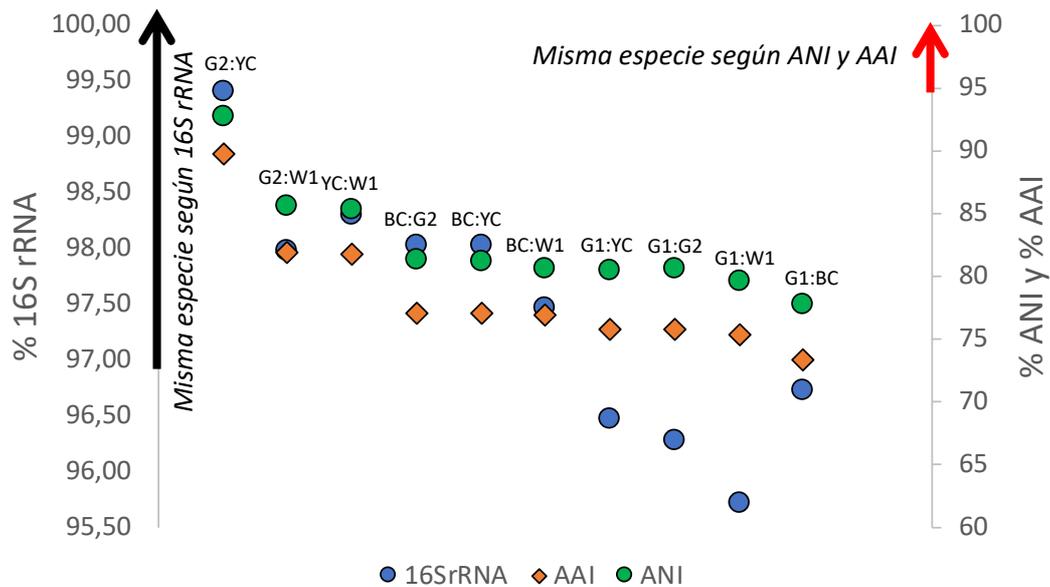


Figura 57. Comparación de las secuencias del 16S rRNA, valores ANI y AAI para los componentes mayoritarios agrupados en el cluster III.

Tras determinar que probablemente todos estos MAGs pertenecen a diferentes especies bacterianas, hemos comparado las secuencias de sus 16S rRNA contra las de las referencias de la base de datos de SILVA\_132. Las secuencias de los componentes mayoritarios se alinearon y se insertaron en un árbol junto con las de referencia. La primera asignación taxonómica para estos componentes se hizo en base a la técnica de amplicones, utilizando la región V3-V4 (411 pb) del gen 16S rDNA que los clasificó con el género *Crossiella*. Al usar la secuencia completa del 16S rRNA, siguieron clasificando dentro de la familia *Pseudonocardiaceae*. Sin embargo, al comparar el 16S rRNA entero de los componentes mayoritarios contra las secuencias de referencia correspondientes en este género *Crossiella*, la identidad fue inferior al 95%. Esta baja identidad se observó también para el resto de secuencias de referencia correspondientes a los demás géneros descritos dentro de esta familia, agrupando las secuencias de los componentes mayoritarios en un género diferente.

# RESULTADOS



Figura 58. Árbol de máxima verosimilitud que muestra las relaciones filogenéticas de los genes del 16S rRNA de los componentes mayoritarios agrupados en el cluster III con las secuencias disponibles de los géneros descritos en la familia *Pseudonocardiaceae*. Se incluyen representantes de secuencias de especies para cada género descrito dentro de esta familia. El árbol se ha construido con el software ARB por el método de PHYML version 20130708 (Guindon and Gascuel, 2003). Los soportes de ramificación determinados por la estimación bayesiana >51 se muestran en los puntos de ramificación. La barra de escala indica 0,1 cambios por nucleótido. El análisis involucró 348 secuencias de nucleótidos. La versión de SILVA utilizada fue SSURef NR\_99\_release\_132.

El árbol que representa la filogenia de los componentes mayoritarios se ha construido con un total de 348 secuencias (Figura 58). Las secuencias de los componentes mayoritarios agruparon con otras 262 secuencias no cultivadas cuyo origen en la mayoría de ellas fueron de colonias visibles (amarillas, blancas o marrón claro). En el árbol que contiene estas secuencias se muestra el origen y tamaño de secuencia de las secuencias más cercanas a los componentes mayoritarios. También observamos que los componentes mayoritarios de las colonias se han detectado en otros estudios de colonias similares a éste.

A continuación, estimamos las distancias de los genomas de referencia de los géneros representados en el árbol con los genomas de los componentes mayoritarios. Se han añadido 2 genomas adicionales que no se detectan en el árbol, correspondientes a los géneros *Actinocrispum* y *Gandjariella*. Estos géneros representan géneros en base a la nomenclatura del NCBI, aunque no en SILVA y por eso no aparecen en el árbol que ha sido construida con la base de datos de SILVA. Aparte, encontramos géneros como *Actinorectispora* y *Thermotunica* que no contienen genomas disponibles en las bases de datos. Estos genomas, 31 en total, se han descargado de la base de datos “genoma” del NCBI y sus números de acceso se encuentran recogidos en el Anexo-Tabla S8. Las distancias entre ellos se calcularon con el programa Mash, el cual estima la fracción de k-mers compartidos entre 2 genomas. Los genomas de los componentes mayoritarios formaron un cluster separado de los genomas de referencia indicando una composición bastante distinta a estos (Figura 59),

tampoco entre ellos presentaron una composición similar, como también se deduce de los valores de ANI y AAI.

Por tanto, los análisis que se obtienen tanto por comparación de la secuencia completa del 16S rRNA como por la comparación de los genomas de los componentes mayoritarios contra las bacterias de referencia en la familia *Pseudonocardia*, indicarían que los componentes mayoritarios recogidos dentro de este cluster, podrían constituir un “nuevo género” dentro de la familia *Pseudonocardia* con 5 especies distintas. Además, este nuevo género estaría compuesto de bacterias especializadas en ambientes de cueva que a día de hoy son prácticamente desconocidas, posiblemente por su incapacidad de recuperarse como cultivos puros, pero son las que durante este tiempo han desarrollado las herramientas necesarias para sobrevivir en estos entornos.

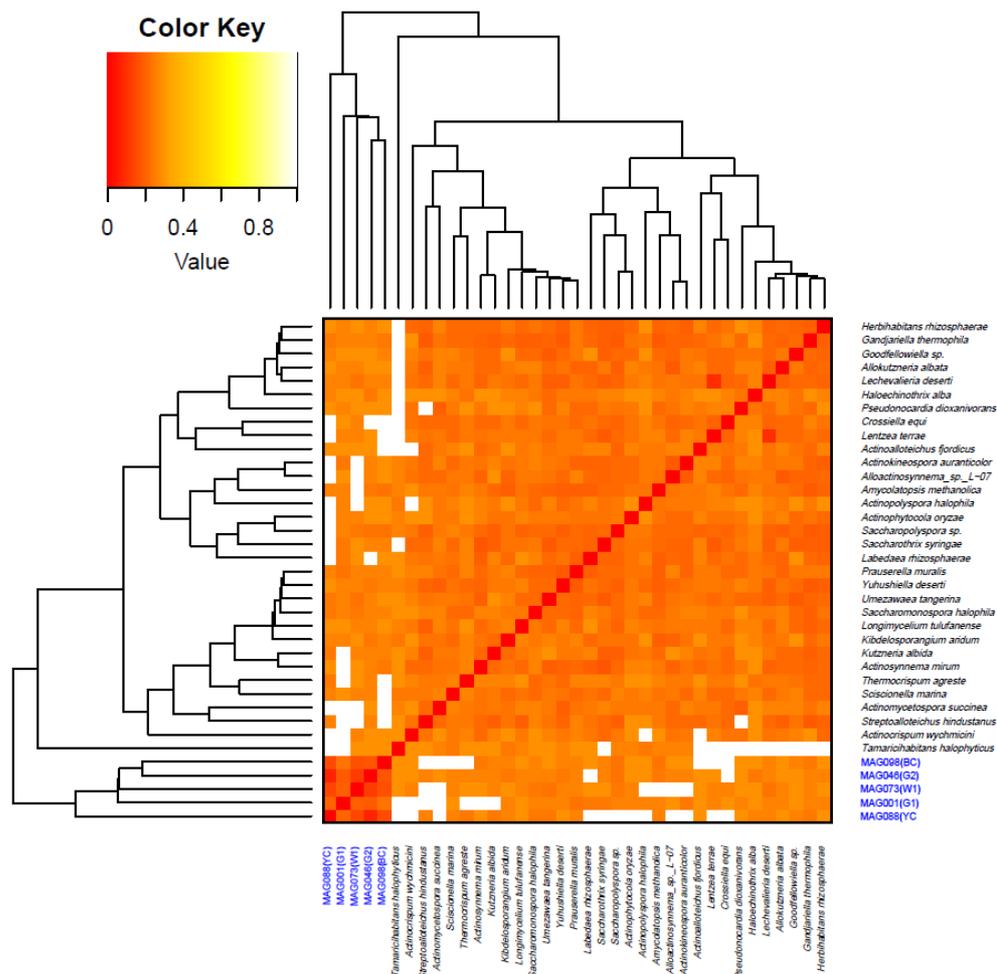


Figura 59. Distancias estimadas entre los genomas de los componentes mayoritarios que agruparon en el cluster III y al menos un representante de los genomas de referencia de los géneros descritos dentro de la familia *Pseudonocardia*.

**Cluster IV**

Dentro de este cluster, sólo agruparon 2 componentes mayoritarios, MAG082 y MAG098, y 2 componentes secundarios en colonias amarillas (MAG090) y en colonias grises (MAG050). En base a la técnica de amplicones y utilizando la región V3-V4 (411 pb) estos tres MAGs se integraron en la misma OTU1 que fue asignada a la familia *Euzebyaceae*, indicando que los 16S rRNA de estos componentes presentaban en ese fragmento una identidad superior al 97%. Para confirmar esta observación se llevó a cabo una comparación de las secuencias totales de DNA y proteína (Figura 61). Los valores de ANI y AAI obtenidos no soportaron esa homogeneidad del cluster. Los MAGs del cluster IV podrían pertenecer a 2 especies diferentes dentro del mismo género, una representada solamente por MAG082 y la otra conteniendo los MAG098, 90 y 50. Además, todos estos MAGs presentaron una copia completa del 16S rRNA y como se ha descrito más arriba, la combinación de los 3 enfoques seguiría reafirmando la hipótesis de la presencia de 2 bacterias diferentes (Figura 60). En este cluster en particular, los valores de ANI y AAI que indicarían que se tratara de la misma especie, correlacionaron con una identidad del 16S rRNA >99,7%.

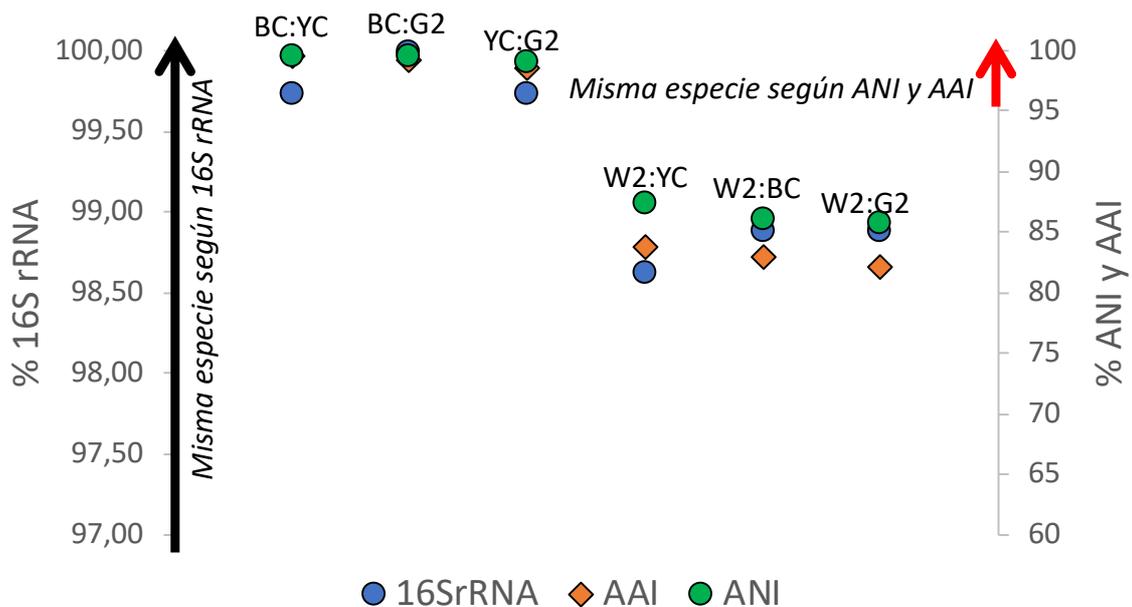
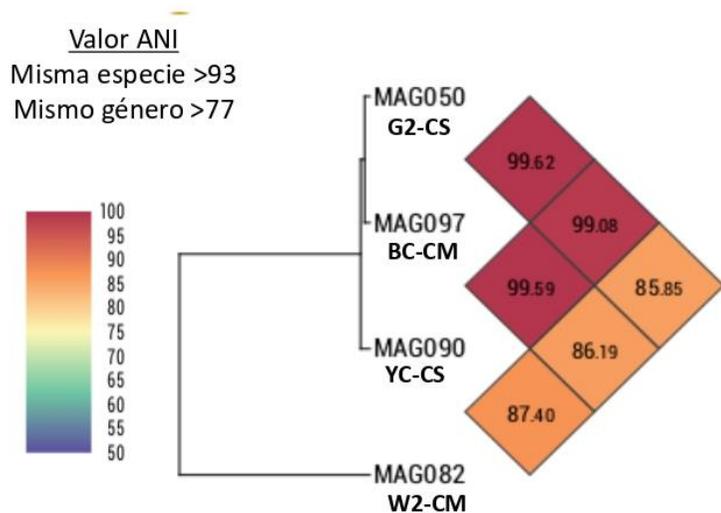


Figura 60. Comparación de las secuencias del 16S rRNA, valores ANI y AAI para los componentes agrupados en el cluster IV.

### Comparación de la secuencia de DNA



### Comparación de la secuencia de proteína

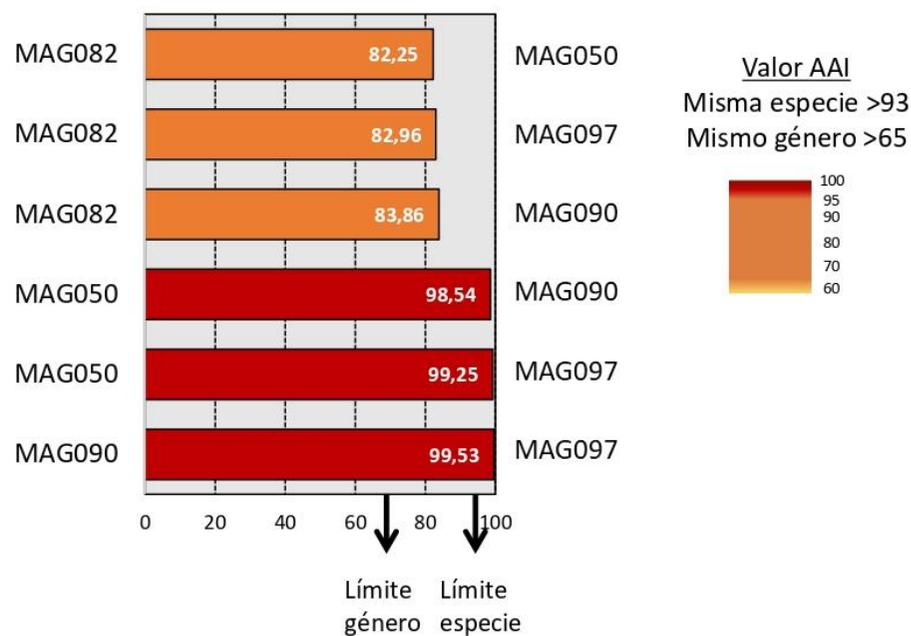


Figura 61. Valores de ANI y AAI de los componentes mayoritarios en las colonias que agruparon en el cluster IV. En este cluster han agrupado también los componentes secundarios MAGs 50 y 90 tratándose de la misma especie que el componente mayoritario MAG097. CM: Componente Mayoritario; CS: Componente Secundario.

La clasificación taxonómica en este cluster ha sido siempre incierta. En base al amplicón obtenido (411pb), se clasificaban dentro de la familia *Euzebyaceae* con una identidad del 97%, pero si consideramos la secuencia completa, la identidad contra las secuencias de este orden es inferior al 92%. Este orden se encuentra dentro de la clase *Nitriliruptoria* que recoge en total 4 órdenes: *Euzebyales*, *Egibacterales*, *Egicoccales* y *Nitriliruptorales*. La identidad de las secuencias del 16S rRNA de los CMs a las secuencias recogidas en los restantes últimos órdenes también es inferior al 94%.

El árbol, por tanto, fue construido con las secuencias recogidas dentro de la clase *Nitriliruptoria*, que son un total de 143 secuencias. Las secuencias de los componentes mayoritarios agruparon dentro de un cluster con 71 secuencias pertenecientes a bacterias no cultivadas (Figura 62). De nuevo, observamos que el origen de estas bacterias cercanas a los componentes en la gran mayoría fue también de colonias visibles similares a las de Altamira.

Si estimamos las distancias entre los genomas de referencia que encontramos de estos 4 órdenes mencionados anteriormente contra los componentes mayoritarios, volvemos a encontrarnos con una muy baja similitud (Figura 63). Estos genomas, 13 en total, se han descargado de la base de datos "NCBI genoma" y se encuentran recogidos en el Anexo-Tabla S8. Esta baja similitud también es correlacionada a que estamos comparando bacterias de órdenes diferentes. También hay que tener en cuenta que el número de genomas de referencia disponibles de cada orden dentro de la clase *Nitriliruptoria* es muy escasa, tan sólo 13 genomas, y, por tanto, la poca información a día de hoy complica la obtención de una taxonomía muy precisa en estos componentes.

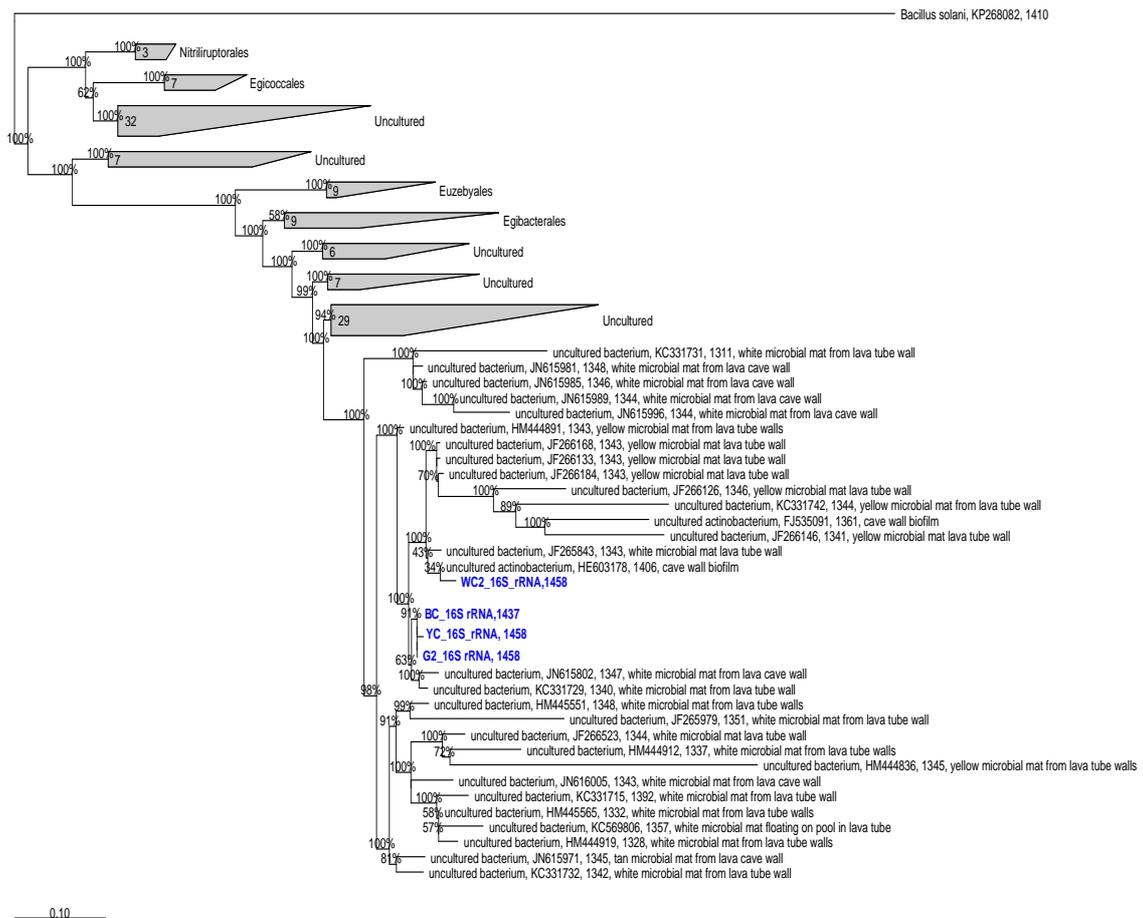


Figura 62. Árbol de máxima verosimilitud que muestra las relaciones filogenéticas de los genes del 16S rRNA de los componentes mayoritarios agrupados en el cluster IV con las secuencias disponibles de los géneros descritos en la clase *Nitriliruptoria*. Se incluyen representantes de secuencias de especies para cada género descrito dentro de esta clase. El árbol se ha construido con el software ARB por el método de PHYML version 20130708 (Guindon and Gascuel, 2003). Los soportes de ramificación determinados por la estimación bayesiana >51 se muestran en los puntos de ramificación. La barra de escala indica 0,1 cambios por nucleótido El análisis involucró 143 secuencias de nucleótidos. La versión de SILVA utilizada fue SSURef NR\_99\_release\_132.

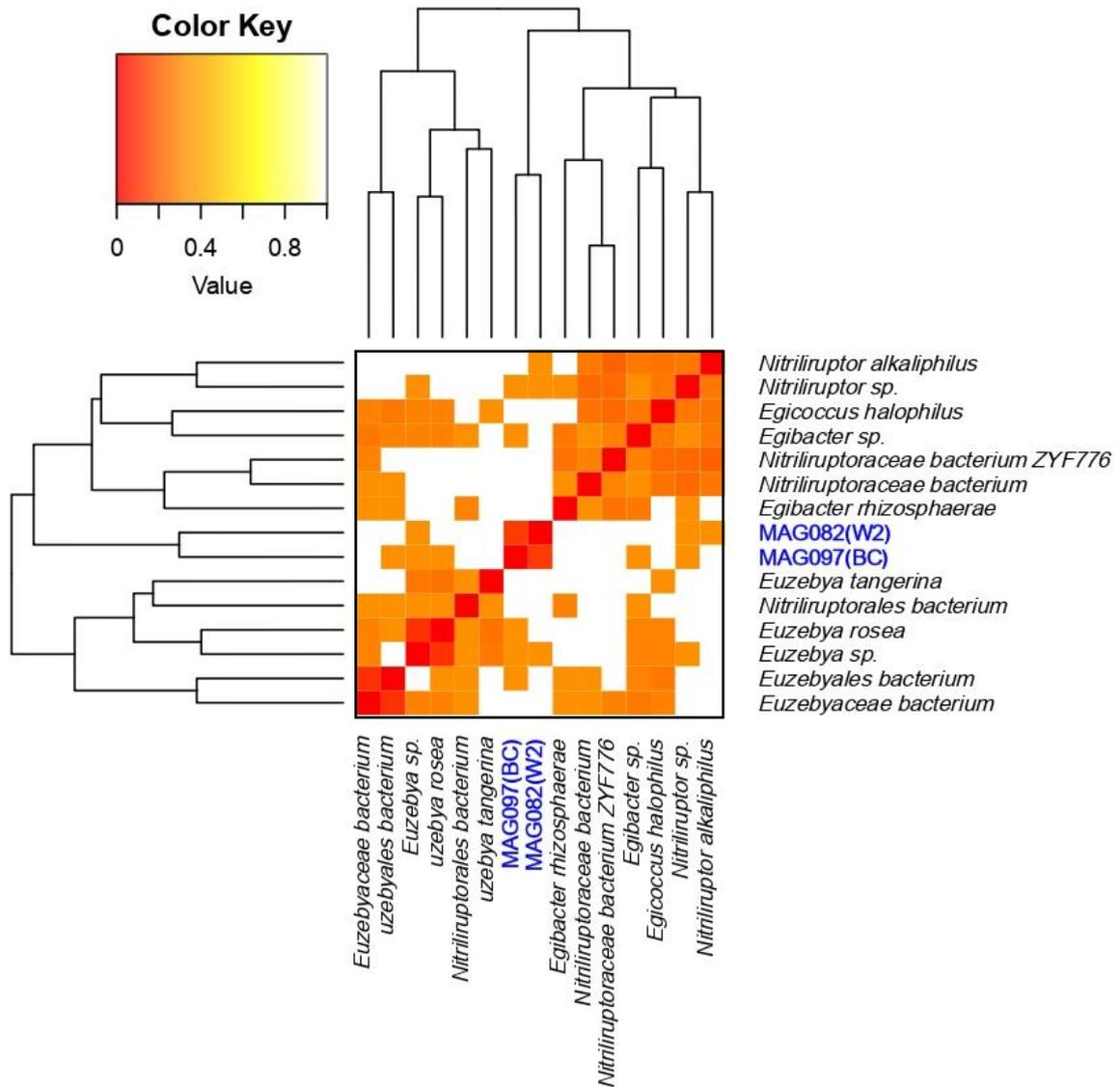


Figura 63. Distancias estimadas entre los genomas de los componentes mayoritarios que agruparon en el cluster IV y los genomas de referencia de los géneros descritos dentro de la clase *Nitriliruptoria*.

### 5.2.5 Métodos para detección de componentes mayoritarios de las colonias mediante el diseño de primers

Los resultados obtenidos a lo largo de esta tesis han permitido determinar que en las colonias siempre encontramos la presencia particular de una o dos bacterias muy abundantes denominadas como componentes mayoritarios de las colonias, y que estos componentes no se encuentran en cantidades apreciables en el soporte carente de colonias. La presencia de estos componentes en la pared por encima de unos niveles basales será indicativa

de la aparición y futuro desarrollo del tipo de colonia característico de este componente.

Aprovechando esta singularidad, y al disponer de borradores muy completos de los genomas de esos componentes, decidimos que el diseño de primers específicos para estos componentes podría proporcionar herramientas para la detección temprana del comienzo de desarrollo de estas colonias en zonas no colonizadas de la cueva de Altamira. De acuerdo con esta estrategia, se han diseñado al menos un par de primers para cada uno de los componentes mayoritarios característicos de cada tipo de colonia. Para el diseño de estos primers se han utilizado genes presentes solamente en cada uno de los genomas de estos componentes (Tabla 5). Los genes seleccionados para diseñar los primers han sido genes codificantes de proteínas, excepto para el MAG082 que en cuyo caso el target elegido fue una región intrónica presente en el 23S rRNA que sólo se ha localizado en este componente mayoritario. Los programas usados y la metodología seguida para la búsqueda de estos targets se encuentran descritos en el apartado de materiales y métodos.

Por tanto, excepto en el caso del MAG082, la elección de los targets se hizo en base a que fuera la única proteína descrita en su familia, sin más proteínas homólogas, siempre cumpliendo que en los metagenomas de Altamira sólo se detectara en el genoma del componente contra el que se estaba diseñando. Para comprobar que estos primers sólo hibridarían con cada componente mayoritario, se hizo un blast de la secuencia de los primers diseñados contra la base de datos que contiene los contigs de todos los metagenomas (YC,W1,W2,G1,G2 y BC) que contiene tanto los contigs de los componentes mayoritarios como los de los secundarios. Las secuencias de cada uno de los primers se recogen en el Anexo-Tabla S9.

Contig	Gen	Tipo de colonia	Tamaño target	Target	Cov	Identidad	ID accesion
<b>CLUSTER PSEUDONOCARDIACEAE</b>							
NODE_101_length_27352_cov_42.187012_25	MAG088_55860	<b>YC</b>	198	Pyrimidine reductase family protein	93	49.59	WP_040836632.1
NODE_4_length_219108_cov_56.835965_118	MAG073_09480	<b>W1</b>	173	UbiA family prenyltransferase	97	62.16	WP_158886769.1
NODE_272_length_17251_cov_43.970852_4	MAG098_00060	<b>BC</b>	297	Phage tail sheath family protein	100	68.41	WP_029337936.1
NODE_118_length_51204_cov_153.155_21	MAG001_13070	<b>G1</b>	230	CopD family protein	93	53.02	WP_037317570.1
NODE_18_length_131575_cov_137.795_15	MAG046_01690	<b>G2</b>	411	Hypothetical protein	100	51,09	WP_162424531.1
<b>CLUSTER NITRILIRUPTORIA</b>							
NODE_4_length_93391_cov_135.934123	-	<b>W2</b>	150	Intron in the 23S rRNA	-	-	-
NODE_470_length_11208_cov_107.744577_6	MAG097_00060	<b>BC</b>	280	Hypothetical protein	82	34.80	MBE1533496.1

Tabla 5. Targets seleccionados para la detección primaria de los componentes mayoritarios que corresponden a cada tipo de colonia.

Cov:Coverage

Los resultados de la comparación con Blastn de los primers diseñados para cada componente fueron muy selectivos. Los primers diseñados para el MAG088 (YC) dieron sólo un hit con el contig 101 que pertenece al componente mayoritario de las colonias amarillas, y el producto de amplificación obtenido fue una banda de 198 pb. Los primers diseñados para el MAG073 (W1) dieron también con un solo hit en el contig 4 que pertenece al componente mayoritario de las colonias blancas subtipo W1, y el producto de amplificación correspondió con una banda de 173 pb. Los primers diseñados para el MAG001 (G1) también tuvieron una alta selectividad dando un hit con el contig 118, que pertenece al componente mayoritario de las colonias grises subtipo G1, y el producto de amplificación fue una banda de 230 pb. Los primers diseñados para el MAG046 (G2) también dieron un hit único con el contig 18, que pertenece al componente mayoritario de las colonias grises subtipo G2, y el producto de amplificación correspondió con una banda de 411 pb.

Los primers seleccionados para los componentes mayoritarios de las colonias beis ofrecen una situación más compleja. Los primers diseñados para el MAG098 se produjeron dos hits, uno con el contig 272 que pertenece al componente mayoritario de las colonias beis y otro con el contig 2453 que pertenece al secundario MAG060 (G2). El producto de amplificación fue una banda de 297 pb. La aparición de dos hits se debe a que este componente mayoritario MAG090 se ha detectado como secundario en el metagenoma de las colonias grises subtipo G2. En el caso del otro componente mayoritario de las colonias beis, el MAG097, se obtuvieron tres hits, uno con el contig 470 que forma parte del componente mayoritario de las colonias beis MAG097, y otros dos con el contig 1018 que forma parte del secundario MAG090 (YC) y con el contig 1501 que forma parte del secundario MAG050 (G2). La aparición de tres hits en este caso se debe a que este componente mayoritario se ha detectado como secundario en los metagenomas de colonias amarillas (MAG090) y en colonias grises subtipo G2 (MAG050). El producto de amplificación para estos últimos primers fue una banda de 280 pb.

Los primers en el componente mayoritario de las colonias blancas subtipo W2, correspondiente al MAG082, fueron diseñados aprovechando la presencia de intrones dentro del gen del 23S rRNA. Estos intrones cuando se eliminan permiten que el 23S rRNA maduro, pueda formar las subunidades ribosomales que darán lugar al ribosoma, esencial para la síntesis de proteínas. Ejemplos concretos de intrones en el gen del 23S rRNA se han detectado en bacterias del filo *Proteobacteria* como *Brucella*, *Coxiella*, *Campylobacter*, *Haemophilus*, *Helicobacter*, *Proteus*, *Salmonella* o *Yersinia* (Pronk and Sanderson, 2001). También se han detectado en otros filos como *Actinobacteria* (Crapoulet et al., 2005), *Cyanobacteria*, *Deinococcus-Thermus* o *Firmicutes* (Pei et al., 2009).

En el componente mayoritario de las colonias blancas subtipo W2, el gen del 23S rRNA (3065 pb sin intrones) se encuentra interrumpido por 2 regiones intrónicas una de 389 pb y una segunda de 309 pb, lo que genera tres posibles exones uno de 2083 pb, otro de 38 pb y un tercero de 944 pb, utilizando como modelo el mejor hit que fue contra *Egibacter rhizosphaerae* strain EGI 80759 con una id de 88,50% (Figura 64). Las comparaciones de la región completa del 23S con el programa blastn indicaban la ausencia de las dos regiones intrónicas en otras secuencias en las bases de datos.

Estos exones tampoco están presentes en el MAG097 que es con el MAG082 el otro componente de la clase *Nitriliruptoria*. Los primers por tanto se han diseñado el forward en el intrón 1 y el reverse en el intrón 2. Los resultados del blastn de estos primers dieron 2 hits, con el contig 4 que pertenece al componente mayoritario de las blancas subtipo W2, y con el contig 4977 que pertenece a colonias amarillas pero no se encuentra bineado en ningún MAG. El producto de amplificación corresponde con una banda específica de 150 pb.

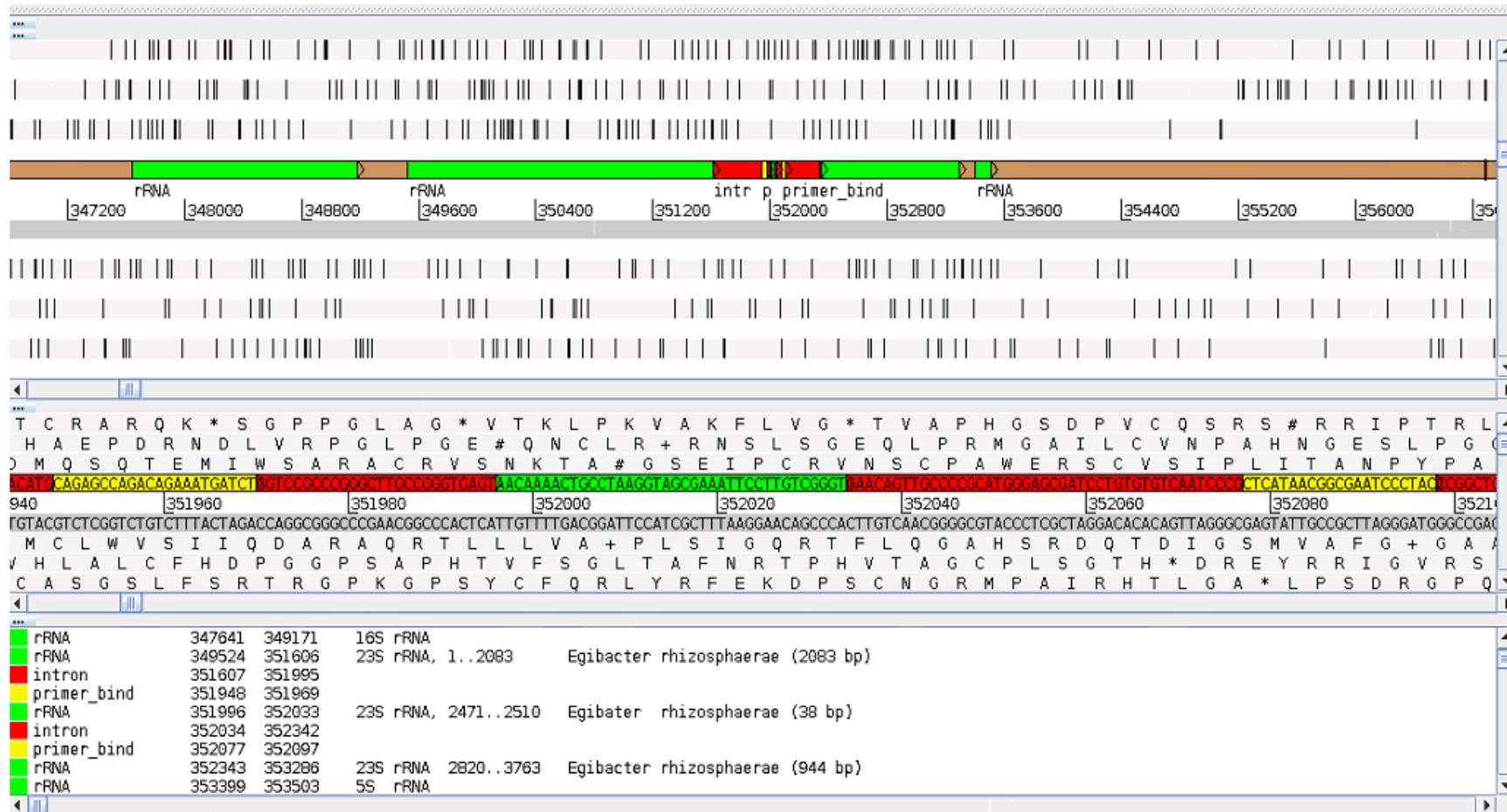


Figura 64. Representación de la estructura del gen 23S rRNA en el genoma del MAG082, componente mayoritario de las colonias W2 intervenido por 2 regiones intrónicas. En color verde los genes ribosomales 16S, 23S y 5S rRNA, en rojo las regiones intrónicas, en amarillo los primers diseñados para su amplificación y en marrón el contig.

Una vez diseñado los primers para la detección de cada componente, se han ensayado en una reacción de PCR con muestras de al menos de un ejemplo de cada subtipo de colonia, para la confirmación de la obtención de una sola banda correspondiente al componente mayoritario. Se han hecho 3 reacciones de PCR independientes, siendo la PCR1 y 2 discriminatoria de los componentes que conforman el cluster III correspondiente a la familia *Pseudonocardiaceae*, y la PCR3 discriminatoria de los componentes que conforman el cluster IV correspondiente a la clase *Nitriliruptoria* (Figura 65).

Los primers para la PCR1 ensayados fueron los primers diseñados contra el MAG088 (componente mayoritario de las colonias amarillas) que da una banda de 198 pb, del MAG073 (componente mayoritario de las colonias blancas subtipo W1) que da una banda de 173 pb, del MAG098 (componente mayoritario de las colonias beis) que da una banda de 297 pb, y además de primers universales (PA8F y 1100R) que dan una banda de 1100 como control de la presencia de DNA cargado en cada muestra analizada. Los primers ensayados para la PCR2 fueron los primers diseñados contra los 2 subtipos de colonias grises, del MAG001 (componente mayoritario de las colonias grises subtipo G1) que da una banda de 230 pb, del MAG046 (componente mayoritario de las colonias grises subtipo G2) que da una banda de 411 pb y los mismos primers universales ensayados para la PCR1 como control. Finalmente, los primers ensayados para la PCR3 fueron los primers del MAG082 (componente mayoritario de las colonias blancas subtipo W2) que da una banda de 150 pb, del MAG097 (componente mayoritario de las colonias beis) que da una banda de 280 pb y de nuevo los mismos primers universales ensayados para la PCR1 y 2 como control.

Cabe destacar que, en estos ensayos, en general los primers seleccionadas fueron diagnóstico selectivo para el tipo de colonia para el que fueron diseñados sin ofrecer problemas de amplificación heteróloga. La selectividad de los primers fue esperada excepto en los casos que implicaban amplificación de los CMs de las colonias beis: MAG098 en la PCR1 y MAG097 en la PCR3.

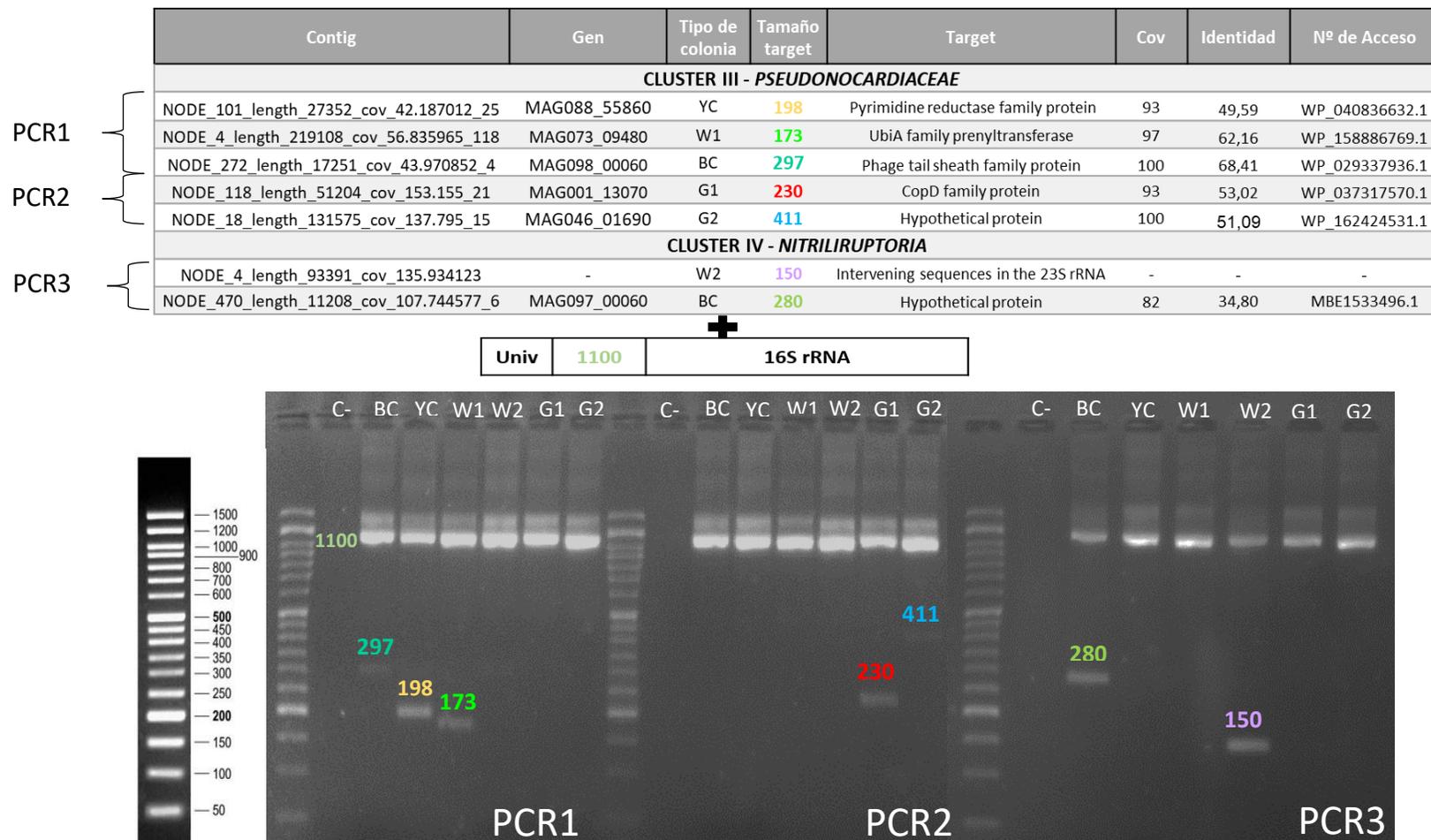


Figura 65. Imagen de los resultados del ensayo de primers diseñados para cada componente mayoritario con las distintas muestras de colonias por la técnica de PCR. PCR1 se ha llevado a cabo con los primers diseñados para la detección del MAG088 (YC), MAG073 (W1) y MAG098 (BC) más primers universales; PCR2 se ha llevado a cabo con los primers diseñados para la detección del MAG001 (G1) y MAG046 (G2) más primers universales; PCR3 se ha llevado a cabo con los primers diseñados para la detección del MAG097 (BC) y MAG082 (W2) más primers universales.

Como se describió más arriba estos primers ofrecieron dos y tres hits respectivamente en ensayos “*in silico*” debido a la aparición doble como CM en colonias beis y como componente secundario en otras colonias. Sin embargo, en los ensayos de PCR realizados únicamente mostraron amplificación en las colonias beis demostrando que pueden ser utilizados como diagnóstico de estas colonias. Para una detección precoz de los componentes MAG097 y MAG098 como componentes secundarios o antes de formar parte del consorcio como colonias visibles, sería posible utilizar una técnica con mayor poder de detección como la PCR a tiempo real y ajustar los ciclos de amplificación y concentración de DNA molde.

Los primers y PCRs descritos constituyen una prueba de concepto para ilustrar las posibilidades que las secuencias de los metagenomas obtenidos pueden ofrecer para diseñar ensayos de análisis o herramientas de control para el estudio o seguimiento de la proliferación de colonias visibles en Altamira y posiblemente en otras localizaciones donde se encuentran.

### **5.2.6 Capacidad codificante y análisis funcional de los metagenomas**

Como paso previo para obtener una imagen completa de las capacidades metabólicas de las comunidades microbianas presentes en las colonias, hemos llevado a cabo una predicción de los ORFs presentes en cada metagenoma seguida de su anotación funcional y clasificación en grupos de ortólogos. Los ORFs predichos por “Prodigal” en cada genoma fueron identificados mediante la asignación de un identificador propio con la siguiente anotación, MAGXYZ\_abcd0, donde MAGXYZ indica el MAG en el que se localiza el ORF y abcd0 es un numeral consecutivo en incrementos de 10 que se asigna secuencialmente.

Para los ORFs codificados en contigs no clasificados en MAGs usamos la anotación MAG0XX\_abcd0, donde simplemente indicamos el metagenoma (tipo de colonia) en la que se localiza cada ORF, así encontraremos MAG0BC para las colonias beis, MAG0YC para las amarillas, y MAG0G1, MAG0G2, MAG0W1 y MAG0W2 para las restantes. Con este procedimiento hemos

anotado la posición y asignado un ID propio a 609.009 ORFs distribuidos en los 6 metagenomas como se muestra en la Tabla 6.

De los 609.009 CDS predichos, 139.876 (23%) no encontraron ningún ortólogo por lo que se clasificaron como “proteínas hipotéticas”. El resto (77%) obtuvieron algún tipo de anotación, aunque en algunos casos solo fuera una indicación de la probable posición taxonómica del ORF. El número de ORFs que finalmente debemos clasificar como proteínas sin función conocida resulta de sumar las que no encontraron ningún ortólogo (proteínas hipotéticas) más las que no obtuvieron clasificación funcional (proteínas con ortólogos, pero sin función) más las que fueron clasificadas en la categoría S (función desconocida) lo que entonces sumaría un total de 258.381 (42,43%) del total de los ORFs predichos.

Estos porcentajes no variaron de forma significativa en los diferentes metagenomas, así los porcentajes de ORFs totales sin función oscilaron entre un 40% mínimo en el metagenoma G1 y un máximo de 45,57% en el metagenoma W1 (Figura 66).

	BC	YC	W1	W2	G1	G2
<b>ORFs predichos</b>	27.140	58.623	57.227	38.159	210.722	217.138
<b>ORFs anotados</b>	21.584	44.647	41.607	29.439	169.051	162.805
<b>Prots hipotéticas</b>	5.556	13.976	15.620	8.720	41.671	54.333
<b>ORFs anotados con función</b>	20.118	41.517	38.949	27.770	159.454	153.489
<b>ORFs anotados sin función</b>	1.466	3.130	2.658	1.669	9.597	9.316
<b>ORFs totales sin función</b>	11.156	25.741	26.079	15.938	84.268	95.199
<b>ORFs totales con función verdadera</b>	15.984	32.882	31.148	22.221	126.454	121.939

Tabla 6. Tabla del número total de ORFs con/sin función en cada metagenoma de colonia. Las proteínas hipotéticas son la diferencia de los ORFs predichos y los ORFs anotados. Los ORFs totales sin función es la suma de las proteínas hipotéticas más las que no obtuvieron clasificación funcional (ORFs anotados sin función) más las que fueron clasificadas en la categoría S (ORFs anotados con función desconocida). Los ORFs con función verdadera es la diferencia de los ORFs predichos y los ORFs totales sin función.

La información recogida en estas anotaciones puede proporcionar información clave sobre ecosistemas dominados por la oligotrofia como ocurre en la cueva de Altamira. Se utilizaron las asignaciones de COG para predecir y clasificar posibles funciones de las secuencias únicas por homología con secuencias de ortólogos. Un total de 350.628 ORFs (57,57%) fueron finalmente mapeados en 19 categorías COG diferentes y se representó su porcentaje normalizado con el número de ORFs predichos en cada metagenoma (Figura 66). Del total de ORFs mapeados en categorías funcionales, únicamente 601 ORFs (0,1%) fueron asignados a categorías de genes eucarióticos (categorías W, Z y B) que no se muestran en la gráfica.

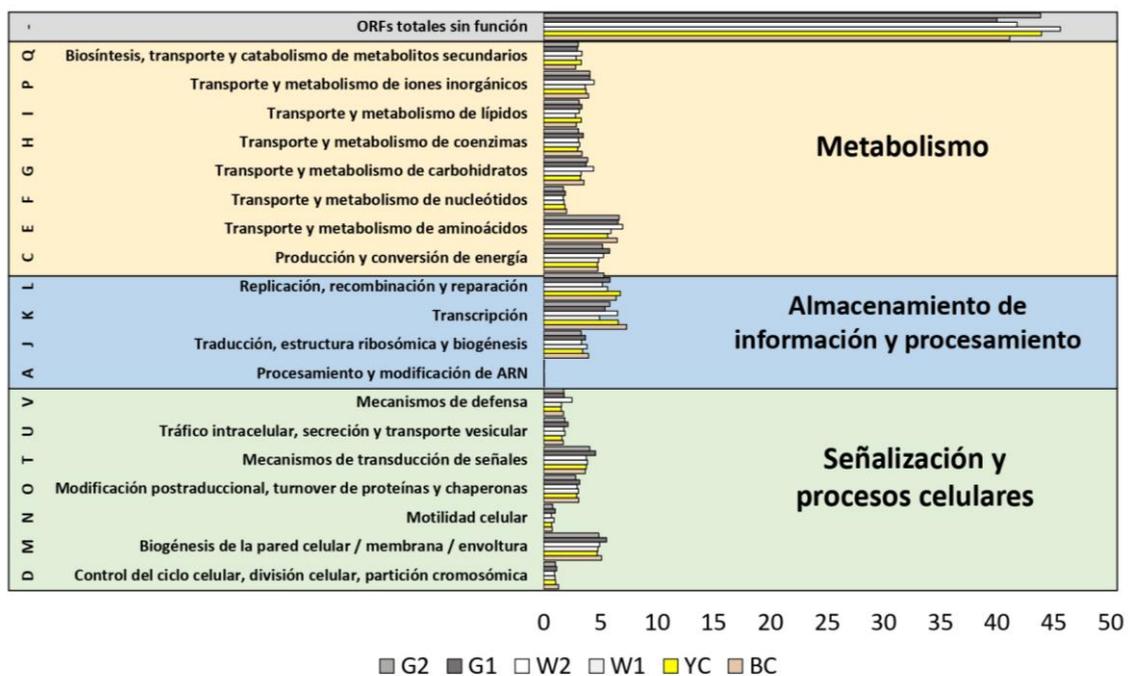


Figura 66. Lista de las categorías funcionales de los grupos de ortólogos anotados en los metagenomas de las colonias visibles.

La comparación funcional general basada en las categorías COG mostró una gran uniformidad en todos los metagenomas. Llama la atención que la distribución de categorías funcionales de los ORFs en los 6 metagenomas es prácticamente igual, lo que sugiere una adaptación muy alta al medio y que las diferentes colonias poseen básicamente las mismas capacidades metabólicas y fisiológicas.

Además del 40% de ORFs sin una asignación funcional clara, es notable la cantidad de genes dedicados al transporte y metabolismo, especialmente a procesos energéticos y transporte y metabolismo de aminoácidos, lo que sugiere que su disponibilidad puede ser esencial en el crecimiento activo de las colonias. También es elevado el número de ORFs dedicados a replicación, recombinación y reparación del DNA. Son muy abundantes los productos implicados en mecanismos de reparación y especialmente los dedicados a superar el stress oxidativo, presumiblemente generado al uso generalizado de la respiración aerobia. En la discusión haré un comentario sobre los genes detectados y su posible papel en el metabolismo de las colonias.

Finalmente, mencionaré la presencia de un 4% de genes involucrados en el metabolismo secundario de los que muchos van a ser responsables de la biosíntesis de antibióticos y otros compuestos bioactivos, que además de jugar un papel esencial en las interacciones entre organismos en la cueva, pueden tener gran interés industrial o biotecnológico.

### **5.3 Cultivo de componentes de las colonias**

Con la metodología descrita se lograron obtener 136 cultivos puros a partir de las muestras de colonias de los cuales 52 se obtuvieron de colonias amarillas, 34 de colonias grises, 31 de colonias blancas y 19 de colonias beis (Anexo-Tabla S10). La mayor parte de estos aislados se obtuvieron en medio de actinomicetos (AIA) con un total de 65, de los cuales 52 fueron recuperados a una temperatura de 28°C, 12 restantes lo fueron a una temperatura de 16°C y 80% de humedad y 1 a una temperatura de 16°C. Otro de los medios empleados fue TSA, en el que se obtuvieron 36 aislados, 35 a una temperatura de 28°C y sólo un aislado a una temperatura de 16°C. Finalmente, en medio TSA 0,1X se obtuvieron 23 aislados, 18 a una temperatura de 28°C y 5 a una temperatura de 26°C, y TSA 0,1X con NaHCO<sub>3</sub> también a 28°C con 10 aislados y a 16°C con 2 aislados.

Se amplificó y secuenció el 16S RNA de estos aislados por la técnica de Sanger. Las secuencias de los 16S rRNAs obtenidas fueron comparadas con las presentes en las bases de datos para proceder a su asignación taxonómica. En cuanto a los taxones obtenidos entre estos cultivables, *Actinobacteria* fue el filo más aislado con 95 bacterias. Dentro de este filo se detectaron 5 órdenes que fueron *Streptomycetales* (53 aislados), *Corynebacteriales* (19), *Micrococcales* (16), *Pseudonocardiales* (6) y *Micromonosporales* (1). Dentro del orden *Streptomycetales*, todos los aislados pertenecieron al género *Streptomyces* (53). En el orden *Corynebacteriales* se detectaron 2 géneros *Nocardia* (12) y *Rhodococcus* (7). En el orden *Micrococcales* fueron 5 géneros correspondientes a *Microbacterium* (5), *Agromyces* (4), *Micrococcus* (4), *Kocuria* (2) y *Frigoribacterium* (1). Dentro del género *Pseudonocardiales* se detectaron 2 géneros que fueron *Crossiella* (3) y *Pseudonocardia* (3). Finalmente, en el orden *Micromonosporales* sólo se obtuvo un aislado correspondiente con la bacteria *Micromonospora matsumotoense*.

El segundo filo más abundante correspondió con *Firmicutes* con 25 aislados de los cuales la gran mayoría pertenecieron al orden *Bacillales* (24) y sólo 1 a *Lactobacillales*. Dentro de la clase *Bacillales* se detectaron 3 géneros correspondientes con *Bacillus* (20), *Paenibacillus* (2), *Psychrobacillus* (1) y *Sporosarcina* (1), mientras que en el orden *Lactobacillales* correspondió con la bacteria *Facklamia tabacinasalis*.

El tercer filo fue *Proteobacteria* con 15 aislados dónde se detectaron las clases *Alphaproteobacteria* (11) y *Gammaproteobacteria* (4). Dentro de la clase *Alphaproteobacteria* se detectaron 3 órdenes que fueron *Rhizobiales* (6), *Rhodospirillales* (3) y *Sphingomonadales* (2). Dentro del orden *Rhizobiales* se detectaron 3 géneros *Phyllobacterium* (4), *Aminobacter* (1) y *Agrobacterium* (1). En el orden *Rhodospirillales*, 2 aislados fueron del género *Inquillinus*, mientras que un aislado clasificaba al 90% de identidad con 2 géneros *Dongia* y *Reyranella*. En el orden *Sphingomonadales* fueron 2, de los cuales uno correspondió con el género *Sphingobium* mientras que el otro aislado clasificaba al 99,43% de identidad con 2 géneros *Sphingomonas* y

*Blastomonas*. Por el contrario, dentro de la clase *Gammaproteobacteria* todos los aislados pertenecieron al género *Pseudomonas* (4).

El cuarto filo detectado fue *Bacteroidetes* con un solo representante correspondiente a la bacteria *Sphingobacterium faecium*, aislado en medio AIA a 16°C y 80 % humedad.

La mayor parte de los aislados del filo *Actinobacteria* (95) se recuperaron en medio AIA con 49 cultivables, donde se detectaron los órdenes *Pseudonocardiales* (6), *Micromonosporales* (1), la mayor parte de los aislados de *Micrococcales* (8) y *Corynebacteriales* (13), y 21 del orden *Streptomyetales*. El segundo medio que más actinomicetos aisló fue el medio rico TSA con 21 aislados, entre los que se detectaron los órdenes *Micrococcales* (3) y *Corynebacteriales* (2), siendo la mayoría de aislados en este medio al orden *Streptomyetales* (16). El tercer medio fue TSA diluido 0,1X con 19 cultivables, entre los que se detectaron los órdenes *Streptomyetales* (13), *Micrococcales* (4) y *Corynebacteriales* (2). El medio TSA diluido 0,1X con  $\text{NaHCO}_3^-$  tuvo 6 aislados de *Streptomyetales* (3), *Corynebacteriales* (2) y *Micrococcales* (1).

Respecto al filo *Firmicutes* (25), el medio que más bacterias aisló de estos cultivables fue TSA (15), en el que se detectaron los órdenes *Bacillales* (14) y *Lactobacillales* (1). El resto de aislados del orden *Bacillales* se detectaron en AAI (7) y en TSA diluido 0,1X (3).

Dentro de la clase *Alphaproteobacteria* (11), el medio más satisfactorio para su aislamiento fue TSA diluido 0,1X con  $\text{NaHCO}_3^-$  en el que se detectaron los órdenes *Rhizobiales* (4) y *Sphingomonadales* (2). El medio AIA fue el segundo que más aisló esta clase detectándose los órdenes *Rhizobiales* (2) y *Rhodospirillales* (2). Finalmente, en medio TSA 0,1X sólo se recuperó un aislado del orden *Rhodospirillales*. Dentro de la clase *Gammaproteobacteria* (4) todos los aislados del género *Pseudomonas* se recuperaron en el medio AIA.

Las secuencias de los 16S RNAs de los aislados fueron alineadas con el programa ClustalW junto con los 16S rRNAs de bacterias de referencia más cercanas a ellas. Después con las secuencias alineadas, se ha construido un árbol filogenético por el método de máxima verosimilitud basado en el modelo Tamura-Ne con el software de MEGA v7.0.26 (Figura 67). La información de las bacterias de referencia utilizadas en este árbol se encuentra recogida en el Anexo-Tabla S11.

Ninguno de los aislados obtenidos en los intentos de cultivo coincide con alguno de los componentes esenciales descritos como componentes tanto mayoritarios como secundarios para las colonias, a juzgar por los bajos índices de identidad que mostraron por la comparación de sus 16S rRNAs (<97%). Esto no quiere decir que las bacterias recuperadas de muestras de colonias no puedan ser posibles componentes minoritarios de las colonias, aunque creemos mucho más probable que se trate de bacterias que forman parte de la microbiota del material soporte de la pared.

Por su interés y especial relación con componentes coloniales hemos seleccionado 4 bacterias para la secuenciación total de su genoma. Dos de ellas que se identificaron con las bacterias *Crossiella cryophila* (VC61) y *Pseudonocardia eucalypti* (VC74) fueron seleccionadas principalmente por pertenecer a la misma familia que los CMs, concretamente del cluster III. La tercera bacteria seleccionada fue clasificada con la especie *Inquilinus ginsengisoli* (VC62), y su selección se debió a que un componente secundario de las colonias grises subtipo G1 (MAG055) fue asignado a este mismo género, aunque comparaciones posteriores de la secuencia del genoma reflejan un valor de ANI del 89,95%, que sugiere que no se tratan de la misma bacteria (Figura 55). La última bacteria seleccionada fue clasificada con la especie *Pseudomonas chlororaphis* (VC102) y su selección fue debida al pigmento de color anaranjado que secreta durante su crecimiento en medio de cultivo que lo hacía reminiscente de las colonias amarillas.

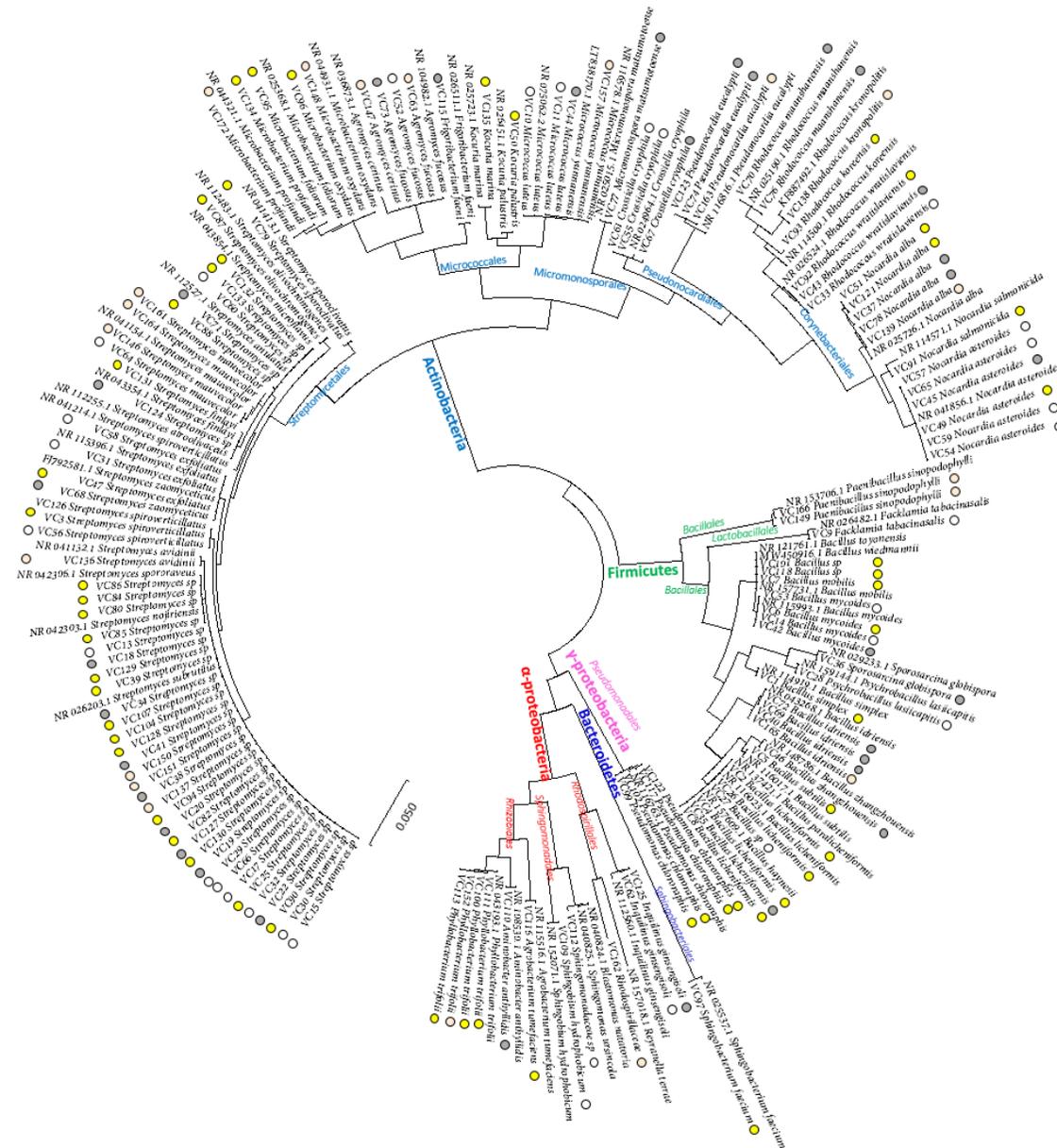


Figura 67. Análisis filogenético molecular por el método de máxima verosimilitud con el software de MEGA7 de los 16S rRNA de los aislados cultivables de las colonias visibles junto con las bacterias de referencias más cercanas de la base de datos del NCBI. La historia evolutiva se infirió utilizando el método de máxima verosimilitud basado en el modelo Tamura-Nei. Se muestra el árbol con la mayor probabilidad logarítmica. El árbol está dibujado a escala, con las longitudes de las ramas medidas en el número de sustituciones por sitio. La proporción de sitios donde al menos 1 base inequívoca está presente en al menos 1 secuencia para cada clado descendiente se muestra junto a cada nodo interno en el árbol. El análisis involucró 196 secuencias de nucleótidos.



## 6. Discusión



## 6.1 La cueva de Altamira y su situación actual

La cueva de Altamira destaca en su ámbito por la singularidad y la calidad del arte que alberga, así como por la buena conservación que presentan sus pinturas, por lo que a día de hoy ostenta la categoría de icono del arte rupestre paleolítico. Además, fue el primer lugar en el mundo donde se identificó arte rupestre del periodo Paleolítico Superior. El descubrimiento de las pinturas en la cueva fue tan impactante para la sociedad en ese momento, que con los años la cueva se convirtió en objeto de estudio para científicos de una gran variedad de campos y desde el principio atrajo la atención de visitantes y curiosos. Como consecuencia, la cueva ha sido constantemente modificada estructuralmente y estudiada con el objetivo de una mayor comprensión (Figura 10). Muchas de estas modificaciones estructurales en la primera etapa del descubrimiento de la cueva fueron orientadas a las visitas que comenzaron a partir del año 1875. Estas visitas, especialmente en los primeros años, se tradujeron en entradas masivas y constantes a la sala de las pinturas.

Por otro lado, la cueva de Altamira, como cualquier otro entorno natural de tipo kárstico sufre un proceso continuo de alteración geoquímica. La contemplación de veladuras de carbonato cálcico en muchos puntos de la cueva, así como la infiltración de agua de escorrentía por la gran cantidad de fisuras o grietas que presentan los techos de la cueva, favorecen la pérdida del pigmento por lavado o la desaparición del mismo. Estos fenómenos naturales no son controlables y en su mayor parte son irreversibles.

Igualmente, las cuevas también son hábitats naturales para microorganismos que juegan un papel importante en la dinámica y los ciclos geoquímicos de éstas. Lo que vamos a encontrar en la cueva son microorganismos bien adaptados al ambiente subterráneo, que han evolucionado hasta alcanzar un equilibrio estable pero delicado. Mientras las condiciones sean constantes, las poblaciones microbianas no experimentan grandes cambios. Pero de la misma manera, alteraciones ambientales o nutricionales pueden dar lugar a cambios en la microbiota normal con efectos impredecibles. La cueva de Altamira tras su descubrimiento, ha experimentado estos cambios, y así ha sufrido

colonizaciones espontáneas de microorganismos como la aparición de comunidades fotosintéticas en zonas de pinturas (algas o cianobacterias) o la presencia de rizomorfos en estructuras artificiales. Como colonización microbiana más relevante se detecta alrededor del año 1999 la presencia de colonias visibles desarrolladas en los techos y paredes de la cueva, que se originaron y son más abundantes cerca de la entrada, pero que con los años su desarrollo ha ido en avance hacia el interior de la cueva alcanzando incluso la sala de Polícromos. El análisis de riesgo elaborado en el Plan de Conservación Preventiva de Altamira considera que la presencia de colonias visibles es el mayor riesgo biológico que amenaza las pinturas de la cueva. Inicialmente las colonias visbles en la cueva de Altamira eran de tres tipos denominadas en función de la pigmentación que presentaban como colonias amarillas, grises y blancas. En el transcurso de este trabajo se identifica un cuarto tipo de colonia visible con un color distintivo diferente, al cual se ha denominado colonias beis.

La preocupación por los cambios microbiológicos que está sufriendo Altamira no son nuevos. En los años 70 y 80 del siglo veinte se impulsaron los primeros estudios microbiológicos en la cueva, los cuales consistían generalmente en el cultivo de los componentes microbianos presentes en muestras de agua, tierra o soporte, así como el análisis del efecto que las entonces masivas visitas pudieran tener sobre los microorganismos presentes. Con la presencia de las colonias visibles, se amplió el estudio a ellas, pero pronto se observó que las técnicas habituales de cultivo no eran satisfactorias para la descripción precisa de los habitantes microbianos de Altamira.

Por ello, estudios posteriores llevados a cabo a partir del año 2000, sobre todo para estudiar las colonias visibles, se hicieron con técnicas de secuenciación sin necesidad de cultivo. Estos primeros análisis describieron la presencia de bacterias de los filos *Actinobacteria* y *Proteobacteria* como claros protagonistas en estas muestras. Otros filos de menor abundancia detectados en las colonias fueron *Acidobacteria*, *Bacteroidetes*, *Chloroflexi*, *Gemmatimonadetes*, *Firmicutes*, *Candidate NC10*, *Nitrospirae*, *Planctomycetes*, *Verrucomicrobia*, y especialmente en colonias blancas *Cyanobacteria* y *Candidate division SPAM*, y en colonias amarillas *Candidate phylum SBR1093*. A parte, se siguieron

haciendo intentos de cultivo de componentes de las colonias lo que permitió la descripción de nuevas especies de bacterias aisladas de colonias grises *Nocardia altamirensis* (*Actinobacteria*, Jurado et al., 2008), y de colonias blancas *Aurantimonas altamirensis* (*Alphaproteobacteria*, Jurado et al., 2006) y *Hoyosella altamirensis* (*Actinobacteria*, Jurado et al., 2009b).

A día de hoy, con las nuevas técnicas de secuenciación masiva se puede obtener una mayor cantidad de secuencias por muestra lo que supone una diferencia cualitativa bastante grande en comparación con las primeras técnicas utilizadas. Esta diferencia es una ventaja en la resolución de problemas microbiológicos como los que atañen a la cueva de Altamira. Esta tesis se plantea con el objetivo de mejorar nuestro conocimiento sobre la composición de los componentes de las colonias empleando nuevas técnicas de secuenciación masiva hasta ahora no utilizadas en la cueva de Altamira. Estas técnicas fueron la técnica de amplicones, en la que se secuencian el gen del 16S rRNA de todos los componentes presentes en la muestra si el esfuerzo de secuenciación empleado es el adecuado, y la técnica de shotgun que nos permitirá obtener borradores, al menos de los genomas de los componentes más abundantes en las colonias.

## 6.2 Composición de las colonias

Las técnicas empleadas en este trabajo nos han proporcionado una cantidad de datos de secuencia muy superior al total de la disponible hasta ahora sobre la población microbiana de Altamira. Por la técnica de amplicones hemos obtenido 5.697.876 lecturas totales (de unos 400 nt cada una) correspondientes a los segmentos hipervariables V3-V4 del RNA ribosomal 16S. La técnica de secuenciación total de shotgun nos ha proporcionado más de 500 Mb de secuencias genómicas en fragmentos mayores de 1 kb correspondientes a los metagenomas de los seis tipos de colonia visible analizadas.

En el caso de la técnica de amplicones, los pasos previos a la secuenciación como son la amplificación por la técnica de PCR de la región de interés y la

preparación de librerías, y pueden introducir errores que se traducirán en unos resultados erróneos sobre la diversidad microbiológica o la abundancia relativa de los componentes de las muestras. Todos nuestros análisis han incluido controles experimentales consistentes en mezclas comerciales con distinta abundancia y composición de bacterias (ZymoBiomics Microbial Community Standards). Los resultados que se obtuvieron sistemáticamente con estos controles incluidos en todos nuestros ensayos fueron los esperados para la composición conocida de las muestras control, sirviendo al menos en parte para asegurar la fidelidad de nuestros resultados.

Este trabajo parte con el objetivo más esencial de averiguar la diversidad real presente en las colonias visibles, saberlo será fundamental para el desarrollo de medidas de control. En estudios previos en Altamira en colonias, la riqueza detectada en colonias amarillas fue de 42 OTUs, sin embargo, el número esperado de especies basado en abundancias con los índices ACE y Chao1 fue de 291 y 172 respectivamente (Porca et al., 2012). Los estudios de rarefacción aplicados a estos resultados mostraban que las curvas de acumulación no llegaban a alcanzar una saturación, sugiriendo que se necesitaba un mayor esfuerzo de secuenciación para alcanzar una descripción completa. En otro estudio en Altamira en colonias blancas, el número de OTUs detectado fue 47 y tampoco la curva de rarefacción alcanzó una asíntota (Portillo Guisado et al., 2009). Estas estimaciones de la diversidad en colonias concuerdan con el resultado final que hemos obtenido tras el análisis detallado de OTUs y que formaron un core en las colonias, aunque no todas las OTUs del core han sido detectadas en los estudios previos.

Por la técnica de amplicones en este estudio hemos obtenido 4.288 OTUs (Tabla1) a partir de 57 muestras analizadas, correspondientes a 55 muestras de colonias de los cuatro tipos descritos y dos muestras de soporte sin colonias aparentes. La diversidad obtenida en las colonias en este estudio ha sido bastante superior a lo observado en los estudios previos, y las curvas de rarefacción alcanzaron la saturación en todas las muestras (Figura 29) indicando que la profundidad de secuenciación obtenida había sido suficiente. La riqueza media detectada en colonias beis fue de 323 OTUs, 618 en colonias

grises, 351 en colonias blancas y 382 en colonias amarillas. Los índices de diversidad obtenidos indicaron para las colonias una riqueza media (Figura 30), inferior a la que suele obtenerse habitualmente en muestras de suelo y soporte (Anexo-Figura S2). Además, se detectaron a las colonias grises como las muestras con mayor riqueza y diversidad entre las muestras de colonias.

Este estudio aporta evidencia de la existencia en las colonias de unos componentes (uno o dos) predominantes, al menos 10 veces más abundantes que el resto y que son característicos en cada colonia (Figura 34). La técnica de amplicones identifica tres de estos componentes, a los que hemos denominado componentes mayoritarios de las colonias. Su distribución desigual en las colonias han sido la base para diferenciar en 2 subtipos a las colonias grises y blancas. La existencia de estos subtipos quedó confirmada con los análisis de  $\beta$ -Diversidad (Figura 31). Esta heterogeneidad detectada en muestras de colonias aparentemente del mismo color no se ha mencionado anteriormente en otros estudios, indicando que la clasificación por color no es suficiente para describir la diversidad de tipos de colonia existentes.

Aunque el tamaño de los amplicones que se usan en esta técnica es pequeño, permite una razonable asignación taxonómica de las OTUs por comparación con las secuencias de los 16S rRNAs presentes en las bases de datos (SILVA). Tras esta asignación tenemos una imagen general de la composición en taxones de las colonias que esta dominada por el filo *Actinobacteria* (43,36% de las secuencias), seguido de *Proteobacteria* (25,87%), *Acidobacteria* (9,70%) y *Plantomyces* (6,41%). Filos con menor frecuencia detectados fueron *Dadabacteria* con 3,22%, *Chloroflexi* con 3,08%, *Rokubacteria* con 2,34% y *Nitrospirae* con 1,20%. Es de notar que ninguno del resto de los Phyla detectados alcanza una riqueza equivalente al 5% de las secuencias. Las concentraciones de la mayor parte de estos componentes minoritarios son equivalentes en las muestras de colonias y las muestras de sustrato sin colonias, lo que arroja dudas sobre su pertenencia real a los consorcios de las colonias.

Por esta razón y por una tendencia a considerar las colonias como consorcios no excesivamente complejos, seleccionamos un core de 36 OTUs, que cumplieren la condición de que su abundancia media en cada subtipo de colonia fuera >1% y que la OTU estuviera presente en al menos 2 muestras de colonias (Figura 34). Dentro de esta selección encontramos como las más abundantes de todas, a las OTUs correspondientes a los componentes mayoritarios clasificándolas dentro del filo *Actinobacteria* con la familia *Euzebyaceae* (OTU1) y con el género *Crossiella* (OTUs 2 y 4).

La presencia de colonias visibles, al igual que los componentes identificados no son exclusivos de la cueva de Altamira, sino que también se han detectado previamente en otros estudios en cuevas. Los trabajos anteriores en colonias de Altamira habían identificado varios de los componentes que describimos ahora. Secuencias con una identidad mayor al 99% con la OTU2 se obtuvieron de colonias amarillas en la cueva de Altamira (clones AY960224.1, AY960256.1, AY960236.1, Portillo et al., 2008) y en otro estudio hecho en colonias amarillas en 3 cuevas distintas entre las que se encuentra nuevamente Altamira (clones HE602849.1 y HE602845.1, Porca et al., 2012), así como en otras cuevas como en la cueva de Llonín en pigmentos rojos de pintura sin aparente colonización (clon AJ421159.1, Schabereiter-Gurtner et al., 2004). La OTU1 se ha detectado con una identidad del 100% en una mezcla de colonias de color blanco, amarillo, gris, azul grisáceo y rosa en una cueva de Eslovenia (clon FJ535092.1, Pasić et al., 2010) o previamente en Altamira con una identidad del 99,51% en colonias grises (clon HE603178.1, Cuezva et al., 2012). La OTU4 es la menos detectable en cuevas, aunque igualmente con una identidad del 99,76% se ha detectado previamente en la Cueva de Altamira (clon EF188338.1, Portillo et al., 2009) o con una identidad del 98% de nuevo en colonias blancas en una cueva volcánica en Nuevo Méjico, USA (clones KC331738.1, KC331736.1).

Dentro de los componentes secundarios que aparecen en este core es notable la presencia particular de la OTU3, prácticamente presente en todas las muestras que hemos analizado y asignada taxonómicamente a la clase *Gammaproteobacteria* y orden *Wb1-P19*. También se detectó como la segunda

OTU más abundante en muestras de suelos y soportes en la cueva (Figura 36). Este componente además ha sido detectado previamente en Altamira y en otros análisis de cuevas. Se ha descrito antes como uno de los 3 miembros claves en el análisis de 128 muestras de cueva (0,32% aire; 5,20% agua, 4,42% roca; 6,15% sedimento) en 8 cuevas kársticas en China (Zhu et al., 2019), y filogenéticamente se ha agrupado en un cluster relacionado con bacterias autótrofas oxidantes de nitrito y/o azufre. Además, esta OTU se ha detectado anteriormente con una identidad del 99,05% en colonias blancas grisáceas en la cueva de Tito Bustillo (clones AJ421204.1, AJ421216.1, Schabereiter-Gurtner et al., 2002) y con una identidad del 99,54% en colonias amarillas conformando uno de los cores en 3 cuevas distintas en Eslovenia, República checa y España, de nuevo en la cueva de Altamira (clones HE602901.1, HE602902.1, Porca et al., 2012). También se ha detectado con una identidad del 99,54% en pinturas con arte prehistórico en la cueva de Magura en Bulgaria (clon HE653868) relacionados con la familia *Chromatiaceae* con un 92-95% de identidad, las cuales son bacterias oxidadoras de sulfuro de hidrógeno (Ivanova et al., 2013). En cuevas de origen volcánico se han obtenido clones con una identidad del 99,54% en colonias marrones (clones LR130704-06) y en colonias amarillas (clones LT854971.1, JN850513.1, JN672508.1, JN850448.1, JN850498.1, HQ721141.1).

A pesar de la gran diversidad obtenida en los análisis, consideramos que existe un conjunto esencial y discreto de taxones que son los verdaderos componentes de las colonias visibles de Altamira. Los filos encontrados en este core han sido *Acidobacteria*, *Actinobacteria*, *Chloroflexi*, *Candidatus Dadabacteria*, *Candidatus Rokubacteria*, *Proteobacteria* y *Nitrospirae*. La gran mayoría de estos filos ya han sido detectados en las colonias en Altamira, a excepción de *Candidatus Dadabacteria* y *Candidatus Rokubacteria*. Todos estos filos suelen estar asociados a entornos de cueva como por ejemplo en muestras de biofilms o espelotemas (Lavoie et al., 2017; Oliveira et al., 2017; Ortiz et al., 2014; Xian et al., 2020). Los filos detectados *Rokubacteria* y *Dadabacteria* se han descrito como taxones involucrados en ciclos biogeoquímicos del carbono, nitrógeno y azufre (Hug et al., 2016). Bacterias del filo *Rokubacteria* son capaces de reducir completamente sulfato y oxidar

thiosulfato a sulfato, oxidar nitrato a nitrito y fijar carbono, mientras que bacterias del filo *Dadabacteria* son capaces de reducir nitrato a amonio, así como sulfato o tiosulfato. También tiene un papel importante en la nitrificación las bacterias correspondientes al filo *Nitrospirae* (Lavoie et al., 2017).

La composición final de este core, a pesar de las 4.288 OTUs detectadas en un inicio, nos permite proponer una composición simple para cada subtipo de colonia que presentaremos un poco más adelante. No sólo los componentes mayoritarios sino muchos de los componentes secundarios también se están detectando en entornos similares, reforzando la idea sugerida de la existencia de determinadas bacterias adaptadas a su desarrollo en cuevas. Estas bacterias constituirían lo que podría considerarse como “microbiota de cuevas” y una información más detallada de las mismas proveniente necesariamente de estudios metagenómicos será esencial para una mejor descripción y una mayor comprensión de sus condiciones de crecimiento.

La secuenciación del DNA total de muestras de las colonias descritas por la técnica de shotgun es una aportación nueva a la caracterización de la microbiota de Altamira y una de las primeras que se lleva a cabo sobre microbiota de cuevas. Como se adelantó más arriba con los datos obtenidos hemos sido capaces de generar seis metagenomas con un tamaño total de 529 megabases en secuencias mayores de 1 kb. Los metagenomas individuales obtenidos oscilaron entre los 25,3 Mb (colonias beis) y los 188,2 Mb (colonias grises subtipo G1).

La clasificación de las secuencias de cada metagenoma en “bines” nos ha permitido recomponer 90 borradores de genomas (MAGs) correspondientes hipotéticamente a otros tantos componentes de las colonias. El recuento de secuencias simples (reads) que mapean en cada uno de los MAGs reconstruidos demuestra la presencia de MAGs mayoritarios con los que alinean hasta el 80% de las secuencias totales (Figura 38). La asignación taxonómica de estos MAGs nos muestra una buena coincidencia con las OTUs que presentaron abundancias altas por la técnica de amplicones. Lo que la

secuenciación total no nos muestra fue una diversidad tan alta como la que ofrecía el análisis de amplicones.

Frente a las más de 4000 OTUs detectadas, solamente se obtuvieron 90 MAGs por la técnica de shotgun, lo que es consecuencia de la diferente tecnología. Unas pocas secuencias de amplicones bastan para definir una OTU mientras que un MAG representa la reconstrucción de un genoma completo. Es evidente que solo los genomas más abundantes van a recuperarse como MAGs y que las bacterias contaminantes del soporte no podrán ser recuperadas como MAGs. De hecho, cerca de dos tercios del total de secuencia ensamblada se encuentra en contigs de menos de 1 kb y por tanto está excluida de los MAGs. La mala calidad de ensamblado de toda esta información se debe a que proviene de genomas poco abundantes presumiblemente contaminantes de suelo o soporte.

Desde este punto de vista el análisis metagenómico por shotgun nos libra probablemente de contaminantes de suelo muy presentes entre las OTUs y aunque podríamos perder algunos componentes poco abundantes, nos ofrece una buena representación de los más abundantes.

No obstante, con 90 MAGs recuperados, obtenemos unos resultados sobre la complejidad de las colonias similares a los que obtuvimos usando el core de las 36 OTUs más abundantes. El número de componentes estimados por ambas técnicas fue similar, oscilando entre 6 y 34 componentes (Tabla 2). Esta diversidad además fue nuevamente mínima para las colonias beis y más alta para las colonias grises recuperándose un mayor número de MAGs en los metagenomas de estas muestras.

A continuación, vamos a proponer una composición en componentes para cada uno de los tipos de colonia combinando los resultados de secuenciación de amplicones y de DNA total.

### Colonias Beis

Las colonias beis fueron las que generaron los metagenomas más simples resultando en la recuperación de 5 MAGs (Figura 68). Lo más destacable fue la coexistencia de dos componentes mayoritarios en estas colonias que correspondieron con las actinobacterias clasificados con la clase *Nitriliruptoria* (MAG097/OTU1) y la familia *Pseudonocardiaceae* (MAG098/OTU2). El resto de los componentes minoritarios pertenecieron, dos a la clase *Gammaproteobacteria* (uno al orden *PLTA13* y otro al género *Lysobacter*), y el último al filo *Actinobacteria* (orden *Solirubrobacterales* y género *67-14*). Todos estos componentes se confirmaron por ambas técnicas como los componentes más abundantes. Además de los 5 MAGs se detectó en el metagenoma la secuencia completa de un 16S rRNA correspondiente con una *Gammaproteobacteria* (orden *wb1-P19*). Esta secuencia de 16S rRNA es igual al 16S rRNA del componente secundario más abundante en la mayoría del resto de metagenomas, de los que sí que se han recuperado MAGs.

Por ello pensamos que este 16S rRNA pertenezca a un sexto componente de las colonias beis, ya que coincide con la OTU3 que es una de las que se obtiene con mayor abundancia en esta muestra por la técnica de amplicones. El que el proceso de bisección no haya producido un MAG que corresponda a este sexto componente puede deberse a su menor abundancia en la colonia lo que va a dificultar su ensamblado frente a genomas más abundantes. También es posible que los contigs de este genoma sean similares en composición y abundancia a otro y de lugar a un MAG muy contaminado. Aunque la OTU3 sea muy abundante puede ocurrir también que corresponda a más de una especie como ocurre con los CMs clasificados con la OTU2, y por eso este componente por la técnica de amplicones muestra una abundancia más alta de la que en realidad presenta y que dificulta su recuperación.

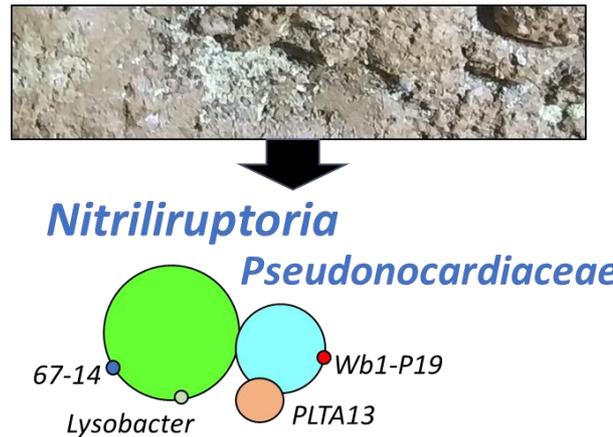


Figura 68. Propuesta de componentes finales para las colonias beis. Cada componente está representado por un círculo de diferente color y tamaño aproximadamente proporcional a su abundancia. Los CMs llevan el nombre de un tipo mayor y de color azul.

### Colonias amarillas

Las colonias amarillas presentaron una diversidad también baja con 9 componentes (Figura 69). Como ocurre en todas las colonias lo más característico fue el componente mayoritario clasificado dentro de la familia *Pseudonocardiaceae* (MAG088/OTU2). El segundo componente más abundante fue una *Gammaproteobacteria*, que en base a la comparación de su secuencia del 16S rRNA se clasificó en el orden *wb1-P19* y en base a sus ORFs, pertenecería al orden *Chromatiales*. Este componente es muy interesante, ya que se observa en casi todos los metagenomas. No hay en las bases de datos muchos genomas disponibles que se encuentren clasificados en taxones cercanos a él, sólo uno correspondiente con la bacteria *Chromatiales bacterium USCg\_Taylor* (GenBank assembly accession: GCA\_002007425.1) secuenciada de muestras de suelo antártico (Figura 47) y que aparece como no clasificado dentro del orden *Chromatiales*.

Como componentes menos abundantes en las colonias amarillas se recuperó una *Actinobacteria* (clase *Nitriliruptoria*) que por comparación de las 3 aproximaciones mencionadas a lo largo de esta tesis (secuencia de DNA y de proteína y secuencia del 16S rRNA) entre los componentes clasificados en esta misma clase, se trató del componente mayoritario de las colonias beiges (Figura 61). Este dato fue de los primeros indicios de que los componentes

mayoritarios también podían ser secundarios en otras colonias. El resto de componentes en las colonias amarillas clasificaron como *Gammaproteobacteria* (orden *PLTA13*), *Actinobacteria* (familia *Mycobacteriaceae*), *Acidobacteria subgroup 6*, *Alphaproteobacteria* (orden *Rhizobiales*), *Actinobacteria* y *Bacteroidetes*. Todos estos taxones fueron detectados igualmente por la técnica de amplicones, a excepción de los componentes clasificados con *Mycobacteriaceae* y *Actinobacteria*.

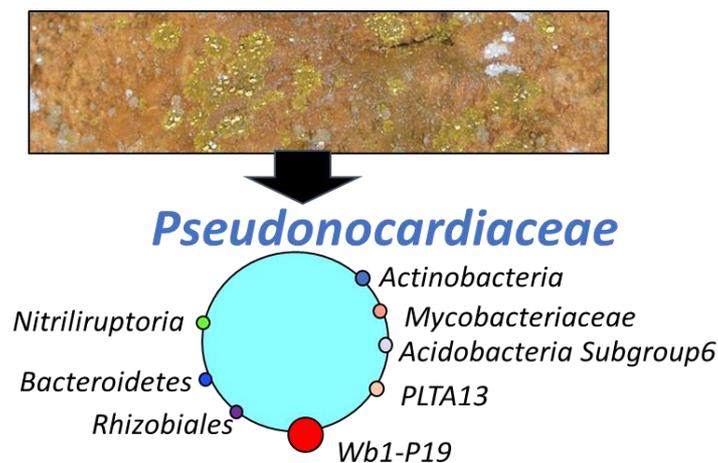


Figura 69. Propuesta de componentes finales de las colonias amarillas. Cada componente está representado por un círculo de diferente color y tamaño aproximadamente proporcional a su abundancia. El CM lleva el nombre de un tipo mayor y de color azul.

#### Colonias blancas (subtipos W1 y W2)

Las colonias blancas también presentan una composición poco compleja. En el subtipo W1 hemos identificado 9 componentes y en el subtipo W2 solamente 6 (Figura 70). Los dos subtipos se diferenciaron en base al componente mayoritario, en ambos pertenecientes al filo *Actinobacteria*, siendo el del subtipo W1 de la familia *Pseudonocardiaceae* (MAG073/OTU2), y el del subtipo W2 de la clase *Nitriliruptoria* (MAG082/OTU1). Aparte, ambos subtipos compartieron como secundario más abundante la *Gammaproteobacteria* clasificada con el orden *Wb1-P19* según SILVA y orden *Chromatiales* en el NCBI, que ya se ha identificado en las colonias anteriores.

De entre los componentes secundarios de las colonias blancas del subtipo W1 cabe destacar la presencia de una arquea de la familia *Nitrosopumilaceae*. Otros componentes pertenecieron a una *Rokubacteria* (clase NC10, orden *Methylomirabilales*, familia *Methylomirabilaceae* y género *wb1-A12*), *Alphaproteobacteria* (orden *Rhodospirillales* y familia *Rhodospirillaceae*), *Nitrospirae*, *Candidatus Dadabacteria*, *Deltaproteobacteria* y *Acidobacteria* (familia *Blastocatellia* y orden 11-24). La arquea detectada en estas colonias también se detectó en colonias grises subtipo G1 correspondiendo con las únicas arqueas en estos metagenomas. La presencia de todos estos componentes se confirmó por ambas técnicas de secuenciación.

Los otros componentes secundarios de las colonias blancas del subtipo W2 correspondieron a una *Actinobacteria* (clase *Thermoleophilia* y orden *Solirubrobacterales*), *Parcubacteria*, *Betaproteobacteria* y *Acidobacteria*. Todos han sido detectados por la técnica de amplicones, a excepción de los componentes clasificados con el superfilo *Parcubacteria* y con el orden *Solirubrobacterales*. En el caso del componente clasificado con el superfilo *Parcubacteria*, su secuencia del 16S rRNA se encuentra interrumpido por 3 intrones (Figura 50) y esto hace que no se detecte por la técnica de amplicones. Los hits más cercanos para el 16S asignaron a esta bacteria a los filios *Candidatus Nomurabacteria* y *Candidatus Campbellbacteria*, aunque con una identidad inferior al 80%. En la literatura, hay muy poca información publicada sobre estos taxones y no se han detectado antes en cuevas probablemente porque muchos de estos estudios se basan en la amplificación de regiones variables del gen 16S rRNA. Sí se ha detectado en ambientes marinos en muestras de sedimento (León-Zayas et al., 2017) y en muestras de suelo forestal (Pu et al., 2019), cuevas y en depósitos provenientes de vermiculaciones (Addesso et al., 2020).

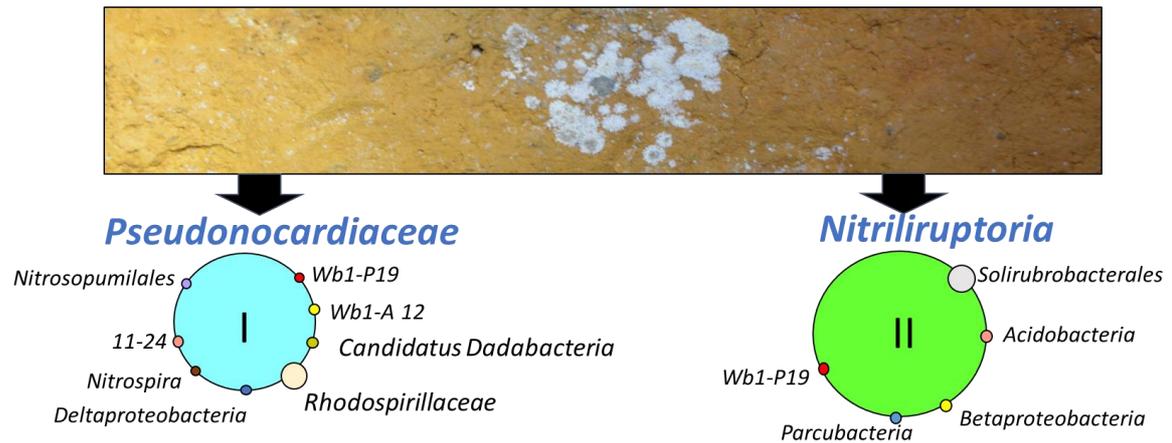


Figura 70. Propuesta de componentes finales para las colonias blancas subtipos W1 y W2. Cada componente está representado por un círculo de diferente color y tamaño aproximadamente proporcional a su abundancia. Los CMs llevan el nombre de un tipo mayor y de color azul.

### Colonias grises (subtipos G1 y G2)

Las colonias grises fueron las más diversas de todas las analizadas estimando en ellas un mayor número de componentes como se muestra en la figura 71 (34 en el subtipo G1 y 27 en el subtipo G2). En ambos subtipos hubo 2 MAGs que fueron descartados por presentar alta contaminación y otros 2 por tener poca completitud, no facilitando una taxonomía significativa.

En ambas colonias, el componente mayoritario fue clasificado dentro de la familia *Pseudonocardiaceae*, tratándose de 2 especies diferentes (MAG001/OTU4 en G1 y MAG046/OTU2 en G2). Como componentes secundarios encontramos 7 MAGs pertenecientes al filo *Actinobacteria* con los órdenes *Pseudonocardiales* (1 en G1 y 2 en G2), *Micromonosporales* (1 en G1) y la clase *Nitriliruptoria* (1 en G2). Se detectaron los componentes mayoritarios de las colonias beis (MAG097/98) como secundarios en las colonias G2 (MAG050/60). Algo también notable en estas colonias fue la abundancia de componentes del filo *Proteobacteria* de las clases *alpha*, *beta* y *delta*, de las que se recuperaron 17 MAGs en G1 y 14 en G2. Los filios restantes pertenecieron a *Acidobacteria* (4 en G1 y 1 en G2), *Nitrospirae* (1 en G1 y 2 en G2), *Chloroflexi* (2 en cada subtipo), *Candidatus Rokubacteria* y *Candidatus*

Dadabacteria (1 en cada subtipo). Sólo en el subtipo G1 se detectó un componente clasificando con el superfilo *Parcubacteria* (detectada en las colonias blancas subtipo W2), el filo Candidate division NC10 (orden *Methylomirabiales*, familia *Methylomirabilaceae* y género *wb1-A12*) y la arquea *Thaumarchaeota* (familia *Nitrosopumilaceae*) también detectada en las colonias blancas subtipo W1.

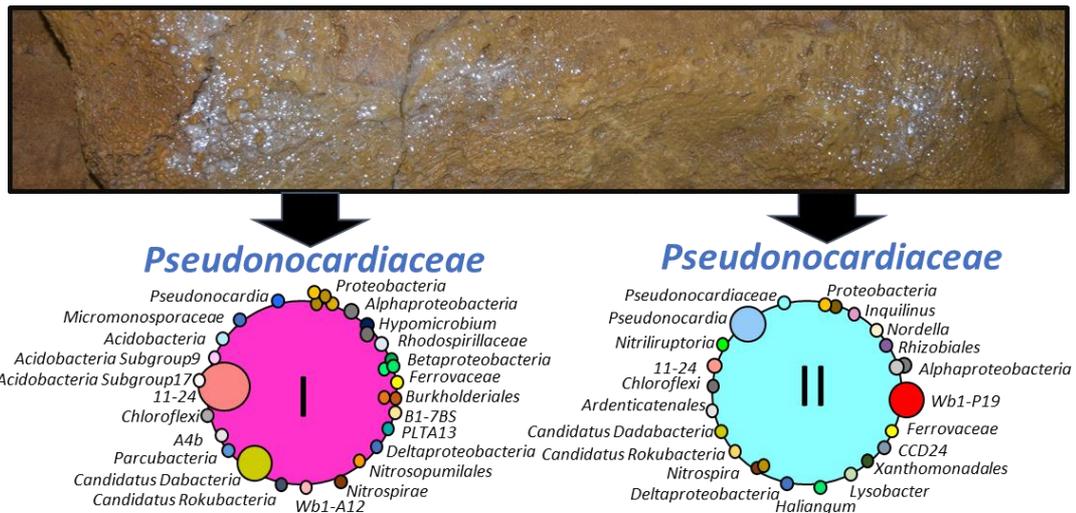


Figura 71. Propuesta de componentes finales para las colonias grises subtipos G1 y G2. Cada componente está representado por un círculo de diferente color y tamaño aproximadamente proporcional a su abundancia. Los CMs llevan el nombre de un tipo mayor y de color azul.

### 6.3 Novedad taxonómica de los componentes de las colonias

De todos los componentes descritos, los componentes mayoritarios, debido a su abundancia, son los que nos han proporcionado MAGs de mayor calidad (más completos y menos contaminados), lo que nos permite profundizar con bastante detalle, tanto en la posición taxonómica como en las capacidades metabólicas y fisiológicas de dichos componentes. Disponemos de la secuencia de la mayoría de los genes que codifican proteínas, de los clusters completos de los rRNA ribosomales (16S, 23S y 5S) y sus correspondientes regiones intergénicas. La comparación de los genomas (MAGs completos) de estos componentes nos muestra que forman dos clusters diferentes

clasificando dentro del filo *Actinobacteria*. Los análisis filogenéticos basados en comparaciones del 16S rRNA o de genomas completos muestran la ausencia de genomas muy relacionados, o lo que es lo mismo, sugiere que se representan taxones nuevos no definidos dentro de las actinobacterias. Los componentes mayoritarios de las colonias G1, G2, W1, YC y BC formaron un cluster (Cluster III) separado de los géneros descritos dentro la familia *Pseudonocardiaceae*. La comparación de los 16S rRNAs de estos componentes reveló identidades entre ellos del orden del 97%, la secuenciación de amplicones solamente resuelve dos OTUs en este grupo. Por este criterio se deberían considerar como pertenecientes a una misma especie. Sin embargo, la comparación de los genomas completos y sobre todo un análisis de las familias de ortólogos presentes en cada MAG, nos revelan una mayor heterogeneidad con un “core” que contendría un 30% de los genes de cada MAG, siendo el resto de genes específico de cada uno de ellos (datos no mostrados). Por lo tanto, nos inclinamos por la propuesta de considerar a estos cinco genomas pertenecientes a un nuevo género en la familia *Pseudonocardiaceae* en el que representarían posiblemente 5 especies diferentes.

En el caso del cluster IV, donde agruparon los componentes de las colonias W2 y BC también formaron un cluster separado de los órdenes descritos de la clase *Nitriliruptoria*, por lo que igualmente se justificaría la propuesta de un nuevo orden en el que situaríamos estos dos MAGs.

Las secuencias de 16S rRNAs de estos componentes no tienen homólogos cercanos en las bases de datos de secuencias completas o taxonómicamente bien clasificadas, sin embargo, podemos encontrar secuencias muy próximas a ellas en bases de datos de proyectos de secuenciación de amplicones o clones de origen metagenómico, que no proporcionan una clasificación taxonómica precisa. Podemos encontrar (octubre de 2020) 262 secuencias relacionadas con los componentes mayoritarios del cluster III y con 71 en el caso del cluster IV. Estas secuencias no clasificadas no nos ayudan a afinar la taxonomía de los MAGs de los componentes mayoritarios, pero si soportarían nuestra propuesta de un nuevo género en la familia *Pseudonocardiaceae* y de un orden

nuevo en la clase *Nitriliruptoria* específico de bacterias de cueva, y para los que estaríamos proporcionando los primeros borradores de genomas de alta calidad.

## 6.4 La formación de colonias visibles como solución universal para los microorganismos de cuevas.

Como resumen y conclusión de la composición de los distintos tipos de colonias vislumbramos un patrón común que se repite y que permite hacer una propuesta sobre la composición general de las colonias visibles coloreadas de cuevas (Figura 72), que representarían un tipo de consorcio bacteriano adaptado a la vida en condiciones de escasa disponibilidad de nutrientes, que son frecuentes en cuevas, y que se localizan con relativa frecuencia en cuevas de diferente naturaleza geoquímica y distribución geográfica universal.

En todas las colonias analizadas se encuentra uno o dos componentes muy abundantes (CMs) del filo *Actinobacteria* (*Pseudonocardiaceae* o *Nitriliruptoria*) y en menor abundancia uno o dos de la clase *Gammaproteobacteria* relacionadas con los órdenes *Chromatiales*, *PLTA13* o *Xanthomonadales* (*Lysobacter*). Como componentes más minoritarios a menudo encontramos otras actinobacterias que puede ser alguno de los CMs (*Pseudonocardiaceae* o *Nitriliruptoria*) u otra diferente (*Pseudonocardia*, *Mycobacteriaceae* o *Solirubrobacterales*), una *Acidobacteria* (11-24), una *Alphaproteobacteria* (*Rhodospirillales* o *Rhizobiales*), una *Betaproteobacteria* (*Burkholderiales*, *Ferrovaceae* o *B1-7BS*) y una *Deltaproteobacteria* (*Myxococcales*). Como componentes menos generalizados se encontraría algún componente relacionado con los filos *Candidatus* Dadabacteria, *Candidatus* NC10, *Candidatus* Rokubacteria, *Nitrospirae*, *Chloroflexi* y *Thaumarchaeota*, y el superfilo *Parcubacteria*.

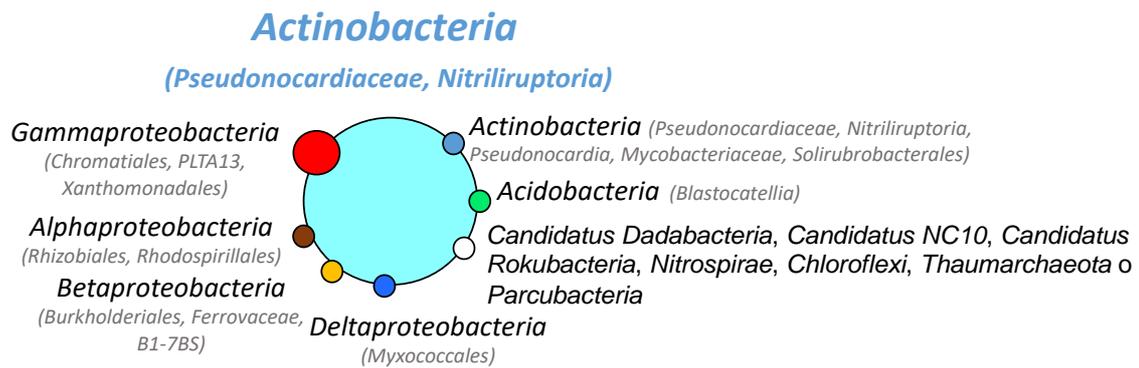


Figura 72. Modelo propuesto de la composición general de una colonia visible del tipo observado habitualmente en cuevas.

De acuerdo con este modelo podemos interpretar la diversidad en las colonias y la aparición de nuevos tipos como resultado de una evolución continua basada en la combinación de determinados componentes de las colonias actuales para el desarrollo de nuevas colonias en la cueva. Esto se ha podido evidenciar en las colonias beis, que fueron las últimas colonias en ser detectadas en la cueva y resultaron ser las más simples de todas. Más de la mitad de sus componentes (4/6) entre los que se incluyen los componentes mayoritarios, son componentes detectados como secundarios en otras colonias como amarillas o grises. Más recientemente se han observado en Altamira colonias con una clara tonalidad rosa que aún no han sido analizadas. Si nuestra hipótesis es correcta, encontraremos que estas colonias están formadas esencialmente por los mismos componentes ya descritos combinados en una nueva proporción dentro de las reglas propuestas en el modelo general para las colonias visibles. Esta posibilidad de selección de componentes para producir combinaciones más adaptadas a la vida en la cueva se traduce en una ventaja y a la vez en un arma que las bacterias tienen para conseguir proliferar a las condiciones de Altamira.

Colonias visibles parecidas a las que se están desarrollando en la cueva de Altamira se observan en otras cuevas en España (Figura15) o en el mundo (Figura 16) con una composición muy similar. Algunas de estas cuevas contienen arte paleolítico, y han sido en mayor o menor medida visitadas por el

público, pero, sin embargo, otras se tratan de cuevas de origen volcánico con condiciones bastante diferentes. Altamira siempre ha estado en debate de permanecer cerrada a visitantes por miedo a que la presencia de estas colonias fuese causada por estas visitas, y que una continuidad de estas desencarrilara una mayor colonización. Los resultados obtenidos en esta tesis sugieren que los componentes detectados en las colonias de Altamira se detectan igualmente en las colonias de otras cuevas en otros lugares del mundo, sin relación alguna con Altamira. Además, también confirma la idea de un core de componentes que se encuentran presentes dentro de este tipo de muestras biológicas y sin una relación aparente con bacterias asociadas habitualmente con el hombre. No hay duda de que el trato que se le ha dado a la cueva anteriormente, así como acciones desarrolladas en sus alrededores pueden haber acelerado el desarrollo de estas colonias naturales en ella. Lo que nuestros resultados sugieren es que esta proliferación ha sido debida a cambios nutricionales o ambientales y que aparentemente no ha habido aporte externo significativo de microorganismos. Por ello, el foco del control o prevención, debe colocarse sobre todo en el mantenimiento de las condiciones ambientales y en evitar en lo posible el aporte de materia orgánica del exterior y en un continuado seguimiento de parámetros ambientales y microbiológicos que permitan controlar la constancia de las condiciones y que no se alteren los delicados equilibrios que gobiernan la vida de la cueva.

## **6.5 El metabolismo de las colonias de la cueva de Altamira**

Desde un punto de vista ecológico las cuevas proporcionan ambientes estables y protegidos donde el desarrollo de la vida microbiana se halla resguardada frente a condiciones adversas cambiantes, por el contrario, son ambientes que pueden ser muy aislados y esto se puede traducir en una disponibilidad limitante de nutrientes.

La cueva de Altamira es un ejemplo claro en cuanto a la constancia de las condiciones ambientales y lo que cabe suponer ante esta constancia, es que el principal factor que va a determinar el tipo de seres vivos que proliferen en la

cueva va a ser la disponibilidad de nutrientes. De este modo la microbiota que encontremos en la cueva va a depender directamente de la disponibilidad de nutrientes disponibles en cada momento.

En los años anteriores al año 2000, la cueva de Altamira era considerada un ambiente poco oligotrófico (Luque and Labrada, 2016) en el que el interior de la cueva presentaba grandes cantidades de materia orgánica procedente de una actividad grande de microvertebrados, insectos, etc. Además, la cueva venía de pasar por un régimen de visitas muy fuerte, había estado iluminada artificialmente e incluso se mantenía actividad ganadera en el entorno de la cueva. En estas circunstancias no es de extrañar que en el periodo de 2003 a 2005 la cueva sufriera una colonización visible en la sala de las pinturas por cianobacterias del orden *Chroococcales* y *Oscillatoriales*; y de algas unicelulares eucariotas en concreto de los géneros *Cyanidium*, *Chladmydomonas* y *Chlorococcum* en menor frecuencia (Informe CSIC 2003-2005 pag 88-91). La causa de esta colonización fue la presencia de focos de iluminación fija en la sala, que al retirarse provocaron la desaparición de estos biofilms fototróficos.

La presencia de luz permitía la formación de este tipo de biopelículas en las que podemos encontrar estos taxas (grupos fototróficos como cianobacterias o algas), que son más favorecidos y considerados como colonizadores primarios debido a su metabolismo más sencillo, los cuáles obtienen de la luz la energía necesaria para fijar CO<sub>2</sub> a partir del que producir moléculas más complejas como carbohidratos que son alimento para el desarrollo heterótrofo (Jurado et al., 2020; Popović et al., 2020). Por tanto, la actividad de estos productores primarios permite el desarrollo de microorganismos heterótrofos. Si no hay luz, los grupos fototróficos no se desarrollan y encontramos entonces como microorganismos quizás más elementales a los microorganismos quimioautótrofos considerados como los principales productores de la materia orgánica en estos biofilms de cueva. Otros microorganismos como hongos, amebas, metazoos, bacterias o arqueas dependerán energéticamente de los autótrofos o de la materia orgánica que entre a la cueva por otros medios, y en

consecuencia juegan un papel menos importante en estos biofilms autotróficos. (Popović et al., 2017; Urzì et al., 2010).

Sin embargo, las condiciones energéticas de Altamira han variado mucho hasta el día de hoy. La presencia de murciélagos hace tiempo que no se detecta, en cuanto a los roedores se tiene evidencia de su entrada, ya que en alguna ocasión se detectan algunas huellas, pero sin otra evidencia visible de los mismos. Tampoco es ya frecuente la presencia de cadáveres de animales, aunque sí alguna vez de artrópodos del orden *Trichoptera* o de la clase *Arachnida*. En cuanto a la materia orgánica que se puede detectar en suelos exteriores y que puede conllevar al arraste de estos nutrientes al interior de la cueva por acción del agua circulante, análisis de compuestos orgánicos en estos suelos en el periodo de 2007 a 2009 fueron altos y de nitratos del 2003 al 2009, aunque se puede observar que con los años su concentración ha ido disminuyendo sugiriendo un proceso de dilución de estos compuestos en los suelos por la acción de la lluvia (Figura 9). La actividad ganadera en los alrededores de la cueva hace tiempo que ha desaparecido, y es probable que el resultado de estas acciones hace que el aporte de materia orgánica en el agua de infiltración haya disminuido. La retirada de la iluminación artificial estable de la cueva ha determinado también la práctica desaparición de biofilms verdes fototróficos.

Actualmente, la cueva sin aporte de luz natural y cerrada excepto a visitas controladas y a trabajos de investigación/mantenimiento, se vigila con mayor detenimiento llevándose a cabo entradas de control de estado por personal del museo con periodicidad semanal, en las que se presta especial atención a eliminar al momento cualquier posible fuente de alimento que soporte el crecimiento de microorganismos. Esta vigilancia se complementa con un especial cuidado en no introducir nada en la cueva que no esté escrupulosamente limpio de materia orgánica y exento de contaminantes biológicos, así como de prevenir acciones en los alrededores que favorezcan la entrada de nutrientes en la cueva.

Como se ha mencionado, la falta de iluminación ha conducido a la desaparición casi total de los organismos fototróficos y de biofilms formados por los mismos. No fue así en el caso de las manchas verdosas tenues detectadas en 2016 en el pasillo entre la Sala de los Muros y la Hoya, coincidiendo con un punto de iluminación temporal durante las visitas experimentales. En estas manchas fue posible detectar un alga mixotrófica clasificada con la especie *Jenufa aroterrestrica*. Esta alga tiene la capacidad de alternar fases autotróficas mientras haya luz disponible con fases oscuras en las que el alga se comporta como un organismo heterótrofo. La eliminación de la iluminación en este punto ha producido la disminución del color verde, aunque con certeza el alga persiste en el techo de la zona en una forma de vida heterótrofa, o latente.

En otras cuevas como la cueva de Nerja (Málaga), la ausencia de organismos fototróficos se ha atribuido a la presencia de una mayor abundancia de amebas o bacterias depredadoras (Jurado et al., 2020) que conducen a la desaparición de las algas. En principio la ausencia de luz en Altamira parece ser la causa más probable de la desaparición (o poca prevalencia) de organismos fotosintéticos.

Todos estos datos nos permiten plantear que en la actualidad el ambiente de la cueva de Altamira es nutricionalmente oligotrófico, y esta circunstancia como decíamos al principio va a determinar la microbiota de la cueva. Una estrategia microbiana de supervivencia en ambientes oligotróficos es la formación de estos biofilms en los que participan varias especies diferentes. Ninguna de las especies que forman parte del biofilm es capaz por ella sola de proliferar en las condiciones disponibles, sino que es la cooperación metabólica entre los miembros del biofilm la que permite la suficiente capacidad metabólica para poder desarrollarse en este ambiente. Una consecuencia de esta estrategia es la dificultad para conseguir la recuperación de componentes de los biofilms en cultivos puros. Por eso no es extraño nuestra repetida falta de éxito intentando cultivar componentes de las colonias coloreadas de Altamira.

En Altamira se detectan este tipo de biofilms en forma de lo que hemos llamado “colonias visibles” microbianas de distintos colores (blancas, grises, amarillas y

beis). Con características externas similares a las de Altamira, este tipo de biofilms se describen con mucha frecuencia en cuevas de todo tipo y localización, convirtiéndose en el mecanismo principal de supervivencia de bacterias que habitan en estos entornos subterráneos. En Altamira las colonias se desarrollan en las paredes y techos de la cueva, su descripción comienza cerca del año 1999, son mucho más frecuentes cerca de la entrada y como se dijo anteriormente se consideran el principal riesgo para el biodeterioro de las pinturas.

Los resultados presentados muestran que las colonias visibles están compuestas de comunidades microbianas con gran dominancia de bacterias y menor de arqueas. Se descarta la presencia de hongos en las colonias. Experimentos iniciales que trataban de amplificar ITS fúngicas a partir de DNA total de las colonias nunca dieron resultado. Por otra parte, no se obtienen secuencias de origen fúngico en los metagenomas. Otro tanto podemos decir de la presencia de protozoos, de los que únicamente tuvimos evidencia en el análisis de una muestra de "moonmilk", un tipo de espeleotema de posible origen biológico, completamente separado de las colonias visibles, que también se encuentra en algunas zonas de la cueva de Altamira.

Una de las cuestiones relevantes sobre la microbiología de Altamira no despejada en esta tesis es el soporte energético y nutricional para el desarrollo de las colonias visibles. La hipótesis dominante hasta la fecha propone que estas colonias son consorcios heterótrofos que prosperan gracias a la materia orgánica disuelta (MOD) detectada en aguas de infiltración, y que disponen de la cantidad suficiente de nitrógeno y otros elementos a partir del agua de infiltración o de los materiales de suelos y paredes. Por lo general, la gran mayoría de cuevas se consideran ambientes oligotróficos con bajas cantidades de recursos tróficos en forma de sustancias orgánicas oxidables. Estos ambientes suelen presentar una concentración inferior a 2 mg/l de carbono orgánico en las aguas de goteo. En estas condiciones y teniendo además en cuenta la alta disponibilidad de CO<sub>2</sub> en la cueva, no es descabellado pensar en cierta capacidad autotrófica de fijación de CO<sub>2</sub>, que debería estar energizada por alguna sustancia química susceptible de oxidación dada la ausencia de luz.

Entre los sustratos que podrían ser oxidados para generar el potencial reductor necesario para la asimilación del carbono a partir de dióxido de carbono ( $\text{CO}_2$ ), y a falta de evidencias que soporten el uso del azufre o el hierro como energizantes, cabe pensar en la utilización de gases atmosféricos como monóxido de carbono (CO) o hidrógeno ( $\text{H}_2$ ). Un tercer compuesto candidato en esta categoría es el amonio ( $\text{NH}_4^+$ ) que podría ser oxidado parcialmente hasta nitrito ( $\text{NO}_2^-$ ) o completamente hasta nitrato ( $\text{NO}_3^-$ ) como han demostrado ser capaces de hacer las bacterias del grupo “Comammox” (Complete ammonium oxidizers).

Algunos de los componentes identificados en las colonias son conocidos por tener capacidad quimioautotrófica. En especial, las actinobacterias, que son el *phylum* mayoritario en las colonias presentan capacidad demostrada para fijar  $\text{CO}_2$  utilizando donadores inorgánicos de energía. Algunas especies del género *Pseudonocardia* se han demostrado capaces de fijar  $\text{CO}_2$  usando el enzima Ribulosa bis-fosfato carboxilasa (RuBisCo). *P. dioxanivorans*, además de fijar  $\text{CO}_2$  por ciclo de Calvin-Benson-Bassham, posee la capacidad de oxidar hidrógeno atmosférico por medio de una hidrogenasa, de manera que es capaz de crecer de manera autotrófica en un medio con sólo  $\text{H}_2$ -bicarbonato (Grostern and Alvarez-Cohen, 2013). Otras especies de *Pseudonocardia* son igualmente fijadoras de  $\text{CO}_2$  pero usan alternativamente monóxido de carbono o amonio como fuente de potencial reductor (Park et al., 2008). La presencia de hidrogenasas es habitual en especies de micobacterias, lo que dota este grupo de una capacidad energética importante que va a ser explotada para la capacidad de las bacterias de este género para causar infecciones persistentes y estados de latencia (Greening et al., 2014).

Las arqueas detectadas en algunos metagenomas corresponden con el filo *Thaumarchaeota* que es bien conocido por ser arqueas oxidadoras de amonio (AOA), descritas principalmente como aerobios quimiolitotrófos oxidantes de amonio, aunque también pueden ser mixótrofos o heterótrofos (Bartossek et al., 2012). También hemos encontrado miembros del filo *Nitrospirae*, detectados en las colonias como componentes secundarios minoritarios. Bacterias de este género juegan un papel importante en la nitrificación como bacterias aerobias

quimiolitotróficas oxidantes de nitritos (Daims and Wagner, 2018). Recientemente se ha reconocido su capacidad para llevar a cabo la oxidación completa de amonio hasta nitrato por lo que las coloca en un grupo (Comammox) de gran importancia en el ciclo del nitrógeno (Daims et al., 2015; van Kessel et al., 2015).

De los borradores de los genomas obtenidos de los componentes de las colonias hemos obtenido información sobre sus capacidades metabólicas, y estos datos podrían darnos alguna respuesta sobre las capacidades metabólicas de las colonias e indirectamente sobre las reacciones básicas que pueden estar soportando en el interior de la cueva.

El catálogo de genes que hemos obtenido puede ser interrogado para encontrar en él genes marcadores de las reacciones metabólicas en las que se basaría la capacidad quimioautotrófica (Tabla S7). En primer lugar, hemos investigado la presencia de las diferentes rutas posibles para la fijación de CO<sub>2</sub> conocidas en bacterias. De ellas, solamente hemos encontrado de forma consistente evidencias de la presencia del ciclo de Calvin-Benson-Bassham, basadas en la presencia de los genes *rbcLS*, que codifican las dos subunidades catalíticas del enzima RuBisCo (EC:4.1.1.39, K01601, K01602). Estos dos genes se han localizado en todos los MAGs de los componentes mayoritarios y también en algunos de los componentes secundarios. En los CMs, los genes *rbcLS* generalmente se encuentran formando parte de clusters de genes con otras funciones auxiliares en la regulación o ensamblado del enzima, lo que mejora la evidencia de su funcionalidad.

Un gen asociado a la funcionalidad de fijación de CO<sub>2</sub> es el que codifica la anhidrasa carbónica. Este enzima que suele formar parte de los carboxisomas, orgánulos especializados en fijación de carbono, asegura la disponibilidad de CO<sub>2</sub> a partir de bicarbonato, un compuesto que está disponible *ad libitum* en la cueva de Altamira. Este enzima es ubicuo en todos los metagenomas. Se encuentra en todos los componentes mayoritarios y es a su vez muy abundante en componentes secundarios (Tabla 7).

Sustrato	Gen	Enzima	EC number	KOs	YC	BC	W1	W2	G1	G2
CO	coxL	aerobic carbon-monoxide dehydrogenase large subunit	EC:1.2.5.3	K03520	13	3	19	27	99	104
	coxM	aerobic carbon-monoxide dehydrogenase medium subunit	EC:1.2.5.3	K03519	4	4	11	11	62	55
	coxS	aerobic carbon-monoxide dehydrogenase small subunit	EC:1.2.5.3	K03518	13	6	14	18	98	89
H <sub>2</sub> CO <sub>3</sub>	cynT, can	carbonic anhydrase	EC:4.2.1.1	K01673	18	9	10	7	53	48
Formate	fdoG	formate dehydrogenase major subunit	EC:1.17.1.9	K00123	25	8	18	18	89	74
	fdoH	formate dehydrogenase iron-sulfur subunit	EC:1.17.1.9	K00124	7	3	6	4	34	29
	fdol	formate dehydrogenase subunit gamma	EC:1.17.1.9	K00127	5	1	4	2	23	22
H <sub>2</sub>	hyaB, hybC	hydrogenase large subunit	EC:1.12.99.6	K06281	9	3	5	6	24	22
	hyaA, hybO	hydrogenase small subunit	EC:1.12.99.6	K06282	11	2	5	5	26	19
NH <sub>3</sub>	amoA/pamoA	methane/ammonia monooxygenase subunit A	EC:1.14.99.39	K10944	1	0	5	1	7	5
	ureC	urease alpha subunit	EC:3.5.1.5	K01428	3	1	11	0	25	25
	amt	ammonium transporter		K03320	26	13	27	15	120	92
CO <sub>2</sub> fixation	rbcL	ribulose-bisphosphate carboxylase large chain	EC:4.1.1.39	K01601	5	4	3	2	12	11
	rbcS	ribulose-bisphosphate carboxylase small chain	EC:4.1.1.39	K01602	3	3	2	2	12	8
N <sub>2</sub> fixation	nifD	nitrogenase molybdenum-iron protein alpha chain	EC:1.18.6.1	K02586	0	0	0	0	3	0
	nifK	nitrogenase molybdenum-iron protein beta chain	EC:1.18.6.1	K02591	0	0	0	0	3	0
	nifH	nitrogenase iron protein NifH	EC:1.18.6.1	K02588	0	0	0	0	1	1

Tabla 7. Lista de marcadores de genes autotróficos y número de genes detectados en cada metagenoma.

Por lo que respecta a posibles reacciones de oxidación que proporcionen el potencial reductor necesario para la fijación oscura de CO<sub>2</sub>, encontramos los genes que podrían catalizar las reacciones de este tipo habituales en actinobacterias. Todos los componentes mayoritarios y algunos de los componentes secundarios contienen genes para las dos subunidades de la hidrogenasa (K06281 y K06282), y también frecuentemente los necesarios para el ensamblado y maduración del enzima. Los genes *coxL* y *coxS* (EC:1.2.5.3, K03520, K03518), que codifican las dos subunidades de una deshidrogenasa de monóxido de carbono, se encuentran en todas las colonias, aunque no siempre en los componentes mayoritarios, de hecho, está ausente en MAG082 y MAG097 que son los miembros de la clase *Nitriliruptoria*.

En principio, y casi exclusivamente gracias a los componentes mayoritarios las colonias visibles podrían tener al menos etapas de su vida en condiciones quimioautotróficas. La disponibilidad de CO<sub>2</sub> en la cueva de Altamira no es un factor limitante sobre todo en la época de invierno. No es tan clara respecto a las concentraciones disponibles de H<sub>2</sub> o de CO, gases presentes en concentraciones mucho menores. Sin embargo, los gases atmosféricos han demostrado ser suficientes para energizar la fijación de carbono por parte de microorganismos en ambientes antárticos extremos (Ji et al., 2017) o en ecosistemas de suelo (Bay et al., 2021).

Una posibilidad adicional sería la utilización de la oxidación del amonio por parte de las colonias visibles como reacción energizante. En todas las colonias, con excepción de las colonias beis, detectamos genes de oxidación del amonio a hidroxilamina (*amoA/pamoA*, EC:1.14.99.39, K10944), que es el primer paso del proceso de nitrificación al que seguiría su oxidación a nitritos. La presencia de oxidasas de amonio se complementa con dos genes que pueden ser necesarios para asegurar la disponibilidad de este ion, si es necesario para el crecimiento de las colonias. Estos dos sistemas complementarios son los transportadores de amonio (*amt*, K03320) y la metaloenzima ureasa (*ureC*, EC:3.5.1.5, K01428), capaz de hidrolizar urea y proporcionar CO<sub>2</sub> y amonio, necesarios para el crecimiento de las colonias.

En este caso no son los componentes mayoritarios los que albergan estos genes, sino que los encontramos en componentes secundarios como las arqueas (Tolar et al., 2017) o los miembros del *phylum Nitrospira* (Palomo et al., 2018). Es probable que esta reacción sea importante en el metabolismo global de las colonias lo que nos daría una justificación nutricional para la necesidad de la formación de consorcios microbianos para una forma de vida autónoma. En todas las colonias se detectan genes relacionados con el metabolismo oxidativo (respiración aerobia) y transporte electrónico usando oxígeno como aceptor. Igualmente se han detectado la mayoría de los genes implicados en las reacciones catabólicas centrales necesarias para la oxidación de moléculas orgánicas como la glucosa, ácidos grasos y/o aminoácidos, para la obtención final de energía y compuestos más simples utilizados para procesos de biosíntesis o asimilatorios ( $\text{H}_2\text{O}$ ,  $\text{CO}_2$  o  $\text{NH}_4^+$ ).

Este análisis del potencial metabólico se presenta para abrir una hipótesis de trabajo que propondría la capacidad de las colonias visibles para crecer de forma autotrófica en las condiciones naturales en Altamira al menos durante la parte del año (otoño-invierno) en la que la concentración de  $\text{CO}_2$  atmosférico es más alta y la actividad vegetal en la cobertera más baja, y al tiempo, la disponibilidad de sustratos orgánicos oxidables en el agua de infiltración más limitante (Figura 73).

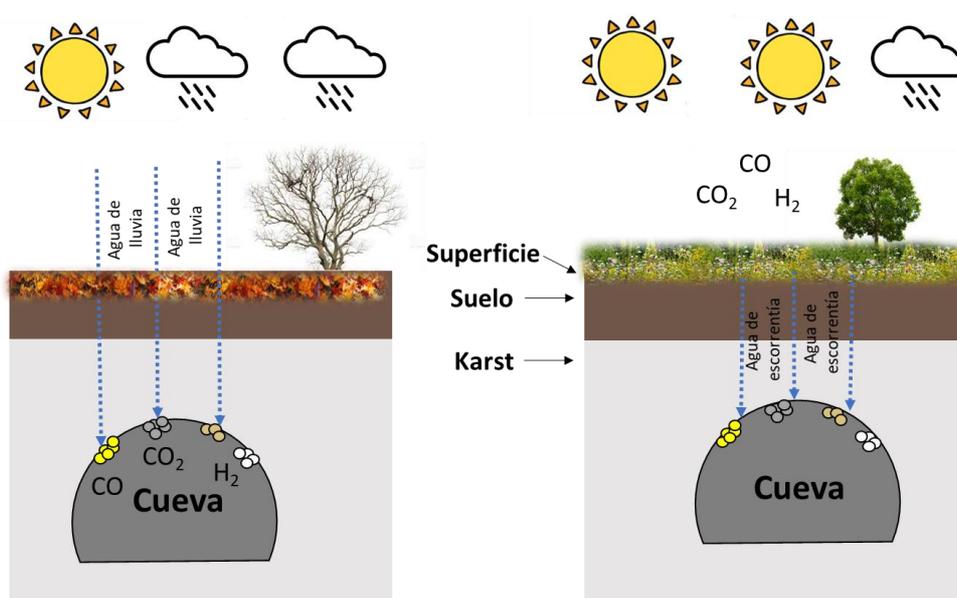


Figura 73. Hipótesis del potencial metabólico sugerido para las colonias. En épocas de invierno el agua de infiltración es pobre en materia orgánica y las colonias optan por una forma de vida autotrófica (izq) favorecida por las altas concentraciones de CO<sub>2</sub>. En verano la mayor concentración de materia orgánica disuelta (MOD) permite una forma de vida heterotrófica.

Aunque la evidencia presentada de la abundancia de genes esenciales para una etapa quimioautotrófica es grande, necesitará una demostración al menos de la expresión de esos genes por análisis transcripcional o una demostración fisiológica de las reacciones que se reclaman. Igualmente ha quedado fuera del alcance de esta tesis un análisis en profundidad de las capacidades metabólicas de las colonias, que podría ser abordado a partir de los datos metagenómicos disponibles, que en el caso de los MAGs de mayor calidad podría proporcionar casi una reconstrucción metabólica completa de sus dueños.

## 6.6 Recuperación en cultivo puro de las bacterias de Altamira

Una buena manera para el estudio de los componentes de las colonias sería poder obtenerlos en cultivos puros, pero ninguno de los intentos hasta ahora en las colonias de Altamira ha permitido el cultivo de los componentes mayoritarios y secundarios detectados en este estudio por las técnicas moleculares empleadas.

Los intentos de cultivo en esta tesis de muestras de los seis tipos de colonias detectadas han proporcionado 136 cultivos puros. Algunos de estos aislamientos correspondientes a los géneros *Crossiella* o *Pseudonocardia* pertenecían a la misma familia que los componentes mayoritarios, aunque basados en las secuencias del 16S rRNA, ninguno resultó ser alguno de ellos. Estudios de cultivo anteriores permitieron el aislamiento de bacterias pertenecientes a los géneros *Acinetobacter*, *Agromyces*, *Afipia*, *Aneurinibacillus*, *Arthrobacter*, *Bacillus*, *Brevibacillus*, *Brevundimonas*,

*Catellatospora, Cellulomonas, Gordonia, Hydrogenophaga, Inquilinus, Janibacter, Kibdelosporangium, Lechevalieria, Methylobacterium, Methylobacterium, Micromonospora, Mycobacterium, Nocardia, Sphingomonas, Streptomyces, Ochrobactrum, Paenibacillus, Pseudomonas, Pseudonocardia, Microbacterium, Ralstonia, Rhodococcus, Sinorhizobium, Staphylococcus, Stenotrophomonas, Tsukamurella, Variovorax* y algunas otras sin afiliación cercana a bacterias cultivables (Informe CSIC 2007-2009, pag 352-357). Muchos de estos taxones se han vuelto a obtener como aislados puros en este estudio.

Por el contrario, las nuevas especies obtenidas de colonias visibles en Altamira (*Hoyosella, Nocardia, Aurantimonas altamirensis*) no se han conseguido recuperar en los intentos de cultivo o en la lista de componentes más abundantes por las técnicas moleculares realizados en esta tesis, indicando que podían tratarse de componentes minoritarios de las colonias o de la microbiota adicional del soporte. En otro estudio hecho en las colonias presentes en la cueva de Altamira y en la cueva Tito Bustillo se subraya la dificultad de crecer los componentes de las colonias (Laiz et al.,2008). De nuevo, los resultados de cultivo obtenidos en esta tesis siguen apoyando la idea de que los componentes de las colonias son difíciles de crecer, y de la diferencia tan grande que hay entre lo que se obtiene por métodos clásicos de cultivo y lo que nos muestran los métodos de la metagenómica.

La incultivabilidad de los componentes de las colonias, así como la poca información presente en las bases de datos sobre ellos, plantea que los métodos de control microbiológicos empleados hasta ahora, aunque proporcionan información válida del estado microbiológico de la cueva no se están dirigiendo adecuadamente al control de las colonias que solo podrán ser seguidos por métodos moleculares sin cultivo o por técnicas de imagen.

La obtención de borradores de buena calidad de los componentes más predominantes en las colonias ha permitido el desarrollo de primers específicos contra estos componentes y la capacidad de poder detectarlos de una manera rápida, sencilla y poco costosa (Figura 66). Además, el uso de estos primers

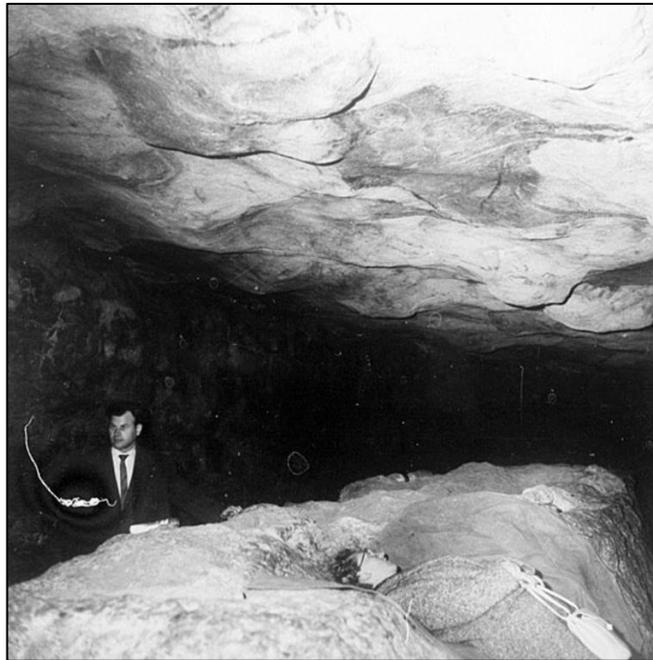
también se podría emplear en otras técnicas como PCR cuantitativa o FISH entre otras como complementación para una mayor detección de estos componentes.

La estructura de las colonias visibles se ha estudiado previamente por la técnica de microscopía electrónica de barrido ambiental que revela detalles estructurales y algún detalle funcional asociado con procesos de biomineralización (Cañaveras et al., 2001; Sanchez-Moral et al., 2003).

El uso de primers como sondas fluorescentes por la técnica FISH y su observación bajo un microscopio de fluorescencia revelaría la organización espacio-temporal de los componentes de las colonias e incluso podría llegar a ofrecer detalles de posibles relaciones nutricionales o fisiológicas entre ellos. El planteamiento de estrategias destinadas al control frente al avance de estas colonias en las paredes y techos de Altamira podrían derivarse de un estudio más exhaustivo de los borradores de estos genomas al igual que de la anotación funcional de las comunidades presentes en ellas, prestando atención a genes implicados en las interacciones sociales de los componentes como son los sistemas CRISPR o sistemas toxina-antitoxina.

De momento, el control y seguimiento de parámetros ambientales y microbiológicos, así como el cuidado de no alterar el ambiente ni introducir nada exógeno a la cueva son las mejores medidas preventivas para evitar un crecimiento indeseado de estas colonias.





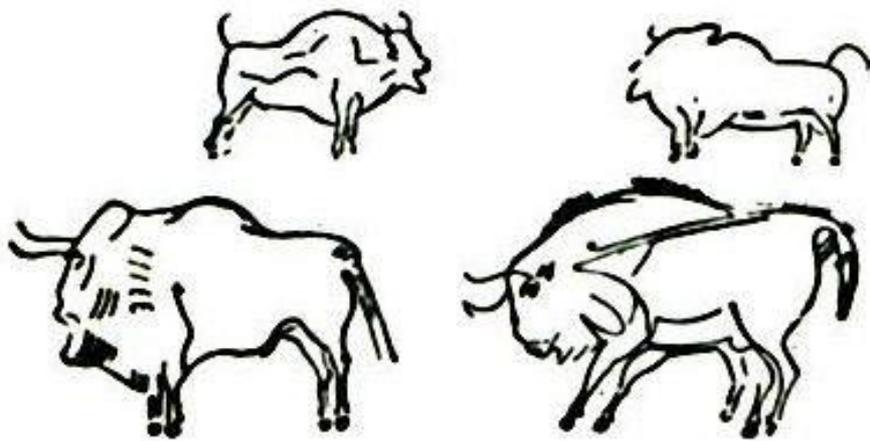
## 7. Conclusiones



- La secuenciación masiva nos ha proporcionado una imagen completa de la composición de las colonias visibles de Altamira.
- Las colonias de un mismo color no son homogéneas, las colonias blancas y grises se han dividido en 2 subtipos, cada una en base a los componentes que las forman.
- La diversidad en las colonias es media-baja en comparación con lo habitual en muestras de suelos. La composición es bastante diferente entre ellas y muy distinta de la composición del soporte adyacente.
- El número de componentes considerados más abundantes en las colonias es bajo, oscila entre 6 y 34 componentes siendo las colonias beis las más sencillas y las grises las más diversas.
- Las colonias presentan uno o dos componentes muy abundantes a los que hemos llamado componentes mayoritarios, suelen representar más del 20% de las células totales y son característicos para cada tipo de colonia.
- Los componentes mayoritarios de las colonias clasifican dentro del filo *Actinobacteria*, en dos taxones diferentes: la familia *Pseudonocardiaceae* y la clase *Nitriliruptoria*. Constituyen un grupo bacteriano muy adaptado a la formación de biofilms coloreados en cuevas.
- Los componentes mayoritarios de la familia *Pseudonocardiaceae* representan probablemente 5 especies distintas dentro de un género nuevo.

## CONCLUSIONES

- Los componentes mayoritarios de la clase *Nitriliruptoria* representan 2 especies pertenecientes a géneros diferentes formando posiblemente un nuevo orden en la clase.
- Los componentes de las colonias visibles de Altamira presentan alta identidad de secuencia con bacterias detectadas en otras cuevas del mundo, sugiriendo que existe un grupo de bacterias muy adaptado a la formación de biofilms coloreados en cuevas.
- Las comparaciones de los diferentes tipos de colonias de Altamira sugieren un modelo general de composición de colonia aplicable a las colonias de Altamira y probablemente a cualquier biofilm similar formado en cuevas.
- Combinaciones diferentes de los mismos componentes pueden dar lugar a nuevos tipos (colores) de colonia.
- El potencial metabólico de las colonias soporta la posibilidad de una forma de vida mixotrófica, alternando periodos heterotróficos y autotróficos, dependiendo de la disponibilidad de nutrientes.
- Los intentos de cultivo de los componentes de las colonias no han permitido recuperar ninguno de los componentes mayoritarios o secundarios detectados por las técnicas de secuenciación masiva.
- Se han diseñado primers que amplifican y permiten identificar de forma específica los componentes mayoritarios de las colonias.
- El diseño de primers frente a los componentes mayoritarios, proporciona un método de detección precoz para el control del biodeterioro y una enorme capacidad analítica para el estudio de las colonias.



## 8. Referencias



- Addesso, R., Gonzalez-Pimentel, J.L., D'Angeli, I.M., De Waele, J., Saiz-Jimenez, C., Jurado, V., Miller, A.Z., Cubero, B., Vigliotta, G., and Baldantoni, D. (2020). Microbial community characterizing vermiculations from karst caves and its role in their formation. *Microb. Ecol.*
- Alfreider, A., Vogt, C., Geiger-Kaiser, M., and Psenner, R. (2009). Distribution and diversity of autotrophic bacteria in groundwater systems based on the analysis of RubisCO genotypes. *Syst. Appl. Microbiol.* *32*, 140–150.
- 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.
- Barton, H.A., and Jurado, V. (2007). What's up down there? Microbial diversity in caves. *Microbe* *2*: 132-138
- Barton, H.A., Giarrizzo, J.G., Suarez, P., Robertson, C.E., Broering, M.J., Banks, E.D., Vaishampayan, P.A., and Venkateswaran, K. (2014). Microbial diversity in a Venezuelan orthoquartzite cave is dominated by the *Chloroflexi* (Class *Ktedonobacterales*) and *Thaumarchaeota* Group I.1c. *Front. Microbiol.* *5*, 615.
- Bartossek, R., Spang, A., Weidler, G., Lanzen, A., and Schleper, C. (2012). Metagenomic analysis of ammonia-oxidizing archaea affiliated with the soil group. *Front. Microbiol.* *3*, 208.
- Bastian, F., Jurado, V., Nováková, A., Alabouvette, C., and Saiz-Jimenez, C. (2010). The microbiology of Lascaux Cave. *Microbiol. Read. Engl.* *156*, 644–652.
- Bay, S.K., Dong, X., Bradley, J.A., Leung, P.M., Grinter, R., Jirapanjawan, T., Arndt, S.K., Cook, P.L.M., LaRowe, D.E., Nauer, P.A., et al. (2021). Trace gas oxidizers are widespread and active members of soil microbial communities. *Nat. Microbiol.* *6*, 246–256.
- Berdy, B., Spoering, A.L., Ling, L.L., and Epstein, S.S. (2017). *In situ* cultivation of previously uncultivable microorganisms using the ichip. *Nat. Protoc.* *12*, 2232–2242.
- Biswas, J., Sharma, K., Harris, K.K., and Rajput, Y. (2013). Biodeterioration agents: Bacterial and fungal diversity dwelling in or on the pre-historic rock-paints of Kabra-pahad, India. *Iran. J. Microbiol.* *5*, 309–314.
- Blehert, D.S., Hicks, A.C., Behr, M., Meteyer, C.U., Berlowski-Zier, B.M., Buckles, E.L., Coleman, J.T.H., Darling, S.R., Gargas, A., Niver, R., et al. (2009). Bat white-nose syndrome: an emerging fungal pathogen? *Science* *323*, 227.
- Bollmann, A., Lewis, K., and Epstein, S.S. (2007). Incubation of environmental samples in a diffusion chamber increases the diversity of recovered isolates. *Appl. Environ. Microbiol.* *73*, 6386–6390.
- Branda, S.S., Vik, S., Friedman, L., and Kolter, R. (2005). Biofilms: the matrix revisited. *Trends Microbiol.* *13*, 20–26.
- Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., and Banfield, J.F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* *523*, 208–211.

- Bruns, A., Cypionka, H., and Overmann, J. (2002). Cyclic AMP and acyl homoserine lactones increase the cultivation efficiency of heterotrophic bacteria from the central Baltic Sea. *Appl. Environ. Microbiol.* *68*, 3978–3987.
- Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* *12*, 59–60.
- Bukin, Y.S., Galachyants, Y.P., Morozov, I.V., Bukin, S.V., Zakharenko, A.S., and Zemskaya, T.I. (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci. Data* *6*, 190007.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* *13*, 581–583.
- Cañaveras, J.C., Hoyos Gómez, M., Sánchez Moral, S., Sanz Rubio, E., Bedoya, J., Hoyos, V., Groth, I., Schumann, P., and Laiz Trobajo, L. (1999). Microbial communities associated with hydromagnesite and needle-fiber aragonite deposits in a karstic cave (Altamira, Northern Spain). *Geomicrobiology Journal* *16* (1): 9-25
- Cañaveras, J.C., Sánchez Moral, S., Sloer, V., and Sáiz-Jiménez, C. (2001). Microorganisms and microbially induced fabrics in cave walls. *Geomicrobiology Journal* *18*(3): 223-240
- Cao, C., Jiang, J., Sun, H., Huang, Y., Tao, F., and Lian, B. (2016). Carbonate mineral formation under the influence of limestone-colonizing *Actinobacteria*: morphology and polymorphism. *Front. Microbiol.* *7*.
- Castelle, C.J., and Banfield, J.F. (2018). Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* *172*, 1181–1197.
- Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., Frischkorn, K.R., Tringe, S.G., Singh, A., Markillie, L.M., et al. (2015). Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol. CB* *25*, 690–701.
- Chakravorty, S., Helb, D., Burday, M., Connell, N., and Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* *69*, 330–339.
- Chaudhary, D.K., Khulan, A., and Kim, J. (2019). Development of a novel cultivation technique for uncultured soil bacteria. *Sci. Rep.* *9*, 6666.
- Chelius, M., Beresford, G., Horton, H., Quirk, M., Selby, G., Simpson, R., Horrocks, R., and Moore, J. (2009). Impacts of alterations of organic inputs on the bacterial community within the sediments of Wind Cave, South Dakota, USA. *Int. J. Speleol.* *38*.
- Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A.M., and Banfield, J.F. (2019). Accurate and complete genomes from metagenomes. *BioRxiv* 808410.
- Chronáková, A., Horák, A., Elhottová, D., and Kristůfek, V. (2009). Diverse archaeal community of a bat guano pile in Domica Cave (Slovak Karst, Slovakia). *Folia Microbiol. (Praha)* *54*, 436–446.

- Codrón, J.C.G. (1983). La caliza y el karst en la Península Ibérica. *Anales de geografía de la Universidad Complutense*, 3, 277.
- Coenye, T., Goris, J., Spilker, T., Vandamme, P., and LiPuma, J.J. (2002). Characterization of unusual bacteria isolated from respiratory secretions of cystic fibrosis patients and description of *Inquilinus limosus* gen. nov., sp. nov. *J. Clin. Microbiol.* 40, 2062–2069.
- Crapoulet, N., Robineau, S., Raoult, D., and Renesto, P. (2005). Intervening sequence acquired by lateral gene transfer in *Tropheryma whipplei* Results in 23S rRNA Fragmentation. *Appl. Environ. Microbiol.* 71, 6698.
- Cross, K.L., Campbell, J.H., Balachandran, M., Campbell, A.G., Cooper, S.J., Griffen, A., Heaton, M., Joshi, S., Klingeman, D., Leys, E., et al. (2019). Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat. Biotechnol.* 37, 1314–1321.
- Cuezva, S., Sanchez-Moral, S., Saiz-Jimenez, C., and Cañaveras, J. (2009). Microbial communities and associated mineral fabrics in Altamira Cave, Spain. *Int. J. Speleol.* 38.
- Cuezva, S., Fernandez-Cortes, A., Porca, E., Pašić, L., Jurado, V., Hernandez-Marine, M., Serrano-Ortiz, P., Hermosin, B., Cañaveras, J.C., Sanchez-Moral, S., et al. (2012). The biogeochemical role of *Actinobacteria* in Altamira Cave, Spain. *FEMS Microbiol. Ecol.* 81, 281–290.
- Czerwik-Marcinkowska, J., and Mrozińska, T. (2011). Algae and *Cyanobacteria* in caves of the Polish Jura. *Polish Botanical Journal*.
- Daims, H., and Wagner, M. (2018). *Nitrospira*. *Trends Microbiol.* 26, 462–463.
- Daims, H., Lebedeva, E.V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatinszky, M., Vierheilig, J., Bulaev, A., et al. (2015). Complete nitrification by *Nitrospira* bacteria. *Nature* 528, 504–509.
- De Mandal, S., Chatterjee, R., and Kumar, N.S. (2017). Dominant bacterial phyla in caves and their predicted functional roles in C and N cycle. *BMC Microbiol.* 17.
- Díaz Herráiz, M., Jurado, V., Cuezva, S., Laiz Trobajo, L., Pallecchi, P., Tiano, P., Sánchez Moral, S., and Sáiz-Jiménez, C. (2013). The actinobacterial colonization of etruscan paintings.
- Docampo, S., Trigo, M.M., Recio, M., Melgar, M., García-Sánchez, J., and Cabezudo, B. (2011). Fungal spore content of the atmosphere of the Cave of Nerja (southern Spain): Diversity and origin. *Sci. Total Environ.* 409, 835–843.
- Dong, X., and Strous, M. (2019). An integrated pipeline for annotation and visualization of metagenomic Contigs. *Front. Genet.* 10.
- D’Onofrio, A., Crawford, J.M., Stewart, E.J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., and Lewis, K. (2010). Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* 17, 254–264.
- 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, 2194–2200.

- Elshahed, M.S., Senko, J.M., Najar, F.Z., Kenton, S.M., Roe, B.A., Dewers, T.A., Spear, J.R., and Krumholz, L.R. (2003). Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. *Appl. Environ. Microbiol.* *69*, 5609–5621.
- Engel, A.S., Porter, M.L., Stern, L.A., Quinlan, S., and Bennett, P.C. (2004). Bacterial diversity and ecosystem function of filamentous microbial mats from aphotic (cave) sulfidic springs dominated by chemolithoautotrophic “*Epsilonproteobacteria*”. *FEMS Microbiol. Ecol.* *51*, 31–53.
- Enning, D., and Garrelfs, J. (2014). Corrosion of iron by sulfate-reducing bacteria: New Views of an Old Problem. *Appl. Environ. Microbiol.* *80*, 1226–1236.
- Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015). Anvi’o: an advanced analysis and visualization platform for ‘omics data. *PeerJ* *3*.
- Fernández Cortés, A., Cuezva, S., Sánchez Moral, S., Cañaveras, J.C., Porca, E., Jurado, V., Martín Sánchez, P.M., and Sáiz-Jiménez, C. (2011). Detection of human-induced environmental disturbances in a show cave. *Environ Sci Pollut Res Int.*;18(6):1037-45.
- Ferrari, B.C., Binnerup, S.J., and Gillings, M. (2005). Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. *Appl. Environ. Microbiol.* *71*, 8714–8720.
- García Antón, E., Fernández-Cortés, A., Álvarez-Gallego, M., and Sánchez Moral, S. (2013). Entry and dispersion of microorganisms inside Altamira Cave: New evidences from aerobiological and atmospheric gases surveys (Taylor & Francis). *Science and Technology for the Conservation of Cultural Heritage*: 55-58.
- García Sánchez, A., Ariza Astolfi, C., Úbeda Ontiveros, J.M., Martín Sánchez, P.M., Jurado, V., Bastian, F., Alabouvette, C., and Sáiz-Jiménez, C. (2013). Free-living amoebae in sediments from the Lascaux Cave in France. *International Journal of Speleology*; *42* (1) 9-13.
- Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* *17*, 175–188.
- Gázquez, F., Quindós-Poncela, L., Sainz-Fernández, C., Fernández-Villar, A., Fuente-Merino, I., and Celaya-Gonzalez, S. (2016). Spatiotemporal distribution of  $\delta(13)CCO_2$  in a shallow cave and its potential use as indicator of anthropic pressure. *J. Environ. Manage.* *180*, 421–432.
- Geesink, P., Wegner, C.-E., Probst, A.J., Herrmann, M., Dam, H.T., Kaster, A.-K., and Küsel, K. (2020). Genome-inferred spatio-temporal resolution of an uncultivated Roizmanbacterium reveals its ecological preferences in groundwater. *Environ. Microbiol.* *22*, 726–737.
- Goldscheider, N., Chen, Z., Auler, A.S., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Stevanovic, Z., et al. (2020). Global distribution of carbonate rocks and karst water resources. *Hydrogeol. J.* *28*, 1661–1677.
- Gómez-Alarcón, G., Muñoz, M.L., and Flores, M. (1994). Excretion of organic acids by fungal strains isolated from decayed sandstone. *Int. Biodeterior. Biodegrad.* *34*, 169–180.
- Gonzalez, J.M., Portillo, M.C., and Saiz-Jimenez, C. (2006). Metabolically active *Crenarchaeota* in Altamira Cave. *Naturwissenschaften* *93*, 42–45.

- Gonzalez-Pimentel, J.L., Miller, A.Z., Jurado, V., Laiz, L., Pereira, M.F.C., and Saiz-Jimenez, C. (2018). Yellow coloured mats from lava tubes of La Palma (Canary Islands, Spain) are dominated by metabolically active *Actinobacteria*. *Sci. Rep.* *8*, 1944.
- Gray, D.A., Dugar, G., Gamba, P., Strahl, H., Jonker, M.J., and Hamoen, L.W. (2019). Extreme slow growth as alternative strategy to survive deep starvation in bacteria. *Nat. Commun.* *10*, 1–12.
- Greening, C., Berney, M., Hards, K., Cook, G.M., and Conrad, R. (2014). A soil actinobacterium scavenges atmospheric H<sub>2</sub> using two membrane-associated, oxygen-dependent [NiFe] hydrogenases. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 4257–4261.
- Grobbelaar, J.U. (2000). Algae: a major threat to show caves. *SIL Proc.* 1922-2010 *27*, 916–919.
- Groster, A., and Alvarez-Cohen, L. (2013). RubisCO-based CO<sub>2</sub> fixation and C<sub>1</sub> metabolism in the actinobacterium *Pseudonocardia dioxanivorans* CB1190. *Environ. Microbiol.* *15*, 3040–3053.
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* *52*, 696–704.
- Gupta, R.Kr., and Srivastava, R. (2012). Resuscitation promoting factors: a family of microbial proteins in survival and resuscitation of dormant *Mycobacteria*. *Indian J. Microbiol.* *52*, 114–121.
- Hathaway, J.J.M., Sinsabaugh, R.L., Dapkevicius, M.D.L.N.E., and Northup, D.E. (2014). Diversity of ammonia oxidation (*amoA*) and nitrogen fixation (*nifH*) genes in Lava Caves of Terceira, Azores, Portugal. *Geomicrobiol. J.* *31*, 221–235.
- He, H., Fu, L., Liu, Q., Fu, L., Bi, N., Yang, Z., and Zhen, Y. (2019). Community structure, abundance and potential functions of bacteria and archaea in the Sansha Yongle Blue Hole, Xisha, South China Sea. *Front. Microbiol.* *10*.
- Hernández, M., Quijada, N.M., Rodríguez-Lázaro, D., and Eiros, J.M. (2020). Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico clínico. *Rev. Argent. Microbiol.* *52*, 150–161.
- Herrmann, M., Wegner, C.-E., Taubert, M., Geesink, P., Lehmann, K., Yan, L., Lehmann, R., Totsche, K.U., and Küsel, K. (2019). Predominance of *Cand.* *Patescibacteria* in groundwater is caused by their preferential mobilization from soils and flourishing under oligotrophic conditions. *Front. Microbiol.* *10*.
- Hess, W.H. (1900). The origin of nitrates in cavern earths. *J. Geology.* *8*, 129–134.
- Hill, C.A. (1981). Origin of Cave Saltpeter. *J. Geology.* *89*, 252–259.
- Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* *34*, 2115–2122.
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical,

functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* *47*, D309–D314.

Hug, L.A., Thomas, B.C., Sharon, I., Brown, C.T., Sharma, R., Hettich, R.L., Wilkins, M.J., Williams, K.H., Singh, A., and Banfield, J.F. (2016). Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ. Microbiol.* *18*, 159–173.

Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* *180*, 4765–4774.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* *11*, 119.

Ibal, J.C., Pham, H.Q., Park, C.E., and Shin, J.-H. (2019). Information about variations in multiple copies of bacterial 16S rRNA genes may aid in species identification. *PLoS ONE* *14*.

Ivanova, V., Tomova, I., Kamburov, A., Tomova, A.B., Vasileva-Tonkova, E., and Kambourova, M. (2013). High phylogenetic diversity of bacteria in the area of prehistoric paintings in magura cave, bulgaria. *Journal of Cave & Karst Studies* . Vol. 75 Issue 3, p218-228. 11p.

Jennings, L.K., Storek, K.M., Ledvina, H.E., Coulon, C., Marmont, L.S., Sadovskaya, I., Secor, P.R., Tseng, B.S., Scian, M., Filloux, A., et al. (2015). Pel is a cationic exopolysaccharide that cross-links extracellular DNA in the *Pseudomonas aeruginosa* biofilm matrix. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 11353–11358.

Jensen, L.J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., and Bork, P. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* *36*, D250-254.

Ji, M., Greening, C., Vanwonterghem, I., Carere, C.R., Bay, S.K., Steen, J.A., Montgomery, K., Lines, T., Beardall, J., Dorst, J. van, et al. (2017). Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature*. Dec 21;552(7685):400-40

Jones, W.K., and White, W.B. (2019). Chapter 72 - Karst. In *Encyclopedia of Caves* (Third Edition), W.B. White, D.C. Culver, and T. Pipan, eds. (Academic Press), pp. 609–618.

Jones, D.S., Tobler, D.J., Schaperdoth, I., Mainiero, M., and Macalady, J.L. (2010). Community Structure of Subsurface Biofilms in the Thermal Sulfidic Caves of Acquasanta Terme, Italy. *Appl. Environ. Microbiol.* *76*, 5902–5910.

Jung, H.-M., Lee, J.-S., Bae, H.-M., Yi, T.-H., Kim, S.-Y., Lee, S.-T., and Im, W.-T. (2011). *Inquilinus ginsengisoli* sp. nov., isolated from soil of a ginseng field. *Int. J. Syst. Evol. Microbiol.* *61*, 201–204.

Jurado, V., Gonzalez, J.M., Laiz, L., and Saiz-Jimenez, C. (2006). *Aurantimonas altamirensis* sp. nov., a member of the order *Rhizobiales* isolated from Altamira Cave. *Int. J. Syst. Evol. Microbiol.* *56*, 2583–2585.

Jurado, V., Boiron, P., Kroppenstedt, R.M., Laurent, F., Couble, A., Laiz, L., Klenk, H.-P., González, J.M., Saiz-Jimenez, C., Mouniée, D., et al. (2008). *Nocardia altamirensis* sp. nov., isolated from Altamira cave, Cantabria, Spain. *Int. J. Syst. Evol. Microbiol.* *58*, 2210–2214.

- Jurado, V., Fernandez-Cortes, A., Cuezva, S., Laiz, L., Cañaveras, J.C., Sanchez-Moral, S., and Saiz-Jimenez, C. (2009a). The fungal colonisation of rock-art caves: experimental evidence. *Naturwissenschaften* *96*, 1027–1034.
- Jurado, V., Kroppenstedt, R.M., Saiz-Jimenez, C., Klenk, H.-P., Mounié, D., Laiz, L., Couble, A., Pötter, G., Boiron, P., and Rodríguez-Nava, V. (2009b). *Hoyosella altamirensis* gen. nov., sp. nov., a new member of the order *Actinomycetales* isolated from a cave biofilm. *Int. J. Syst. Evol. Microbiol.* *59*, 3105–3110.
- Kakumanu, M.L., and Williams, M.A. (2012). Soil diffusion system enriches the growth of diverse and previously uncultivated bacterial taxa. *Soil Biology & Biochemistry*.
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* *28*, 27–30.
- Kerkhof, L.J., Dillon, K.P., Häggblom, M.M., and McGuinness, L.R. (2017). Profiling bacterial communities by MinION sequencing of ribosomal operons. *Microbiome* *5*.
- van Kessel, M.A.H.J., Speth, D.R., Albertsen, M., Nielsen, P.H., Op den Camp, H.J.M., Kartal, B., Jetten, M.S.M., and Lüscher, S. (2015). Complete nitrification by a single microorganism. *Nature* *528*, 555–559.
- Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* *64*, 346–351.
- Koch, M., Rudolph, C., Moissl, C., and Huber, R. (2006). A cold-loving crenarchaeon is a substantial part of a novel microbial community in cold sulphidic marsh water. *FEMS Microbiol. Ecol.* *57*, 55–66.
- Kong, H.G., Bae, J.Y., Lee, H.J., Joo, H.J., Jung, E.J., Chung, E., and Lee, S.-W. (2014). Induction of the viable but nonculturable state of *Ralstonia solanacearum* by low temperature in the soil microcosm and its resuscitation by catalase. *PLoS One* *9*, e109792.
- Konstantinidis, K.T., and Tiedje, J.M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol.* *10*, 504–509.
- Kouki, S., Saidi, N., M'hiri, F., Nasr, H., Cherif, H., Ouzari, H., and Hassen, A. (2011). Isolation and characterization of facultative mixotrophic ammonia-oxidizing bacteria from constructed wetlands. *J. Environ. Sci.* *23*, 1699–1708.
- Krakova, L., De Leo, F., Bruno, L., Pangallo, D., and Urzi, C. (2015). Complex bacterial diversity in the white biofilms of the Catacombs of St. Callixtus in Rome evidenced by different investigation strategies. *Environ. Microbiol.* *17*, 1738–1752.
- Kumar, S., Tamura, K., and Nei, M. (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci.* *10*, 189–191.
- Labrada, L., Salgado, J.M., Espinosa, J., and Luque, C.G. (2010). Fauna invertebrada de las cavidades de Cantabria. *Locustella Anu. Nat. Cantab.* *28*–43.
- Laiz, L., Groth, I., Gonzalez, I., and Saiz-Jimenez, C. (1999). Microbiological study of the dripping waters in Altamira cave (Santillana del Mar, Spain). *J. Microbiol. Methods* *36*, 129–138.

- Laiz, L., Gonzalez, J.M., and Saiz-Jimenez, C. Microbial communities in caves: ecology, physiology, and effects on paleolithic paintings. 28.
- Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U. S. A.* 82, 6955–6959.
- Lavoie, K.H., Winter, A.S., Read, K.J.H., Hughes, E.M., Spilde, M.N., and Northup, D.E. (2017). Comparison of bacterial communities from lava cave microbial mats to overlying surface soils from Lava Beds National Monument, USA. *PLoS ONE* 12.
- Lee, I., Ouk Kim, Y., Park, S.-C., and Chun, J. (2016). OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 66, 1100–1103.
- Lee, N.M., Meisinger, D.B., Aubrecht, R., Kovacic, L., Saiz-Jimenez, C., Baskar, S., Baskar, R., Liebl, W., Porter, M.L., and Engel, A.S. (2012). Caves and karst environments. In *Life at extremes: environments, organisms and strategies for survival*, E.M. Bell, ed. (Wallingford: CABI), pp. 320–344.
- León-Zayas, R., Peoples, L., Biddle, J.F., Podell, S., Novotny, M., Cameron, J., Lasken, R.S., and Bartlett, D.H. (2017). The metabolic potential of the single cell genomes obtained from the Challenger Deep, Mariana Trench within the *Candidate* Superphylum Parcubacteria (OD1). *Environ. Microbiol.* 19, 2769–2784.
- Lewin, G.R., Carlos, C., Chevrette, M.G., Horn, H.A., McDonald, B.R., Stankey, R.J., Fox, B.G., and Currie, C.R. (2016). Evolution and ecology of *Actinobacteria* and their bioenergy applications. *Annu. Rev. Microbiol.* 70, 235–254.
- Li, L., Mendis, N., Trigui, H., Oliver, J.D., and Faucher, S.P. (2014). The importance of the viable but non-culturable state in human bacterial pathogens. *Front. Microbiol.* 5, 258.
- Liem, M., Regensburg-Tuink, A.J.G., Henkel, C.V., and Spaink, H.P. (2020). Microbial diversity characterization of seawater in a pilot study using Oxford Nanopore Technologies long-read sequencing. *BMC Research Notes* volume 14, Article number: 42.
- Liu, J., Deng, Y., Li, L., Li, B., Li, Y., Zhou, S., Shirliff, M.E., Xu, Z., and Peters, B.M. (2018). Discovery and control of culturable and viable but non-culturable cells of a distinctive *Lactobacillus harbinensis* strain from spoiled beer. *Sci. Rep.* 8, 11446.
- Lorch, J.M., Palmer, J.M., Lindner, D.L., Ballmann, A.E., George, K.G., Griffin, K., Knowles, S., Huckabee, J.R., Haman, K.H., Anderson, C.D., et al. (2016). First detection of bat white-nose syndrome in western North America. *MSphere* 1.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, null, Buchner, A., Lai, T., Steppi, S., Jobb, G., et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371.
- Macalady, J.L., Dattagupta, S., Schaperdoth, I., Jones, D.S., Druschel, G.K., and Eastman, D. (2008). Niche differentiation among sulfur-oxidizing bacterial populations in cave waters. *ISME J.* 2, 590–601.

- Martín, C. de las H., and Corruchaga, J.A.L. (2009). Conservar y proteger la cueva de Altamira. Medio siglo de arqueología en el Cantábrico Oriental y su entorno: actas del congreso Internacional, 2009, ISBN 978-84-7821-739-7, págs. 1041-1048, (Instituto Alavés de Arqueología), pp. 1041–1048.
- Martínez, J.L. (1984). Disolución de rocas carbonatadas: Cuantificación del proceso actual de karstificación en el macizo de la Piedra de San Martín (Pirineo Occidental). Cuad. Investig. Geográfica Geogr. Res. Lett. 127–138.
- Méheust, R., Burstein, D., Castelle, C.J., and Banfield, J.F. (2019). The distinction of CPR bacteria from other bacteria based on protein family content. Nat. Commun. 10, 4173.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 386.
- Mukamolova, G.V., Turapov, O.A., Kazarian, K., Telkov, M., Kaprelyants, A.S., Kell, D.B., and Young, M. (2002). The *rpf* gene of *Micrococcus luteus* encodes an essential secreted growth factor. Mol. Microbiol. 46, 611–621.
- Mulec, J., Oarga-Mulec, A., Tomazin, R., and Matos, T. (2015). Characterization and fluorescence of yellow biofilms in karst caves, southwest Slovenia. Int. J. Speleol. 44.
- Muyzer, G., Waal, E.C. de, and Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59, 695–700.
- Mysara, M., Vandamme, P., Props, R., Kerckhof, F.-M., Leys, N., Boon, N., Raes, J., and Monsieus, P. (2017). Reconciliation between operational taxonomic units and species boundaries. FEMS Microbiol. Ecol. 93.
- Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinforma. Oxf. Engl. 29, 2933–2935.
- Niederberger, T.D., Sohm, J.A., Gunderson, T., Tirindelli, J., Capone, D.G., Carpenter, E.J., and Cary, S.C. (2015). Carbon-fixation rates and associated microbial communities residing in arid and ephemerally wet antarctic dry valley soils. Front. Microbiol. 6.
- Northup, D.E., Melim, L.A., Spilde, M.N., Hathaway, J.J.M., Garcia, M.G., Moya, M., Stone, F.D., Boston, P.J., Dapkevicius, M.L.N.E., and Riquelme, C. (2011). Lava cave microbial communities within mats and secondary mineral deposits: implications for life detection on other planets. Astrobiology 11, 601–618.
- Nováková, A., Kubátová, A., Sklenář, F., and Hubka, V. (2018). Microscopic fungi on cadavers and skeletons from cave and mine environments. Czech Mycology 70(2): 101–121.
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834.
- Oliveira, C., Gunderman, L., Coles, C.A., Lochmann, J., Parks, M., Ballard, E., Glazko, G., Rahmatallah, Y., Tackett, A.J., and Thomas, D.J. (2017). 16S rRNA gene-based metagenomic analysis of Ozark Cave bacteria. Diversity 9. 9(3):31.

- Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and Phillippy, A.M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* *17*.
- Ortiz, M., Legatzki, A., Neilson, J.W., Fryslie, B., Nelson, W.M., Wing, R.A., Soderlund, C.A., Pryor, B.M., and Maier, R.M. (2014). Making a living while starving in the dark: metagenomic insights into the energy dynamics of a carbonate cave. *ISME J.* *8*, 478–491.
- Paharik, A.E., and Horswill, A.R. (2016). The Staphylococcal Biofilm: Adhesins, Regulation, and Host Response. *Microbiol. Spectr.* *4*.
- Palomo, A., Pedersen, A.G., Fowler, S.J., Dechesne, A., Sicheritz-Pontén, T., and Smets, B.F. (2018). Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox *Nitrospira*. *ISME J.* *12*, 1779–1793.
- Park, S.W., Park, S.T., Lee, J.E., and Kim, Y.M. (2008). *Pseudonocardia carboxydivorans* sp. nov., a carbon monoxide-oxidizing actinomycete, and an emended description of the genus *Pseudonocardia*. *Int. J. Syst. Evol. Microbiol.* *58*, 2475–2478.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* *25*, 1043–1055.
- Pasić, L., Kovce, B., Sket, B., and Herzog-Velikonja, B. (2010). Diversity of microbial communities colonizing the walls of a Karstic cave in Slovenia. *FEMS Microbiol. Ecol.* *71*, 50–60.
- Pei, A., Nossa, C.W., Chokshi, P., Blaser, M.J., Yang, L., Rosmarin, D.M., and Pei, Z. (2009). Diversity of 23S rRNA Genes within Individual Prokaryotic Genomes. *PLOS ONE* *4*, e5437.
- Pootakham, W., Mhuantong, W., Yoocha, T., Putchim, L., Sonthirod, C., Naktang, C., Thongtham, N., and Tangphatsornruang, S. (2017). High resolution profiling of coral-associated bacterial communities using full-length 16S rRNA sequence data from PacBio SMRT sequencing system. *Sci. Rep.* *7*, 1–14.
- Porca, E., Jurado, V., Martin-Sanchez, P.M., Hermosin, B., Bastian, F., Alabouvette, C., and Saiz-Jimenez, C. (2011). Aerobiology: An ecological indicator for early detection and control of fungal outbreaks in caves. *Ecol. Indic.* *11*, 1594–1598.
- Porca, E., Jurado, V., Žgur-Bertok, D., Saiz-Jimenez, C., and Pašić, L. (2012). Comparative analysis of yellow microbial communities growing on the walls of geographically distinct caves indicates a common core of microorganisms involved in their formation. *FEMS Microbiol. Ecol.* *81*, 255–266.
- Portillo, M.C., Gonzalez, J.M., and Saiz-Jimenez, C. (2008). Metabolically active microbial communities of yellow and grey colonizations on the walls of Altamira Cave, Spain. *J. Appl. Microbiol.* *104*, 681–691.
- Portillo, M.C., Saiz-Jimenez, C., and Gonzalez, J.M. (2009). Molecular characterization of total and metabolically active bacterial communities of “white colonizations” in the Altamira Cave, Spain. *Res. Microbiol.* *160*, 41–47.

- Portillo Guisado, M. del C., Sáiz-Jiménez, C., and González Grau, J.M. (2009). Molecular characterization of total and metabolically active bacterial communities of “white colonizations” in the Altamira Cave, Spain. *Research in Microbiology* 160(1): 41-47.
- Pronk, L.M., and Sanderson, K.E. (2001). Intervening sequences in *rrl* genes and fragmentation of 23S rRNA in genera of the family *Enterobacteriaceae*. *J. Bacteriol.* 183, 5782–5787.
- Pruesse, E., Peplies, J., and Glöckner, F.O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28, 1823–1829.
- Pu, G., Lv, Y., Dong, L., Zhou, L., Huang, K., Zeng, D., Mo, L., and Xu, G. (2019). Profiling the bacterial diversity in a typical karst tiankeng of China. *Biomolecules* 9.
- Pulschen, A.A., Bendia, A.G., Fricker, A.D., Pellizari, V.H., Galante, D., and Rodrigues, F. (2017). Isolation of uncultured bacteria from Antarctica using long incubation periods and low nutritional media. *Front. Microbiol.* 8.
- Pusz, W., Ogórek, R., Knapik, R., Kozak, B., and Bujak, H. (2015). The occurrence of fungi in the recently discovered Jarkowicka Cave in the Karkonosze Mts. (Poland). *Geomicrobiol. J.* 32, 59–67.
- Rangseekaew, P., and Pathom-aree, W. (2019). Cave *Actinobacteria* as producers of bioactive metabolites. *Front. Microbiol.* 10.
- Rao, Y., Shang, W., Yang, Y., Zhou, R., and Rao, X. (2020). Fighting mixed-species microbial biofilms with cold atmospheric plasma. *Front. Microbiol.* 11.
- Reynolds, H.T., and Barton, H.A. (2014). Comparison of the white-nose syndrome agent *Pseudogymnoascus destructans* to cave-dwelling relatives suggests reduced saprotrophic enzyme activity. *PLOS ONE* 9, e86437.
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19126–19131.
- Riquelme, C., Marshall Hathaway, J.J., Enes Dapkevicius, M. de L.N., Miller, A.Z., Kooser, A., Northup, D.E., Jurado, V., Fernandez, O., Saiz-Jimenez, C., and Cheeptham, N. (2015). Actinobacterial diversity in volcanic caves and associated geomicrobiological interactions. *Front. Microbiol.* 6.
- Rodriguez-R, L.M., Gunturu, S., Harvey, W.T., Rosselló-Mora, R., Tiedje, J.M., Cole, J.R., and Konstantinidis, K.T. (2018a). The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.*
- Rodriguez-R, L.M., Gunturu, S., Tiedje, J.M., Cole, J.R., and Konstantinidis, K.T. (2018b). Nonpareil 3: Fast estimation of metagenomic coverage and sequence diversity. *MSystems* 3.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584.
- Rosero-Chasoy, G., Rodríguez-Jasso, R.M., Aguilar, C.N., Buitrón, G., Chairez, I., and Ruiz, H.A. (2021). Microbial co-culturing strategies for the production high value compounds, a reliable framework towards sustainable biorefinery implementation – an overview. *Bioresour. Technol.* 321, 124458.

- Rosselló-Móra, R., and Amann, R. (2015). Past and future species definitions for Bacteria and Archaea. *Syst. Appl. Microbiol.* *38*, 209–216.
- Rossmassler, K., Engel, A.S., Twing, K.I., Hanson, T.E., and Campbell, B.J. (2012). Drivers of epsilonproteobacterial community composition in sulfidic caves and springs. *FEMS Microbiol. Ecol.* *79*, 421–432.
- Sainz, C., Rábago, D., Celaya, S., Fernández, E., Quindós, J., Quindós, L., Fernández, A., Fuente, I., Arteché, J.L., and Quindós, L.S. (2018). Continuous monitoring of radon gas as a tool to understand air dynamics in the cave of Altamira (Cantabria, Spain). *Sci. Total Environ.* *624*, 416–423.
- Sainz, C., Rábago, D., Fernández, E., Quindós, J., Quindós, L., Fernández, A., Fuente, I., Arteché, J.L., Quindós, L.S., and Celaya, S. (2020). Variations in radon dosimetry under different assessment approaches in the Altamira Cave. *J. Radiol. Prot. Off. J. Soc. Radiol. Prot.* *40*, 367–380.
- Saiz-Jimenez, C., and Hermosin, B. (1999). Thermally assisted hydrolysis and methylation of dissolved organic matter in dripping waters from the Altamira Cave. *J. Anal. Appl. Pyrolysis* *49*, 337–347.
- Saiz-Jimenez, C., Cuezva, S., Jurado, V., Fernandez-Cortes, A., Porca, E., Benavente, D., Canaveras, J.C., and Sanchez-Moral, S. (2011). Paleolithic art in peril: Policy and science collide at Altamira Cave. *Science* *334*, 42–43.
- Sánchez-Moral, S., Soler, V., Cañaveras, J.C., Sanz-Rubio, E., Van Grieken, R., and Gysels, K. (1999). Inorganic deterioration affecting the Altamira Cave, N Spain: quantitative approach to wall-corrosion (solutional etching) processes induced by visitors. *Sci. Total Environ.* *243–244*, 67–84.
- Sanchez-Moral, S., Canaveras, J.C., Laiz, L., Saiz-Jimenez, C., Bedoya, J., and Luque, L. (2003). Biomediated precipitation of calcium carbonate metastable phases in hypogean environments: A short review. *Geomicrobiol. J.* *20*, 491–500.
- Sand, W. (1997). Microbial mechanisms of deterioration of inorganic substrates—A general mechanistic overview. *Int. Biodeterior. Biodegrad.* *40*, 183–190.
- Sauro, F., Cappelletti, M., Ghezzi, D., Columbu, A., Hong, P.-Y., Zowawi, H.M., Carbone, C., Piccini, L., Vergara, F., Zannoni, D., et al. (2018). Microbial diversity and biosignatures of amorphous silica deposits in orthoquartzite caves. *Sci. Rep.* *8*, 1–14.
- Schabereiter-Gurtner, C., Saiz-Jimenez, C., Piñar, G., Lubitz, W., and Rölleke, S. (2002). Phylogenetic 16S rRNA analysis reveals the presence of complex and partly unknown bacterial communities in Tito Bustillo cave, Spain, and on its Palaeolithic paintings. *Environ. Microbiol.* *4*, 392–400.
- Schabereiter-Gurtner, C., Saiz-Jimenez, C., Piñar, G., Lubitz, W., and Rölleke, S. (2004). Phylogenetic diversity of bacteria associated with Paleolithic paintings and surrounding rock walls in two Spanish caves (Llonín and La Garma). *FEMS Microbiol. Ecol.* *47*, 235–247.
- Sedlar, K., Kupkova, K., and Provaznik, I. (2017). Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. *Comput. Struct. Biotechnol. J.* *15*, 48–55.

Senoh, M., Ghosh-Banerjee, J., Ramamurthy, T., Colwell, R.R., Miyoshi, S.-I., Nair, G.B., and Takeda, Y. (2012). Conversion of viable but nonculturable enteric bacteria to culturable by co-culture with eukaryotic cells. *Microbiol. Immunol.* *56*, 342–345.

Shahi, S.K., Freedman, S.N., and Mangalam, A.K. (2017). Gut microbiome in multiple sclerosis: The players involved and the roles they play. *Gut Microbes* *8*, 607–615.

Sharma, R., Ranjan, R., Kapardar, R.K., and Grover, A. (2005). “Unculturable” bacterial diversity: An untapped resource. *Curr. Sci.* *89*, 72–77.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* *7*, 539.

Somavilla, J.F., Khayyat, N., and Arroyo, V. (19780000). A comparative study of the microorganisms present in the Altamira and La Pasiega caves. *Int. Biodeterior. Bull.* *14*, 103–109.

Spruit, C.J.P., and Wanklyn, J.N. (1951). Iron/sulphide ratios in corrosion by sulphate-reducing bacteria. *Nature* *168*, 951–952.

St. John, E., Liu, Y., Podar, M., Stott, M.B., Meneghin, J., Chen, Z., Lagutin, K., Mitchell, K., and Reysenbach, A.-L. (2019). A new symbiotic nanoarchaeote (*Candidatus* Nanoclepta minutus) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst. Appl. Microbiol.* *42*, 94–106.

Staley, J.T., and Konopka, A. (1985). Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* *39*, 321–346.

Stewart, E.J. (2012). Growing unculturable bacteria. *J. Bacteriol.* *194*, 4151–4160.

Stomeo, F., Portillo Guisado, M. del C., González Grau, J.M., Laiz Trobajo, L., and Sáiz-Jiménez, C. (2008). *Pseudonocardia* in white colonizations in two caves with Paleolithic paintings. *International Biodeterioration and Biodegradation* *62*(4): 483-486

Sturm, A., and Dworkin, J. (2015). Phenotypic Diversity as a Mechanism to Exit Cellular Dormancy. *Curr. Biol. CB* *25*, 2272–2277.

Tanaka, T., Kawasaki, K., Daimon, S., Kitagawa, W., Yamamoto, K., Tamaki, H., Tanaka, M., Nakatsu, C.H., and Kamagata, Y. (2014). A hidden pitfall in the preparation of agar media undermines microorganism cultivability. *Appl. Environ. Microbiol.* *80*, 7659–7666.

Tindall, B.J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W., and Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* *60*, 249–266.

Tolar, B.B., Wallsgrrove, N.J., Popp, B.N., and Hollibaugh, J.T. (2017). Oxidation of urea derived nitrogen by *Thaumarchaeota* dominated marine nitrifying communities. *Environ. Microbiol.* *19*, 4838–4850.

Tolba, M.E.M., Huseein, E.A.M., Farrag, H.M.M., Mohamed, H.E.D., Kobayashi, S., Suzuki, J., Ali, T.A.M., and Sugano, S. (2016). *Allovalkampiella spelaea* causing Keratitis in humans. *PLoS Negl. Trop. Dis.* *10*, e0004841.

- Tomczyk-Żak, K., and Zielenkiewicz, U. (2016). Microbial Diversity in Caves. *Geomicrobiol. J.* *33*, 20–38.
- Ulyantsev, V.I., Kazakov, S.V., Dubinkina, V.B., Tyakht, A.V., and Alexeev, D.G. (2016). MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data. *Bioinforma. Oxf. Engl.* *32*, 2760–2767.
- Vanderwolf, K.J., Malloch, D., Mcalpine, D.F., and Forbes, G. (2013). A world review of fungi, yeasts, and slime molds in caves. *International Journal of Speleology.* *42* (1). 77-96.
- Vartoukian, S.R., Palmer, R.M., and Wade, W.G. (2010). Strategies for culture of ‘unculturable’ bacteria: Culturing the unculturable. *FEMS Microbiol. Lett.* no-no.
- Větrovský, T., and Baldrian, P. (2013). The Variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE* *8*.
- Vigneron, A., Cruaud, P., Langlois, V., Lovejoy, C., Culley, A.I., and Vincent, W.F. (2020). Ultra-small and abundant: Candidate phyla radiation bacteria are potential catalysts of carbon transformation in a thermokarst lake ecosystem. *Limnol. Oceanogr. Lett.* *5*, 212–220.
- Von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F.H., and Dutilh, B.E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* *20*, 217
- Walochnik, J., and Mulec, J. (2009). Free-living amoebae in carbonate precipitating microhabitats of karst caves and a new vahlkampfiid amoeba, *Allovahlkampfia spelaea* gen. nov., sp. nov. *Acta Protozool.* *2009*, 25–33.
- Wei, C., and Zhao, X. (2018). Induction of viable but nonculturable *Escherichia coli* O157:H7 by Low Temperature and Its Resuscitation. *Front. Microbiol.* *9*.
- Wiseschart, A., Mhuantong, W., Tangphatsornruang, S., Chantasingsh, D., and Pootanakit, K. (2019). Shotgun metagenomic sequencing from Manao-Pee cave, Thailand, reveals insight into the microbial community structure and its metabolic potential. *BMC Microbiol.* *19*, 144.
- Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinforma. Oxf. Engl.* *32*, 605–607.
- Wurch, L., Giannone, R.J., Belisle, B.S., Swift, C., Utturkar, S., Hettich, R.L., Reysenbach, A.-L., and Podar, M. (2016). Genomics-informed isolation and characterization of a symbiotic *Nanoarchaeota* system from a terrestrial geothermal environment. *Nat. Commun.* *7*, 12115.
- Xian, W.-D., Salam, N., Li, M.-M., Zhou, E.-M., Yin, Y.-R., Liu, Z.-T., Ming, Y.-Z., Zhang, X.-T., Wu, G., Liu, L., et al. (2020). Network-directed efficient isolation of previously uncultivated *Chloroflexi* and related bacteria in hot spring microbial mats. *Npj Biofilms Microbiomes* *6*, 1–10.
- Yang, C., Zhao, F., Hou, Q., Wang, J., Li, M., and Sun, Z. (2020). PacBio sequencing reveals bacterial community diversity in cheeses collected from different regions. *J. Dairy Sci.* *103*, 1238–1249.

Zhao, R., Wang, H., Yang, H., Yun, Y., and Barton, H.A. (2017). Ammonia-oxidizing archaea dominate ammonia-oxidizing communities within alkaline cave sediments. *Geomicrobiol. J.* 34, 511–523.

Zhu, H.-Z., Zhang, Z.-F., Zhou, N., Jiang, C.-Y., Wang, B.-J., Cai, L., and Liu, S.-J. (2019). Diversity, distribution and co-occurrence patterns of bacterial communities in a karst cave system. *Front. Microbiol.* 10.

**Otros documentos usados como referencia:**

- Soledad Cuezva, tesis, Universidad Complutense de Madrid (2008)  
[www.irnase.csic.es/users/SOM.Science.334.42.2011/tesis.pdf](http://www.irnase.csic.es/users/SOM.Science.334.42.2011/tesis.pdf)
- Informe CSIC Estudio integral del estado de conservación de la cueva de Altamira y su arte paleolítico (2003 - 2005).
- Informe CSIC Estudio integral del estado de conservación de la cueva de Altamira y su arte paleolítico (2007 - 2009). Perspectivas futuras de conservación.  
<http://www.culturaydeporte.gob.es/dam/jcr:907536f2-b5f8-458a-b92a-37359b648699/pdf24-sanchez-moral-2014-conservacion-altamira.pdf>
- Informe final. Servicio de control y seguimiento del biodeterioro y de las condiciones ambientales del agua de infiltración y de las emisiones de gas radón de la cueva de Altamira. nº de expediente: 2016c1000014.
- Informe del Programa de Investigación para la Conservación Preventiva y Régimen de acceso a la cueva de Altamira (2012-2014) Volumen I  
<https://ipce.culturaydeporte.gob.es/investigacion/conservacion-bienes-culturales/proyecto-altamira.html>
- Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available  
at: [www.genome.gov/sequencingcostsdata](http://www.genome.gov/sequencingcostsdata). Accessed [8/10/2021]





## 9. Anexo



Tabla S1. Metadatos de las muestras analizadas por la técnica de amplicones. En color azul se resaltan las muestras analizadas también por la técnica de shotgun.

Muestra	Color	Tipo de colonia	Fecha de muestreo	Sala de la cueva
<b>B11M1</b>	Beis	BC	30/10/2018	Vestíbulo
<b>B11M2</b>	Beis	BC	30/10/2018	Vestíbulo
<b>B12M1</b>	Beis	BC	28/05/2019	Vestíbulo
<b>B12M2</b>	Beis	BC	28/05/2019	Vestíbulo
<b>B12M3</b>	Beis	BC	28/05/2019	Vestíbulo
<b>B12M4</b>	Beis	BC	28/05/2019	Pasillo Polícromos
<b>B12M5</b>	Beis	BC	28/05/2019	Pasillo Polícromos
<b>B12M6</b>	Beis	BC	28/05/2019	Vestíbulo
<b>G6M1</b>	Gris	G1	24/02/2015	Vestíbulo
<b>G8M1</b>	Gris	G1	23/02/2016	Vestíbulo
<b>G8M2</b>	Gris	G1	23/02/2016	Vestíbulo
<b>G9M1</b>	Gris	G1	13/09/2016	Vestíbulo
<b>G9M2</b>	Gris	G1	13/09/2016	Cruce
<b>G10M1</b>	Gris	G1	28/11/2017	Vestíbulo
<b>G10M2</b>	Gris	G2	28/11/2017	Vestíbulo
<b>G10M3</b>	Gris	G1	29/08/2017	Cruce
<b>G10M4</b>	Gris	G2	29/08/2017	Cruce
<b>G11M3</b>	Gris	G2	30/10/2018	Vestíbulo
<b>G11M4</b>	Gris	G1	30/10/2018	Vestíbulo
<b>G11M5</b>	Gris	G2	30/10/2018	Vestíbulo
<b>G11M6</b>	Gris	G2	30/04/2019	Cruce
<b>G11M7</b>	Gris	G1	30/04/2019	Cruce
<b>G12M1</b>	Gris	G2	27/08/2019	Vestíbulo
<b>G12M2</b>	Gris	G2	27/08/2019	Vestíbulo
<b>G12M3</b>	Gris	G2	27/08/2019	Cruce
<b>G12M4</b>	Gris	G1	27/08/2019	Cruce
<b>W6M1</b>	Blanca	W1	24/02/2015	Vestíbulo
<b>W8M1</b>	Blanca	W2	23/02/2016	Vestíbulo
<b>W8M2</b>	Blanca	W2	23/02/2016	Vestíbulo
<b>W9M1</b>	Blanca	W2	13/09/2016	Cruce
<b>W10M1</b>	Blanca	W2	28/11/2017	Cruce
<b>W10M2</b>	Blanca	W2	28/11/2017	Cruce
<b>W10M3</b>	Blanca	W2	29/08/2017	Cruce
<b>W10M4</b>	Blanca	W2	29/08/2017	Cruce
<b>W11M4</b>	Blanca	W2	30/04/2019	Cruce
<b>W11M5</b>	Blanca	W2	30/04/2019	Cruce
<b>W11M6</b>	Blanca	W2	11/07/2018	Vestíbulo
<b>W11M7</b>	Blanca	W2	11/07/2018	Vestíbulo
<b>W12M1</b>	Blanca	W2	27/08/2019	Cruce
<b>W12M2</b>	Blanca	W1	27/08/2019	Cruce
<b>Y6M1</b>	Amarilla	YC	24/02/2015	Vestíbulo
<b>Y8M1</b>	Amarilla	YC	23/02/2016	Vestíbulo
<b>Y8M2</b>	Amarilla	YC	23/02/2016	Vestíbulo
<b>Y9M1</b>	Amarilla	YC	13/09/2016	Vestíbulo
<b>Y9M2</b>	Amarilla	YC	13/09/2016	Vestíbulo

<b>Y10M1</b>	Amarilla	YC	28/11/2017	Vestíbulo
<b>Y10M2</b>	Amarilla	YC	28/11/2017	Vestíbulo
<b>Y10M3</b>	Amarilla	YC	29/08/2017	Vestíbulo
<b>Y10M4</b>	Amarilla	YC	29/08/2017	Vestíbulo
<b>Y10M5</b>	Amarilla	YC	29/08/2017	Vestíbulo
<b>Y11M3</b>	Amarilla	YC	30/04/2019	Vestíbulo
<b>Y11M4</b>	Amarilla	YC	11/07/2018	Vestíbulo
<b>Y11M5</b>	Amarilla	YC	11/07/2018	Vestíbulo
<b>Y12M1</b>	Amarilla	YC	27/08/2019	Vestíbulo
<b>Y12M2</b>	Amarilla	YC	27/08/2019	Vestíbulo
<b>SOP10M1</b>	Control	CONTROL	28/11/2017	Vestíbulo
<b>SOP10M2</b>	Control	CONTROL	28/11/2017	Vestíbulo

Tabla S2. Metadatos de las muestras de suelo y soporte analizadas por la técnica de amplicones.

Muestra	Sala de la cueva	Tipo de Muestra	Estación	Año
<b>HBPI14</b>	HOYA BISONTE	SOPORTE	INVIERNO	2014
<b>HBPI18</b>	HOYA BISONTE	SOPORTE	INVIERNO	2018
<b>HBPO18</b>	HOYA BISONTE	SOPORTE	OTOÑO	2018
<b>HBPP18</b>	HOYA BISONTE	SOPORTE	PRIMAVERA	2018
<b>HBPV19</b>	HOYA BISONTE	SOPORTE	VERANO	2019
<b>HBSI14</b>	HOYA BISONTE	SUELO	INVIERNO	2014
<b>HBSI18</b>	HOTA BISONTE	SUELO	INVIERNO	2018
<b>HBSO15</b>	HOYA BISONTE	SUELO	OTOÑO	2015
<b>HBSO16</b>	HOYA BISONTE	SUELO	OTOÑO	2016
<b>HBSO18</b>	HOYA BISONTE	SUELO	OTOÑO	2018
<b>Hbsp17</b>	HOYA BISONTE	SUELO	PRIMAVERA	2017
<b>HBSV15</b>	HOYA BISONTE	SUELO	VERANO	2015
<b>HBSV16</b>	HOYA BISONTE	SUELO	VERANO	2016
<b>HBSV18</b>	HOYA BISONTE	SUELO	VERANO	2018
<b>HCPI18</b>	HOYA CABRA	SOPORTE	INVIERNO	2018
<b>HCPO16</b>	HOYA CABRA	SOPORTE	OTOÑO	2016
<b>HCPO18</b>	HOYA CABRA	SOPORTE	OTOÑO	2018
<b>HCPP16</b>	HOYA CABRA	SOPORTE	PRIMAVERA	2016
<b>HCPP17</b>	HOYA CABRA	SOPORTE	PRIMAVERA	2017
<b>HCPP19</b>	HOYA CABRA	SOPORTE	PRIMAVERA	2019
<b>HCPV18</b>	HOYA CABRA	SOPORTE	VERANO	2018
<b>HCPV19</b>	HOYA CABRA	SOPORTE	VERANO	2019
<b>HCSI18</b>	HOYA CABRA	SUELO	INVIERNO	2018
<b>HCSP17</b>	HOYA CABRA	SUELO	PRIMAVERA	2017
<b>HCSV17</b>	HOYA CABRA	SUELO	VERANO	2017
<b>HCSV19</b>	HOYA CABRA	SUELO	VERANO	2019
<b>P1PI14</b>	POLICROMOS	SOPORTE	INVIERNO	2014
<b>P1PI15</b>	POLICROMOS	SOPORTE	INVIERNO	2015
<b>P1PO15</b>	POLICROMOS	SOPORTE	OTOÑO	2015
<b>P1PO16</b>	POLICROMOS	SOPORTE	OTOÑO	2016
<b>P1PP16</b>	POLICROMOS	SOPORTE	PRIMAVERA	2016
<b>P1PV15</b>	POLICROMOS	SOPORTE	VERANO	2015
<b>P1SI14</b>	POLICROMOS	SUELO	INVIERNO	2014
<b>P1SI15</b>	POLICROMOS	SUELO	INVIERNO	2015
<b>P1SO15</b>	POLICROMOS	SUELO	OTOÑO	2015
<b>P1SO16</b>	POLICROMOS	SUELO	OTOÑO	2016
<b>P1SP16</b>	POLICROMOS	SUELO	PRIMAVERA	2016
<b>P1SV15</b>	POLICROMOS	SUELO	VERANO	2015
<b>P1SV16</b>	POLICROMOS	SUELO	VERANO	2016
<b>P2PI14</b>	POLICROMOS	SOPORTE	INVIERNO	2014
<b>P2PI15</b>	POLICROMOS	SOPORTE	INVIERNO	2015
<b>P2PO15</b>	POLICROMOS	SOPORTE	OTOÑO	2015
<b>P2PV15</b>	POLICROMOS	SOPORTE	VERANO	2015
<b>P2SO16</b>	POLICROMOS	SUELO	OTOÑO	2016

<b>P2SV15</b>	POLICROMOS	SUELO	VERANO	2015
<b>P2SV16</b>	POLICROMOS	SUELO	VERANO	2016
<b>PPI18</b>	POLICROMOS	SOPORTE	INVIERNO	2018
<b>PPO18</b>	POLICROMOS	SOPORTE	OTOÑO	2018
<b>PPP17</b>	POLICROMOS	SOPORTE	PRIMAVERA	2017
<b>PPP18</b>	POLICROMOS	SOPORTE	PRIMAVERA	2018
<b>PPP19</b>	POLICROMOS	SOPORTE	PRIMAVERA	2019
<b>PPV18</b>	POLICROMOS	SOPORTE	VERANO	2018
<b>PPV19</b>	POLICROMOS	SOPORTE	VERANO	2019
<b>PSI18</b>	POLICROMOS	SUELO	INVIERNO	2018
<b>PSO18</b>	POLICROMOS	SUELO	OTOÑO	2018
<b>PSP17</b>	POLICROMOS	SUELO	PRIMAVERA	2017
<b>PSP18</b>	POLICROMOS	SUELO	PRIMAVERA	2018
<b>PSP19</b>	POLICROMOS	SUELO	PRIMAVERA	2019
<b>PSV18</b>	POLICROMOS	SUELO	VERANO	2018
<b>PSV19</b>	POLICROMOS	SUELO	VERANO	2019
<b>VPI14</b>	VESTÍBULO	SOPORTE	INVIERNO	2014
<b>VPI15</b>	VESTÍBULO	SOPORTE	INVIERNO	2015
<b>VPI18</b>	VESTÍBULO	SOPORTE	INVIERNO	2018
<b>VPO15</b>	VESTÍBULO	SOPORTE	OTOÑO	2015
<b>VPO16</b>	VESTÍBULO	SOPORTE	OTOÑO	2016
<b>VPO18</b>	VESTÍBULO	SOPORTE	OTOÑO	2018
<b>VPP16</b>	VESTÍBULO	SOPORTE	PRIMAVERA	2016
<b>VPP17</b>	VESTÍBULO	SOPORTE	PRIMAVERA	2017
<b>VPP18</b>	VESTÍBULO	SOPORTE	PRIMAVERA	2018
<b>VPP19</b>	VESTÍBULO	SOPORTE	PRIMAVERA	2019
<b>VPV15</b>	VESTÍBULO	SOPORTE	VERANO	2015
<b>VPV16</b>	VESTÍBULO	SOPORTE	VERANO	2016
<b>VPV18</b>	VESTÍBULO	SOPORTE	VERANO	2018
<b>VPV19</b>	VESTÍBULO	SOPORTE	VERANO	2019
<b>VSI14</b>	VESTÍBULO	SUELO	INVIERNO	2014
<b>VSI15</b>	VESTÍBULO	SUELO	INVIERNO	2015
<b>VSI18</b>	VESTÍBULO	SUELO	INVIERNO	2018
<b>VSO15</b>	VESTÍBULO	SUELO	OTOÑO	2015
<b>VSO16</b>	VESTÍBULO	SUELO	OTOÑO	2016
<b>VSP16</b>	VESTÍBULO	SUELO	PRIMAVERA	2016
<b>VSP17</b>	VESTÍBULO	SUELO	PRIMAVERA	2017
<b>VSP18</b>	VESTÍBULO	SUELO	PRIMAVERA	2018
<b>VSP19</b>	VESTÍBULO	SUELO	PRIMAVERA	2019
<b>VSV15</b>	VESTÍBULO	SUELO	VERANO	2015
<b>VSV18</b>	VESTÍBULO	SUELO	VERANO	2018
<b>VSV19</b>	VESTÍBULO	SUELO	VERANO	2019

Tabla S3. Números de secuencias obtenidas para las muestras analizadas por la técnica de amplicones.

Colonia	Muestra	Crudas	Filtradas	No quiméricas
BEIS	B11M1	35714	21216	20396
	B11M2	367220	199316	192065
	B12M1	19102	15415	15074
	B12M2	32330	26950	26202
	B12M3	60225	50217	48882
	B12M4	173936	142010	136350
	B12M5	144610	122342	118743
	B12M6	7450	6176	6027
GRIS	G6M1	378967	36324	24593
	G8M1	260295	24365	13595
	G8M2	227298	24535	14039
	G9M1	325847	29788	24165
	G9M2	1160510	127847	110692
	G10M1	368773	246835	216781
	G10M2	283425	180971	159801
	G10M3	361863	236885	221522
	G10M4	240056	145543	131333
	G11M3	631788	409292	389515
	G11M4	417247	244818	228881
	G11M5	247797	132972	125549
	G11M6	205012	128216	122931
	G11M7	186956	94644	89223
	G12M1	209730	166971	158738
	G12M2	169986	134893	129014
	G12M3	173190	141956	133516
	G12M4	74947	61156	57926
BLANCA	W6M1	345716	23362	15299
	W8M1	232454	40239	22690
	W8M2	158096	25464	13966
	W9M1	483957	82803	69175
	W10M1	166768	124435	112129
	W10M2	97235	68777	61134
	W10M3	146724	99777	94340
	W10M4	177698	108737	100275
	W11M4	340935	220831	214895
	W11M5	248009	158253	154685
	W11M6	546257	362346	346453
	W11M7	287099	186960	180547
	W12M1	137534	114713	108512
	W12M2	75543	60651	57706
AMARILLA	Y6M1	291414	27866	18331
	Y8M1	260781	27628	17158
	Y8M2	241875	20512	12507
	Y9M1	1376991	162500	146962
	Y9M2	766569	77701	64568
	Y10M1	136722	69527	59650
	Y10M2	187942	91523	80960

	Y10M3	137590	60439	55462
	Y10M4	122376	59865	53289
	Y10M5	161535	82443	74058
	Y11M3	307509	179197	173091
	Y11M4	116531	66031	60927
	Y11M5	130418	71557	67217
	Y12M1	153893	125532	119463
	Y12M2	119478	96996	91264
CONTROL	SOP10M1	96624	60637	56194
	SOP10M2	124259	82915	79416
	<b>TOTAL</b>	<b>14940806</b>	<b>6191870</b>	<b>5697876</b>

Tabla S4. Taxonomía asignada a las OTUs que forman parte del core en las muestras de colonias visibles analizadas por la técnica de amplicones.

M.O: número de muestras en las que aparece esa OTU.

OTU ID	Taxonomía	Frecuencia	M.O
OTU1	<i>Bacteria; Actinobacteria; Nitriliruptoria; Euzebyales; Euzebyaceae</i>	1042377	50
OTU2	<i>Bacteria; Actinobacteria; Actinobacteria; Pseudonocardiales; Pseudonocardiaceae; Crossiella</i>	972004	54
OTU3	<i>Bacteria; Proteobacteria; Gammaproteobacteria; Nitrosococcales; Nitrosococcaceae; wb1-P19</i>	306256	57
OTU4	<i>Bacteria; Actinobacteria; Actinobacteria; Pseudonocardiales; Pseudonocardiaceae; Crossiella</i>	249867	36
OTU5	<i>Bacteria; Acidobacteria; Blastocatellia (Subgroup 4); 11-24</i>	104103	48
OTU6	<i>Bacteria; Dadabacteria; Dadabacteriia; Dadabacteriales</i>	90996	51
OTU7	<i>Bacteria; Acidobacteria; Blastocatellia (Subgroup 4); Blastocatellales; Blastocatellaceae; JGI 0001001-H03</i>	53933	24
OTU8	<i>Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Hyphomicrobiaceae</i>	53181	55
OTU9	<i>Bacteria; Proteobacteria; Gammaproteobacteria; PLTA13</i>	47932	37
OTU10	<i>Bacteria; Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; B1-7BS</i>	44215	55
OTU11	<i>Bacteria; Acidobacteria; Blastocatellia (Subgroup 4); 11-24</i>	43008	47
OTU12	<i>Bacteria; Proteobacteria; Gammaproteobacteria; PLTA13</i>	41976	52
OTU13	<i>Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Hyphomicrobiaceae; Hyphomicrobium</i>	40765	52
OTU14	<i>Bacteria; Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; B1-7BS</i>	38082	48
OTU15	<i>Bacteria; Rokubacteria; NC10; Rokubacteriales</i>	37092	53
OTU16	<i>Bacteria; Nitrospirae; Nitrospira; Nitrospirales; Nitrospiraceae; Nitrospira</i>	36796	57
OTU18	<i>Bacteria; Proteobacteria; Gammaproteobacteria; Betaproteobacteriales</i>	31106	46
OTU19	<i>Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Haliangiaceae; Haliangium</i>	30955	31
OTU20	<i>Bacteria; Chloroflexi; Anaerolineae; Ardenticatenales</i>	30493	49
OTU21	<i>Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Lysobacter</i>	30493	39

OTU24	<i>Bacteria; Proteobacteria; Gammaproteobacteria; Steroidobacterales; Steroidobacteraceae; Steroidobacter</i>	28471	47
OTU25	<i>Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiales Incertae Sedis</i>	25515	53
OTU26	<i>Bacteria; Proteobacteria; Gammaproteobacteria</i>	24979	49
OTU27	<i>Bacteria; Chloroflexi; Ktedonobacteria; Ktedonobacterales; Ktedonobacteraceae</i>	24461	12
OTU30	<i>Bacteria; Proteobacteria; Gammaproteobacteria; CCD24</i>	22150	10
OTU32	<i>Bacteria; Acidobacteria; Subgroup 6</i>	21445	34
OTU34	<i>Bacteria; Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; TRA3-20</i>	21359	50
OTU36	<i>Bacteria; Chloroflexi; Anaerolineae; Caldilineales; Caldilineaceae; Litorilinea</i>	19301	47
OTU37	<i>Bacteria; Dadabacteria; Dadabacteriia; Dadabacteriales</i>	18942	51
OTU41	<i>Bacteria; Acidobacteria; Subgroup 6</i>	16940	15
OTU47	<i>Bacteria; Proteobacteria; Alphaproteobacteria; Azospirillales; Inquilinaceae; Inquilinus; Inquilinus limosus</i>	16109	21
OTU50	<i>Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Arenimonas</i>	15016	29
OTU57	<i>Bacteria; Actinobacteria; Thermoleophilia; Solirubrobacterales; 67-14</i>	12827	27
OTU58	<i>Bacteria; Actinobacteria; Actinobacteria; Pseudonocardiales; Pseudonocardiaceae; Pseudonocardia</i>	12797	30
OTU66	<i>Bacteria; Proteobacteria; Deltaproteobacteria</i>	10969	44
OTU159	<i>Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Sphingomonas</i>	3853	34

Figura S1. Representación de los filos más abundantes (con una abundancia >1%) en las muestras de suelo y soporte.

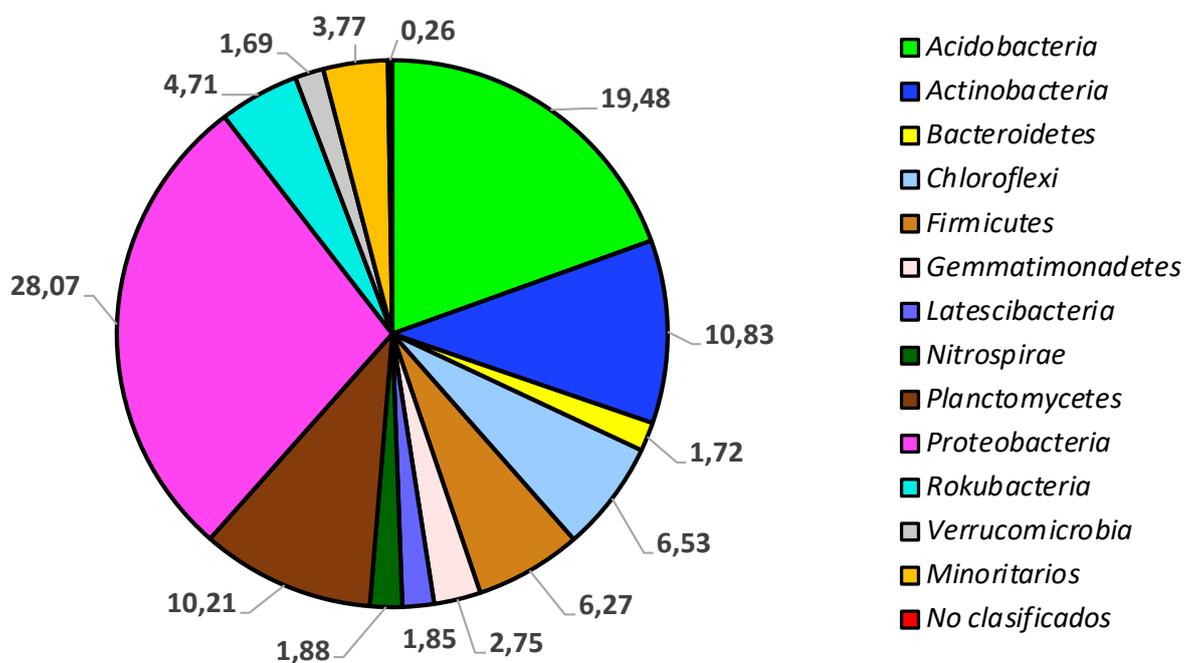


Figura S2. Representación de los parámetros indicadores de  $\alpha$ -diversidad en las muestras de suelos y soportes analizados por la técnica de amplicones. Se representa para cada zona de la cueva dónde se ha recogido muestra tanto de suelo como de soporte el nº de OTUs y los índices de Shannon y de Dominancia.

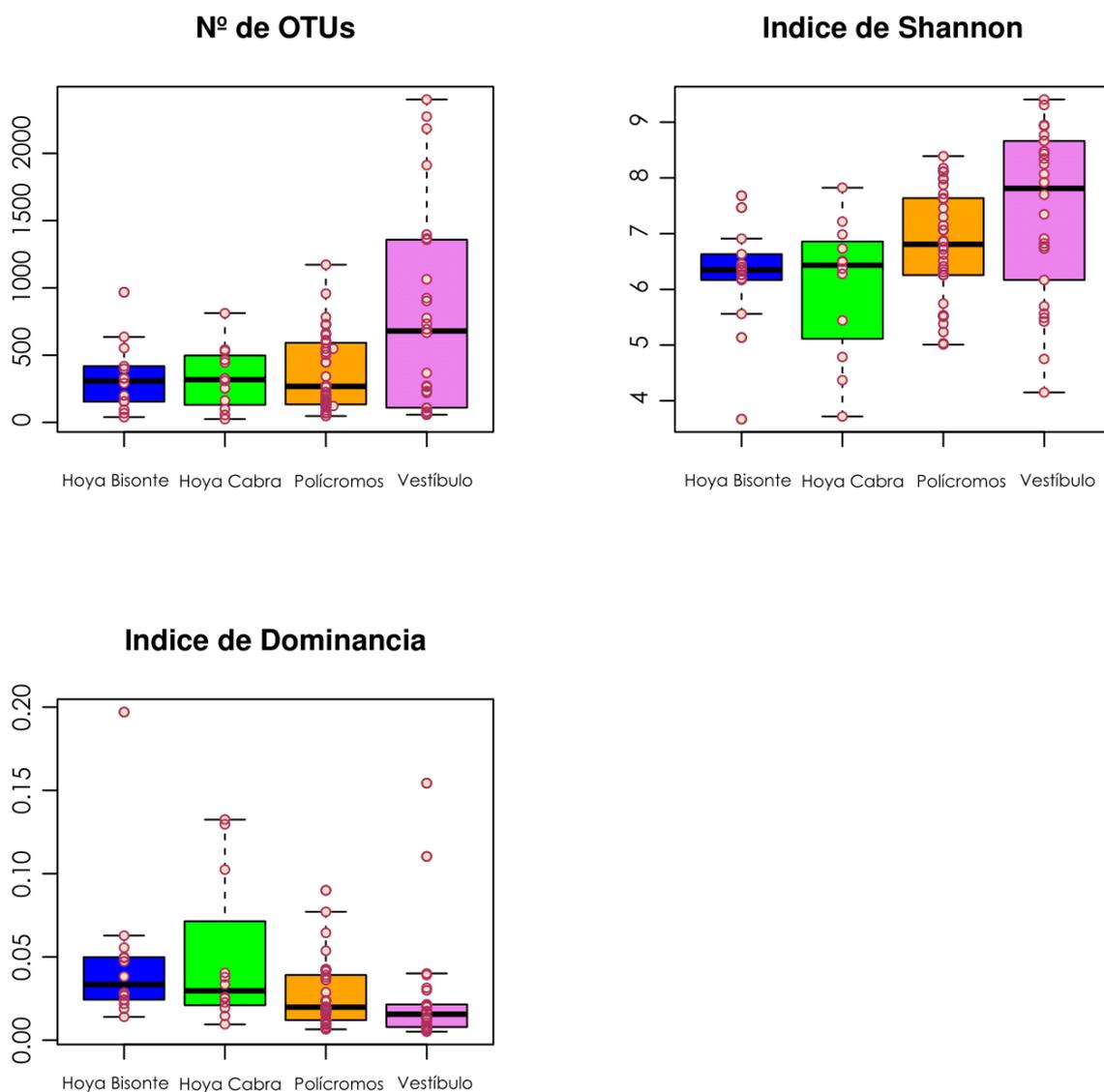


Figura S3. Correlación de los índices de  $\alpha$ -diversidad obtenidos por la técnica de amplicones (índice de Shannon) y shotgun (índice Nd) en las muestras de metagenomas.

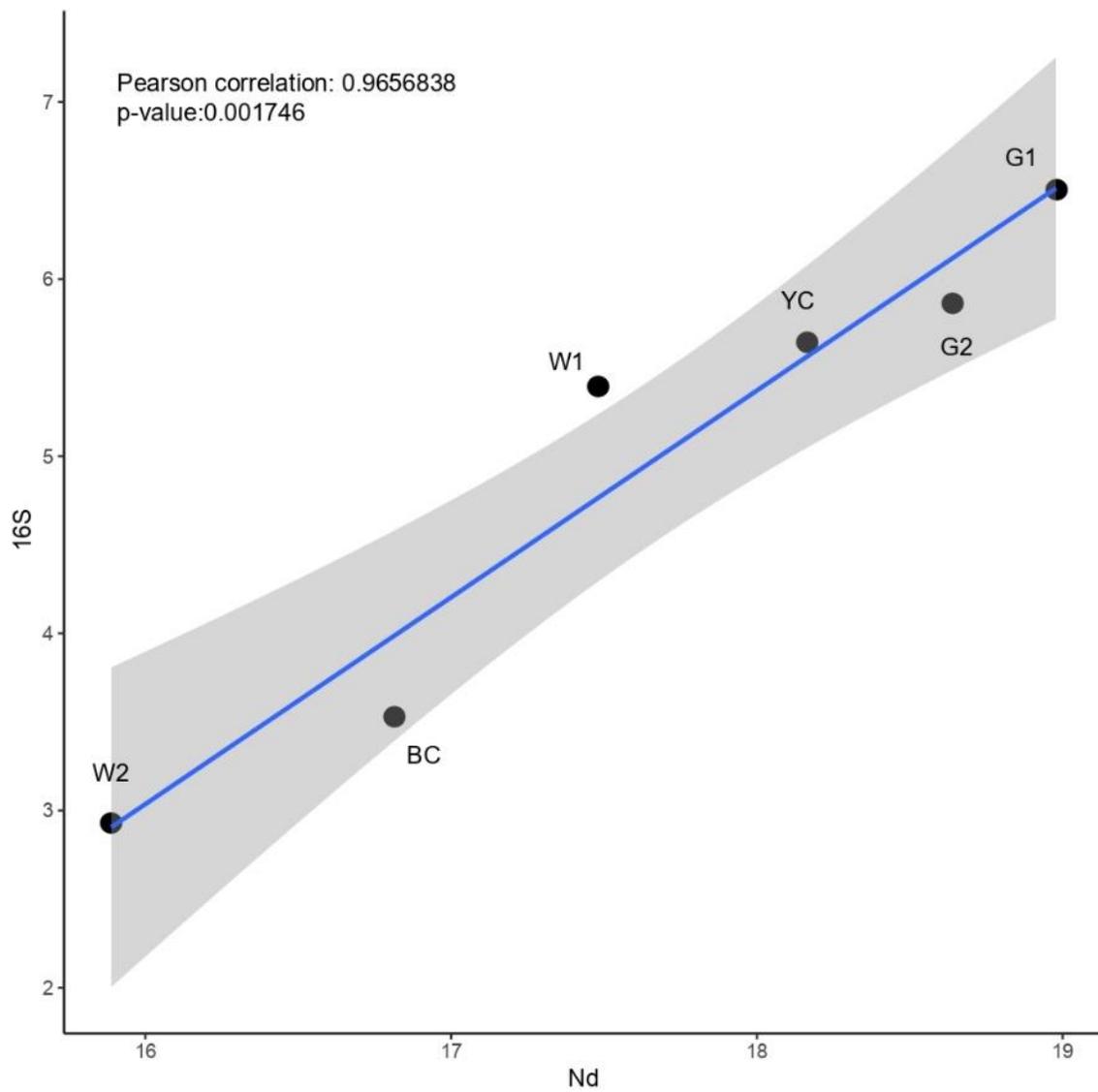


Tabla S5. Resumen de datos de los MAGs obtenidos de los metagenomas de las colonias. Para cada MAG se recoge su nivel de contaminación y completitud obtenidos con el programa CheckM, y su taxonomía asignada primero en base a la comparación de los ORFs contra la base de datos del NCBI y segundo en base a la asignación obtenida por comparación de las secuencias de los 16S rRNAs contra la base de datos de SILVA.

T.C: Tipo de Colonia; Cal: Calidad; M: Media; A: Alta; B: Baja; Compl: Completitud; Cont:Contaminación.

T.C	MAG	Cal	Compl	Cont	Tamaño del genoma	Contig más largo	Nº contigs	Genes	%GC	NCBI taxonomía	SILVA taxonomía	Tamaño 16S rRNA
G1	MAG001	M	100	12,9	7488565	131484	438	8121	0,67	<i>Bacteria;Actinobacteria;Actinobacteria;Pseudonocardiales;Pseudonocardiales bacterium</i>	<i>Bacteria;Actinobacteria;Actinobacteria;Pseudonocardiales;Pseudonocardiales bacterium</i>	1537
	MAG002	A	98,28	2,1	4346002	240073	316	4307	0,50	<i>Bacteria</i>	<i>Bacteria;Dadabacteria;Dadabacteria;Dadabacteriales;Candidatus Dadabacteria bacterium CSP1-2;Candidatus Dadabacteria bacterium CSP1-2</i>	1585
	MAG003	A	100	1,7	2420963	450802	63	2264	0,59	<i>Bacteria;Proteobacteria</i>	<i>Bacteria;Proteobacteria;Gamma proteobacteria;Betaproteobacteriales;Ferrovaceae;Ferrovum</i>	1546
	MAG004	A	94,83	0	11324678	126714	885	9767	0,59	<i>Bacteria;Acidobacteria</i>	<i>Bacteria;Acidobacteria;Blastocatellia(Subgroup4);11-24</i>	1529
	MAG005	M	71,47	6,9	2242966	38247	316	2370	0,70	<i>Bacteria;Proteobacteria;Betaproteobacteria</i>	-	-
	MAG006	M	78,42	0	918735	113411	73	1039	0,47	<i>Bacteria</i>	<i>Bacteria; Parcubacteria</i>	1522

## ANEXO

<b>MAG007</b>	M	51,41	7,7	2790673	49395	692	3348	0,68	<i>Bacteria;Proteobacteria;Betapr oteobacteria</i>	-	-
<b>MAG008</b>	M	80,88	1,2	3299398	41215	479	3489	0,66	<i>Bacteria;Proteobacteria;Betapr oteobacteria</i>	-	-
<b>MAG009</b>	A	96,55	1,9	7720411	84070	694	7283	0,65	<i>Bacteria;Acidobacteria;Acidob acteriabacterium</i>	<i>Bacteria;Acidobacteria;Subgrou p9</i>	368 / 1276
<b>MAG010</b>	M	70,87	5,6	1219028	28867	279	1676	0,34	<i>Archaea;Thaumarchaeota;Tha umarchaeota archaeon CSP1- 1</i>	-	-
<b>MAG011</b>	B	46,03	7,8	3020414	19937	1024	3650	0,71	<i>Bacteria;Proteobacteria</i>	-	-
<b>MAG012</b>	M	84,48	3,8	2683247	82040	422	2721	0,65	<i>Bacteria;Proteobacteria;Alpha proteobacteria;Rhizobiales;Hy phomicrobiaceae;Hyphomicrob ium</i>	<i>Bacteria;Proteobacteria;Alphapr oteobacteria;Rhizobiales;Hypho microbiaceae;Hyphomicrobium/ Bacteria;Proteobacteria;Alphapr oteobacteria;Micavibrionales</i>	1164 / 71
<b>MAG013</b>	M	79,31	0	3562342	64777	268	3659	0,54	<i>Bacteria;Proteobacteria;Deltap roteobacteria</i>	<i>Bacteria;Proteobacteria;Deltapro teobacteria;MBNT15/Bacteria;Pr oteobacteria;Deltaproteobacteria ;Myxococcales;bacteriap25</i>	840/ 1569
<b>MAG014</b>	M	77,98	3,3	5460614	25516	1814	6778	0,72	<i>Bacteria;Actinobacteria;Actino bacteria;Pseudonocardiales;Ps eudonocardiaceae;Pseudonoc ardia</i>	-	-
<b>MAG015</b>	M	81,03	6,2	2228533	37147	368	2396	0,64	<i>Bacteria;Proteobacteria</i>	-	-
<b>MAG016</b>	M	82,29	9,8	4956333	53384	1113	5471	0,66	<i>Bacteria;Proteobacteria</i>	<i>Bacteria;Proteobacteria;Gamma proteobacteria;PLTA13</i>	1573
<b>MAG017</b>	M	75,32	3,5	1200372	11017	554	1589	0,64	<i>Bacteria;Candidate division NC10;Candidate division NC10 bacterium CSP1-5</i>	<i>Bacteria;Rokubacteria;NC10;Me thylomirabiales;Methylomirabila ceae;wb1-A12;Candidate</i>	208

											división NC10 bacterium CSP1-5	
<b>MAG018</b>	M	74,92	4,6	5444691	48659	1736	6549	0,65	<i>Bacteria;Actinobacteria;Actinobacteria;Micromonosporales;Micromonosporaceae</i>	<i>Bacteria;Actinobacteria;Actinobacteria;Micromonosporales;Micromonosporaceae</i>	1459	
<b>MAG019</b>	B	36,42	9,9	1348858	14797	711	1778	0,65	<i>Bacteria;Proteobacteria;Alpha proteobacteria;Rhizobiales;Hyphomicrobiaceae;Hyphomicrobium</i>	-	-	
<b>MAG020</b>	B	47,19	6,1	1874449	17834	1078	2549	0,58	<i>Bacteria;Nitrospirae</i>	-	-	
<b>MAG021</b>	M	84,33	6	4439330	24209	1149	4900	0,58	<i>Bacteria;Chloroflexi</i>	<i>Bacteria;Chloroflexi;Anaerolineae;SBR1031;A4b</i>	1429	
<b>MAG022</b>	M	58,56	5,5	2489691	16671	967	3136	0,45	<i>Bacteria</i>	-	-	
<b>MAG023</b>	B	37,21	6,7	2248755	17324	1250	3116	0,62	<i>Bacteria;Proteobacteria</i>	-	-	
<b>MAG024</b>	B	34,31	7,5	3123590	29230	1605	4417	0,63	<i>Bacteria;Proteobacteria</i>	<i>Bacteria;Proteobacteria;Gamma proteobacteria;Betaproteobacteriales;B1-7BS</i>	407	
<b>MAG025</b>	B	49,75	0	2807380	11772	1157	3452	0,53	<i>Bacteria;Chloroflexi</i>	-	-	
<b>MAG026</b>	M	50,92	1,7	2989753	10472	1399	3688	0,71	<i>Bacteria;Acidobacteria;Acidobacteriabacterium</i>	<i>Bacteria;Acidobacteria;Subgroup17</i>	384	
<b>MAG027</b>	B	36,05	9,3	2215502	9891	1153	3001	0,64	<i>Bacteria;Proteobacteria;Alpha proteobacteria</i>	-	-	
<b>MAG028</b>	B	36,44	5,2	2207395	9890	1078	2835	0,68	<i>Bacteria;Proteobacteria;Alpha proteobacteria;Rhodospirillales</i>	-	-	



<b>MAG050</b>	M	63,95	6,9	4471934	48012	632	4930	0,69	<i>Bacteria;Actinobacteria</i>	<i>Bacteria;Actinobacteria;Nitriliruptoria;Euzebyales;Euzebyaceae</i>	1458
<b>MAG051</b>	M	74,37	6,9	3195188	155875	134	2846	0,65	<i>Bacteria;Proteobacteria;Gamma proteobacteria;Xanthomonadales</i>	-	-
<b>MAG052</b>	M	98,28	5,6	3267255	160999	214	3155	0,47	<i>Bacteria</i>	<i>Bacteria;Dadabacteria;Dadabacteriia;Dadabacteriales;Candidatus Dadabacteria bacterium CSP1-2;Candidatus Dadabacteria bacterium CSP1-2;Candidatus Dadabacteria bacterium CSP1-2</i>	1585
<b>MAG053</b>	M	76,57	9,5	5134356	173801	297	4951	0,68	<i>Bacteria;Proteobacteria;Deltaproteobacteria</i>	-	-
<b>MAG054</b>	A	92,01	0	3440239	143633	199	3453	0,65	<i>Bacteria;Proteobacteria;Gamma proteobacteria</i>	<i>Bacteria;Proteobacteria;Gamma proteobacteria;CCD24</i>	327 / 1262
<b>MAG055</b>	M	94,51	7,8	6022827	101991	779	6246	0,70	<i>Bacteria;Proteobacteria</i>	<i>Bacteria;Proteobacteria;Alphaproteobacteria;Azospirillales;Inquillinae;Inquillinus;Inquillinus limosus DSM16000</i>	1496
<b>MAG056</b>	M	65,52	7,8	1839710	12052	884	2407	0,63	<i>Bacteria;Nitrospirae;Nitrospira</i>	-	-
<b>MAG057</b>	M	90,91	6,1	4689839	75397	495	5015	0,62	<i>Bacteria;Proteobacteria;Alpha proteobacteria;Rhizobiales</i>	<i>Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiales Incertae Sedis;Nordella</i>	836
<b>MAG058</b>	M	59,53	3,5	2901216	42914	610	2928	0,68	<i>Bacteria;Proteobacteria;Gamma proteobacteria;Xanthomonadales;Xanthomonadaceae</i>	<i>Bacteria;Proteobacteria;Gamma proteobacteria;Xanthomonadales;Xanthomonadaceae;Lysobacter;Lysobacter koreensis</i>	1556
<b>MAG059</b>	M	61,58	3,8	2548397	70671	916	3287	0,61	<i>Bacteria;Proteobacteria;Alpha proteobacteria;Rhizobiales</i>	-	-

## ANEXO

<b>MAG060</b>	M	83,04	6,4	5954220	76294	1348	6939	0,68	<i>Bacteria;Actinobacteria;Actinobacteria;Pseudonocardiales;Pseudonocardiales bacterium</i>	-	-
<b>MAG061</b>	M	61,25	2,0	4218948	12065	1816	4685	0,60	<i>Bacteria;Acidobacteria</i>	<i>Bacteria;Acidobacteria;Blastocatellia (Subgroup4);11-24</i>	1492
<b>MAG062</b>	A	93,68	4,7	6888699	47141	2214	8593	0,57	<i>Bacteria;Nitrospirae</i>	<i>Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae;Nitrospira</i>	837 / 675
<b>MAG063</b>	B	43,18	3,5	4005047	57864	1914	5446	0,62	<i>Bacteria;Proteobacteria</i>	-	-
<b>MAG064</b>	M	59,44	0	5498228	18979	1747	5658	0,56	<i>Bacteria;Chloroflexi</i>	-	-
<b>MAG065</b>	B	24,29	7,9	3383476	23221	1695	4733	0,58	<i>Bacteria;Proteobacteria;Alpha proteobacteria</i>	-	-
<b>MAG066</b>	M	61,41	6,1	4065178	17072	1159	4489	0,68	<i>Bacteria;Proteobacteria;Deltaproteobacteria</i>	<i>Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Haliangiaceae;Haliangium</i>	137
<b>MAG067</b>	B	34,48	8,6	2150610	16309	952	2620	0,71	<i>Bacteria;Proteobacteria;Alpha proteobacteria</i>	-	-
<b>MAG068</b>	B	29,89	1,7	2676600	6224	1660	3807	0,71	<i>Bacteria;Candidatus Rokubacteria</i>	-	-
<b>MAG069</b>	M	51,69	5,6	4703340	15844	1977	5849	0,51	<i>Bacteria;Chloroflexi;Chloroflexi bacterium</i>	<i>Bacteria;Chloroflexi;Anaerolineae;Ardenticatenales</i>	728 / 1502
<b>MAG070</b>	M	57,16	7,8	2349258	9498	1239	3244	0,66	<i>Bacteria;Proteobacteria</i>	-	-
<b>MAG071</b>	B	42,15	6,9	1954064	11828	1158	2770	0,45	<i>Archaea/ Bacteria</i>	<i>Archaea;Thaumarchaeota;Nitrososphaeria;Nitrosotaleales;Nitrosotaleaceae</i>	1467



	<b>MAG081</b>	B	41,84	3,3	3358834	13062	2018	4426	0,60	<i>Bacteria;Acidobacteria</i>	<i>Bacteria;Acidobacteria;Blastocatellia(Subgroup4);11-24</i>	779
<b>W2</b>	<b>MAG082</b>	M	96,55	6,9	5463669	126841	268	5477	0,69	<i>Bacteria;Actinobacteria;Actinobacteria</i>	<i>Bacteria;Actinobacteria;Nitriliruptoria;Euzebyales;Euzebyaceae</i>	1531
	<b>MAG083</b>	A	92,95	0,2	3204176	76861	275	3317	0,60	<i>Bacteria;Proteobacteria;Gammaaproteobacteria;Chromatiales;Chromatiales bacterium USCg_Taylor</i>	<i>Bacteria;Proteobacteria;Gammaaproteobacteria;Nitrosococcales;Nitrosococcaceae;wb1-P19</i>	1135
	<b>MAG084</b>	M	62,98	2,6	5231336	28337	1338	6241	0,73	<i>Bacteria;Actinobacteria;Thermoleophilia;Solirubrobacterales</i>	-	-
	<b>MAG085</b>	M	76,54	3,5	1420772	49482	357	1824	0,50	<i>Bacteria</i>	<i>Bacteria; Parcubacteria</i>	1522
	<b>MAG086</b>	M	76,18	3,4	3517794	21168	1350	4424	0,68	<i>Bacteria;Proteobacteria</i>	-	-
	<b>MAG087</b>	M	55,02	1,4	5461098	59338	1503	5883	0,58	<i>Bacteria;Acidobacteria</i>	-	-
<b>YC</b>	<b>MAG088</b>	M	100	10,3	6179751	256681	172	6599	0,68	<i>Bacteria;Actinobacteria;Actinobacteria;Pseudonocardiales;Pseudonocardiales bacterium</i>	<i>Bacteria;Actinobacteria;Actinobacteria;Pseudonocardiales;Pseudonocardiaaceae;Crossiella</i>	1528
	<b>MAG089</b>	A	91,85	4,5	4113265	51226	856	4723	0,60	<i>Bacteria;Proteobacteria;Gammaaproteobacteria;Chromatiales;Chromatiales bacterium USCg_Taylor</i>	<i>Bacteria;Proteobacteria;Gammaaproteobacteria;Nitrosococcales;Nitrosococcaceae;wb1-P19</i>	1543
	<b>MAG090</b>	M	76,49	10,3	3793170	34399	806	4379	0,69	<i>Bacteria;Actinobacteria;Actinobacteria</i>	<i>Bacteria;Actinobacteria;Nitriliruptoria;Euzebyales;Euzebyaceae</i>	1530
	<b>MAG091</b>	M	79,53	6,9	3614549	26977	738	3894	0,66	<i>Bacteria;Proteobacteria</i>	<i>Bacteria;Proteobacteria;Gammaaproteobacteria;PLTA13</i>	275

	<b>MAG092</b>	B	44,81	4,2	2417819	15651	756	3143	0,65	<i>Bacteria;Actinobacteria;Actinobacteria;Corynebacteriales;Mycobacteriaceae</i>	<i>Bacteria;Actinobacteria;Actinobacteria;Corynebacteriales;Mycobacteriaceae;Mycobacterium;Mycobacterium noviomagense</i>	141
	<b>MAG093</b>	M	57,99	2,2	3579880	23930	1798	4553	0,63	<i>Bacteria;Acidobacteria</i>	<i>Bacteria;Acidobacteria;Subgroup 6</i>	998
	<b>MAG094</b>	B	49,03	4,0	2073275	10289	1010	2750	0,69	<i>Bacteria;Proteobacteria;Alpha proteobacteria;Rhizobiales</i>	-	-
	<b>MAG095</b>	B	46,32	1,7	2463887	8806	1550	3647	0,70	<i>Bacteria;Actinobacteria;Actinobacteria</i>	-	-
	<b>MAG096</b>	B	35,52	6,6	2127477	11648	1386	3167	0,39	<i>Bacteria;Bacteroidetes</i>	-	-
<b>BC</b>	<b>MAG097</b>	M	96,55	6,9	5316480	59461	689	5568	0,69	<i>Bacteria;Actinobacteria;Actinobacteria</i>	<i>Bacteria;Actinobacteria;Nitrilitoria;Euzebyales;Euzebyaceae</i>	1219 / 408
	<b>MAG098</b>	A	98,28	2,6	5780955	172933	267	5977	0,67	<i>Bacteria;Actinobacteria;Actinobacteria;Pseudonocardiales;Pseudonocardiales bacterium</i>	<i>Bacteria;Actinobacteria;Actinobacteria;Pseudonocardiales;Pseudonocardiales bacterium</i>	1219 / 406
	<b>MAG099</b>	A	100	0	4345056	165655	292	4239	0,63	<i>Bacteria;Proteobacteria</i>	<i>Bacteria;Proteobacteria;Gamma proteobacteria;PLTA13</i>	1566
	<b>MAG100</b>	A	98,12	0	2257677	37922	322	2462	0,68	<i>Bacteria;Actinobacteria</i>	<i>Bacteria;Actinobacteria;Thermoplasma;Solirubrobacterales;67-14</i>	1155
	<b>MAG101</b>	M	67,89	4,3	2898466	13277	1248	3424	0,67	<i>Bacteria;Proteobacteria;Gamma proteobacteria;Xanthomonadales;Xanthomonadaceae</i>	<i>Bacteria;Proteobacteria;Gamma proteobacteria;Xanthomonadales;Xanthomonadaceae;Lysobacter;Lysobacter koreensis</i>	1556

Tabla S6. Clusters de MAGs en base a su similitud (AAI y ANI).

T.C: Tipo de Colonia; Cal: Calidad; M: Media; A: Alta; B: Baja.

Taxonomía de los MAGs	Cluster	Taxonomía del cluster	MAG	T.C	Calidad	AAI	ANI	
<i>Thaumarchaeota</i>	Cluster I	<i>Thaumarchaeota</i>	MAG010	G1	M	95,64	96,51	
			MAG074	W1	M			
<i>Acidobacteria</i>	Cluster II	<i>Acidobacteria</i>	MAG081	W1	B	>90,09	>94,49	
			MAG004	G1	A			
			MAG061	G2	M			
				MAG093	YC	M		
				MAG087	W2	M		
				MAG009	G1	A		
				MAG032	G2	B		
				MAG026	G1	M		
<i>Actinobacteria</i>	Cluster III	<i>Pseudonocardiaaceae</i>	MAG088	YC	M	>71,66	>77,10	
			MAG046	G2	A			
			MAG098	BC	A			
			MAG060	G2	M			
			MAG001	G1	M			
	Cluster IV	<i>Nitriliruptoria</i>	MAG097	BC	M	>82,25	>85,85	
			MAG090	YC	M			
			MAG050	G2	M			
			MAG082	W2	M			
				MAG014	G1	M		
				MAG048	G2	A		
				MAG084	W2	M		
				MAG095	YC	B		
				MAG092	YC	B		
			MAG100	BC	A			
			MAG018	G1	M			
<i>Proteobacteria</i>				MAG023	G1	B		
				MAG070	G2	M		
				MAG034	G1	B		
				MAG063	G2	B		
				MAG015	G1	M		
				MAG011	G1	B		
<i>Alphaproteobacteria</i>	Cluster V	<i>Rhizobiales</i>	MAG094	YC	B	93,21	95,79	
			MAG067	G2	B			
				MAG065	G2	B		
				MAG059	G2	M		
				MAG057	G2	M		
				MAG019	G1	B		

## ANEXO

			MAG012	G1	M			
			MAG028	G1	B			
			MAG055	G2	M			
			MAG076	W1	M			
			MAG027	G1	B			
<i>Betaproteobacteria</i>	Cluster VI	<i>Ferrovaceae</i>	MAG049	G2	A	99,84	99,81	
			MAG003	G1	A			
				MAG007	G1	M		
				MAG024	G1	B		
	Cluster VII	<i>Betaproteobacteria</i>	MAG086	W2	M	72,33	78,99	
			MAG005	G1	M			
				MAG030	G1	B		
			MAG029	G1	B			
			MAG008	G1	M			
<i>Gammaproteobacteria</i>	Cluster VIII	<i>Lysobacter</i>	MAG101	BC	M	95,41	96,55	
			MAG058	G2	M			
				MAG051	G2	M		
	Cluster IX	<i>Chromatiales</i>	MAG047	G2	M	>86,91	>91,30	
			MAG089	YC	A			
			MAG083	W2	A			
				MAG078	W1	M		
				MAG054	G2	A		
				MAG016	G1	M		
				MAG091	YC	M		
			MAG099	BC	A			
<i>Deltaproteobacteria</i>			MAG013	G1	M			
			MAG053	G2	M			
			MAG066	G1	M			
			MAG080	W1	B			
<i>Nitrospirae</i>	Cluster X	<i>Nitrospira</i>	MAG077	W1	M	91,60	92,31	
			MAG062	G2	A			
				MAG020	G1	B		
			MAG056	G2	M			
<i>Candidate division NC10</i>	Cluster XI	<i>Candidate division NC10</i>	MAG017	G1	M	96,90	98,40	
			MAG075	W1	M			
<i>Candidatus Rokubacteria</i>			MAG068	G2	B			
			MAG033	G1	B			
<i>Parcubacteria</i>	Cluster XII	<i>Parcubacteria</i>	MAG006	G1	M	99,50	99,90	
			MAG085	W2	M			
<i>Candidatus Dadabacteria</i>	Cluster XIII	<i>Candidatus Dadabacteria</i>	MAG002	G1	A	>99,70	>99,80	
			MAG079	W1	A			

			MAG052	G2	M		
<i>Chloroflexi</i>			MAG021	G1	M		
			MAG069	G2	M		
			MAG064	G2	M		
			MAG025	G1	B		
<i>Bacteroidetes</i>			MAG096	YC	B		
No clasificados			MAG072	G2	B		
			MAG022	G1	M		
Bines contaminados			MAG031	G1	B		
			MAG071	G2	B		

Tabla S7. Genomas de referencia utilizados para la comparación a nivel de de los MAGs agrupados o no en clusters y sus números de acceso en GenBank.

Tipo de cluster	Taxonomía de los genomas de referencia	Ensamblados de los genomas de referencia	Número de acceso en GenBank
Cluster I	<i>Nitrosopumilaceae</i> ( <i>Thaumarchaeota</i> )	<i>Nitrosarchaeum</i> sp. AC2	GCA_013407385.1
		<i>Nitrosarchaeum koreense</i> MY1	GCA_000220175.2
		<i>Candidatus Nitrosomarinus</i> sp.	GCA_003702465.1
		<i>Candidatus Nitrosomarinus catalina</i>	GCA_002156965.1
		<i>Nitrosopumilus maritimus</i> SCM1	GCA_000018465.1
		<i>Thaumarchaeota archaeon</i> CSP1-1	GCA_001443365.1
		<i>Candidatus Nitrosopumilus</i> sp. SW	GCA_006740685.1
		<i>Candidatus Nitrosopumilus sediminis</i>	GCA_000299395.1
Cluster II	<i>Blastocatellia</i> ( <i>Acidobacteria</i> )	<i>Chloracidobacterium</i> sp. UBA2026	GCA_002333435.1
		<i>Chloracidobacterium thermophilum</i> B	GCA_000226295.1
		<i>Pyrinomonadaceae</i> bacterium	GCA_013816505.1
		<i>Pyrinomonas methylaliphatogenes</i>	GCA_000820845.2
		<i>Blastocatellia bacterium</i> AA13	GCA_003222245.1
		<i>Blastocatellia bacterium</i>	GCA_011058595.1
Cluster III	<i>Pseudonocardiaceae</i> ( <i>Actinobacteria</i> )	<i>Pseudonocardiales</i> bacterium	GCA_003244055.1
		<i>Pseudonocardiales</i> bacterium	GCA_003244245.1
		<i>Pseudonocardiales</i> bacterium	GCA_013696175.1
		<i>Pseudonocardiales</i> bacterium	GCA_003244235.1
Cluster VI	<i>Ferrovaceae</i> ( <i>Betaproteobacteria</i> )	<i>Ferrovum</i> sp. 34-44-207	GCA_002279055.1
		<i>Ferrovum myxofaciens</i>	GCA_013181395.1
		<i>Ferrovum</i> sp. JA12	GCA_001431705.1
		<i>Ferrovum</i> sp. 37-45-19	GCA_002279305.1
		<i>Ferrovum</i> sp. PN-J185	GCA_001581925.1
		<i>Ferrovum</i> sp. Z-31	GCA_001581915.1
		Uncultured <i>Ferrovum</i> sp Umea3p4_bin-1559	GCA_903826585.1
		Uncultured <i>Ferrovum</i> sp- 7_bin-604	GCA_903829055.1
Cluster VIII	<i>Lysobacter</i> ( <i>Gammaproteobacteria</i> )	<i>Lysobacter soli</i>	GCA_009755585.1
		<i>Lysobacter concretionis</i> Ko07 = DSM 16239	GCA_000768345.1
		<i>Lysobacter capsici</i>	GCA_002355295.1
		<i>Lysobacter antibioticus</i>	GCA_001442745.1
		<i>Lysobacter defluvii</i> IMMIB APB-9 = DSM 18482	GCA_000423325.1
		<i>Lysobacter spongicola</i> DSM 21749	GCA_900167055.1
		<i>Lysobacter alkalisoli</i>	GCA_006547045.1

Cluster IX	<i>Chromatiales</i> ( <i>Gammaproteobacteria</i> )	<i>Chromatiales bacterium</i> <i>USCg_Taylor</i>	GCA_002007425.1
Cluster X	<i>Nitrospira</i> ( <i>Nitrospirae</i> )	<i>Nitrospira</i> sp	GCA_005116745.1
		<i>Nitrospira defluvii</i>	GCA_000196815.1
		<i>Nitrospira moscoviensis</i>	GCA_001273775.1
		<i>Nitrospira japonica</i>	GCA_900169565.1
		<i>Candidatus Nitrospira nitrificans</i>	GCA_001458775.1
		<i>Candidatus Nitrospira nitrosa</i>	GCA_001458735.1
		<i>Candidatus Nitrospira inopinata</i>	GCA_001458695.1
		<i>Nitrospira lenta</i>	GCA_900403705.1
Cluster XI	Candidate division NC10	Candidate division NC10 bacterium CSP1-5	GCA_001443495.1
		Uncultured candidate division NC10 bacterium	GCA_903827505.1
		<i>Candidatus Methylospirillum oxyfera</i>	GCA_000091165.1
		Candidate division NC10 bacterium SPGG6	GCA_001643535.1
		Candidate division NC10 bacterium RIFCSPLOWO2_02_FULL_66_22	GCA_001771245.1
		Candidate division NC10 bacterium RIFCSPLOWO2_12_FULL_66_18	GCA_001771285.1
		Candidate division NC10 bacterium RBG_16_65_8	GCA_001772295.1
		Candidate division NC10 bacterium	GCA_003105305.1
Cluster XII	<i>Parcubacteria</i>	<i>Candidatus Campbellbacteria</i> bacterium GW2011_GWC1_35_31	GCA_000990745.1
		<i>Candidatus Campbellbacteria</i> bacterium RIFCSPHIGHO2_01_FULL_34_10	GCA_001774415.1
		<i>Candidatus Campbellbacteria</i> bacterium RIFCSPHIGHO2_12	GCA_001776215.1
		<i>Candidatus Campbellbacteria</i> bacterium RIFOXYC2_FULL_35_25	GCA_001774495.1
		<i>Candidatus Nomurabacteria</i> bacterium CG_4_9_14_0_2_um_filter_32_10	GCA_002788475.1
		<i>Candidatus Nomurabacteria</i> bacterium GW2011_GWE1_32_28	GCA_000989515.1
		<i>Candidatus Nomurabacteria</i> bacterium RIFOXYD1_FULL_42_14	GCA_001787275.1
		<i>Candidatus Nomurabacteria</i> bacterium UBA5155	GCA_002413705.1

Cluster XIII	<i>Candidatus</i> Dadabacteria	<i>Candidatus</i> Dadabacteria bacterium CSP1-2	GCA_001443445.1
		<i>Candidatus</i> Dadabacteria bacterium RBG_19FT_COMBO_40_33	GCA_001775245.1
		<i>Candidatus</i> Dadabacteria bacterium RIFCSPHIGHO2_12_FULL_53_21	GCA_001775255.1
		<i>Candidatus</i> Dadabacteria bacterium UBA1144	GCA_002311855.1
		<i>Candidatus</i> Dadabacteria bacterium UBA2774	GCA_002352485.1
		<i>Candidatus</i> Dadabacteria bacterium	GCA_014075295.1
No Cluster	<i>Inquilinus</i> (Alphaproteobacteria)	<i>Inquilinus</i> sp.	GCA_013003675.1
		<i>Inquilinus limosus</i>	GCA_002195995.1
		<i>Inquilinus</i> sp. KBS0705	GCA_005938025.2
		<i>Inquilinus ginsengisoli</i>	SRX6876147
		<i>Inquilinus limosus</i> DSM 16000	GCA_000423185.1
		<i>Inquilinus limosus</i> MP06	GCA_000767795.1

Tabla S8. Genomas de referencia utilizados para los análisis de estimación de distancias con los genomas de los componentes mayoritarios y su número de acceso en GenBank. En asterisco se representan sólo los géneros descritos en la base de datos del NCBI y no en SILVA.

Tipo de cluster	Taxonomía de los genomas de referencia	Ensamblados de los genomas de referencia	Número de acceso en GenBank
Cluster III- Pseudonocardiaceae	<i>Actinoalloteichus</i>	<i>Actinoalloteichus fjordicus</i>	GCA_001941625.1
	<i>Actinocrispum</i> *	<i>Actinocrispum wychmicini</i>	GCA_004345645.1
	<i>Actinokineospora</i>	<i>Actinokineospora auranticolor</i>	GCA_002934265.1
	<i>Actinomycetospora</i>	<i>Actinomycetospora succinea</i>	GCA_004363095.1
	<i>Actinophytocola</i>	<i>Actinophytocola oryzae</i>	GCA_004364325.1
	<i>Actinopolyspora</i>	<i>Actinopolyspora halophila</i>	GCA_000371785.1
	<i>Actinorectispora</i>	<i>no genome</i>	-
	<i>Actinosynnema</i>	<i>Actinosynnema mirum</i>	GCA_000023245.1
	<i>Alloactinosynnema</i>	<i>Alloactinosynnema sp. L-07</i>	GCA_900070365.1
	<i>Allokutzneria</i>	<i>Allokutzneria albata</i>	GCA_000720955.1
	<i>Amycolatopsis</i>	<i>Amycolatopsis methanolica</i>	GCA_000739085.1
	<i>Crossiella</i>	<i>Crossiella equi</i>	GCA_002155995.1
	<i>Gandjariella</i> *	<i>Gandjariella thermophila</i>	GCA_005405885.1
	<i>Goodfellowiella</i>	<i>Goodfellowiella sp</i>	GCA_008386585.1
	<i>Haloechothrix</i>	<i>Haloechothrix alba</i>	GCA_900188115.1
	<i>Haloechothrix</i>	<i>Haloechothrix alba</i>	GCA_900188115.1
	<i>Herbihabitans</i>	<i>Herbihabitans rhizosphaerae</i>	GCA_004216555.1
	<i>Kibdelosporangium</i>	<i>Kibdelosporangium phytohabitans</i>	GCA_001302585.1
	<i>Kutzneria</i>	<i>Kutzneria albida</i>	GCA_000525635.1
	<i>Labedaea</i>	<i>Labedaea rhizosphaerae</i>	GCA_004362825.1
	<i>Lechevalieria</i>	<i>Lechevalieria deserti</i>	GCA_003148865.1
	<i>Lentzea</i>	<i>Lentzea terrae</i>	GCA_003265345.1
	<i>Longimycelium</i>	<i>Longimycelium tulufanense</i>	GCA_014645795.1
<i>Prauserella</i>	<i>Prauserella muralis</i>	GCA_007993965.1	
<i>Pseudonocardia</i>	<i>Pseudonocardia dioxanivorans</i>	GCA_000196675.1	

## ANEXO

	<i>Pseudonocardiales</i>	<i>Pseudonocardiales bacterium</i>	GCA_003244255.1
	<i>Saccharomonospora</i>	<i>Saccharomonospora halophila</i>	GCA_000383775.1
	<i>Saccharopolyspora</i>	<i>Saccharopolyspora sp</i>	GCA_012277335.1
	<i>Saccharothrix</i>	<i>Saccharothrix syringae</i>	GCA_009498035.1
	<i>Sciscionella</i>	<i>Sciscionella marina</i>	GCA_000379465.1
	<i>Streptoalloteichus</i>	<i>Streptoalloteichus hindustanus</i>	GCA_900129375.1
	<i>Tamaricihabitans</i>	<i>Tamaricihabitans halophyticus</i>	GCA_004340875.1
	<i>Thermocrispum</i>	<i>Thermocrispum agreste</i>	GCA_000427905.1
	<i>ThermOTUnica</i>	<i>no genome</i>	-
	<i>Umezawaea</i>	<i>Umezawaea tangerina</i>	GCA_003002815.1
	<i>Yuhushiella</i>	<i>Yuhushiella deserti</i>	GCA_004365925.1
<i>Cluster IV- Nitriliruptoria</i>	<i>Egicoccales</i>	<i>Egicoccus halophilus</i>	GCA_004300825.1
	<i>Nitriliruptorales</i>	<i>Nitriliruptor sp</i>	GCA_007693285.1
		<i>Nitriliruptor alkaliphilus</i>	GCA_000969705.1
		<i>Nitriliruptoraceae bacterium ZYF776</i>	GCA_009720505.1
		<i>Nitriliruptorales bacterium</i>	GCA_007119035.1
		<i>Nitriliruptoraceae bacterium</i>	GCA_003565155.1
	<i>Euzebyales</i>	<i>Euzebya sp</i>	GCA_003344865.1
		<i>Euzebyaceae bacterium</i>	GCA_013812395.1
		<i>Euzebyales bacterium</i>	GCA_013817125.1
		<i>Euzebya rosea</i>	GCA_003073135.1
		<i>Euzebya tangerina</i>	GCA_003074135.1
	<i>Egibacterales</i>	<i>Egibacter sp</i>	GCA_003553795.1
		<i>Egibacter rhizosphaerae</i>	GCA_004322855.1

Tabla S9. Lista de primers y su secuencia diseñados para identificar los componentes mayoritarios de las colonias.

Primer	Secuencia	Tipo de colonia	Tipo de cluster	
MAG001F	GTTCTATAGCGTGGCGAGTCAT	G1	III <i>Pseudonocardia</i>	
MAG001R	CTTGGAGTAGCGAGGCAATAC			
MAG046F	GGCCGACAACCTGTACGA	G2		
MAG046R	CTCACCTCGCTTAGGTCCTG			
MAG073F	GTCGCCTTAGCCTGGTTCTAC	W1		
MAG073R	ACCAAAAACAGTGGGACACAG			
MAG088F	TACGACGTGTACTGTGCGAAG	YC		
MAG088R	ACAGACAATCGTAGCTGGACAA			
MAG098F	GGGTTATTGTGTTTGTTCGATCTT	BC		IV <i>Nitriliruptoria</i>
MAG098R	CATTAATGTCCTGCGGGTCT			
MAG097F	ATACGGGTCGTAGGTGGTGA			
MAG097R	CTACTCGCCCTCGAGCTATG			
MAG082F	CAGAGCCAGACAGAAATGATCT	W2		
MAG082R	GTAGGGATTTCGCCGTTATGAG			

Tabla S10. Lista de los aislados cultivables de las colonias visibles.  
T.C: Tipo de colonia; C.C: Condiciones de Crecimiento; H: Humedad.

ID	T.C	Sala cueva	Medio aislado	C.C	Fecha de siembra	Tamaño	Identidad	Taxonomía
VC1	YC	Vestibulo	TSA	28°C	13/9/16	954	99,58	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus simplex</i>
VC2	YC	Vestibulo	TSA	28°C	13/9/16	1471	99,86	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus licheniformis</i>
VC3	YC	Vestibulo	TSA	28°C	13/9/16	1442	99,09	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces spiroverticillatus</i>
VC4	YC	Vestibulo	TSA	28°C	13/9/16	1471	99,86	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus licheniformis</i>
VC5	YC	Vestibulo	TSA	28°C	13/9/16	1467	99,93	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus subtilis</i>
VC6	YC	Vestibulo	TSA	28°C	13/9/16	1473	99,86	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus mycoides</i>
VC7	YC	Vestibulo	TSA	28°C	13/9/16	1473	99,00	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus mobilis</i>
VC8	YC	Vestibulo	TSA	28°C	13/9/16	952	99,89	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus licheniformis</i>
VC9	WC	Cruce	TSA	28°C	13/9/16	1477	99,38	<i>Firmicutes; Bacilli; Lactobacillales; Aerococcaceae; Facklamia; Facklamia tabacinasalis</i>
VC10	WC	Cruce	TSA	28°C	13/9/16	1446	99,58	<i>Actinobacteria; Actinobacteria; Micrococcales; Micrococccaceae; Micrococcus; Micrococcus luteus</i>

<b>VC11</b>	WC	Cruce	TSA	28°C	13/9/16	1441	99,58	<i>Actinobacteria; Actinobacteria; Micrococcales; Micrococaceae; Micrococcus; Micrococcus luteus</i>
<b>VC13</b>	WC	Cruce	TSA	28°C	13/9/16	1440	99,79	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC14</b>	WC	Cruce	TSA	28°C	13/9/16	1473	99,86	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus mycoides</i>
<b>VC15</b>	WC	Cruce	TSA	28°C	13/9/16	1447	99,79	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC17</b>	WC	Cruce	TSA	28°C	13/9/16	1440	99,58	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC18</b>	GC	Cruce	TSA	28°C	13/9/16	1448	99,65	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC19</b>	GC	Cruce	TSA	28°C	13/9/16	1447	99,79	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC20</b>	GC	Cruce	TSA	28°C	13/9/16	1448	99,65	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC22</b>	GC	Cruce	TSA	28°C	13/9/16	1447	99,58	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC25</b>	YC	Vestibulo	TSA	28°C	13/9/16	1441	99,65	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC26</b>	YC	Vestibulo	TSA	28°C	13/9/16	948	99,89	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus licheniformis</i>
<b>VC27</b>	WC	Cruce	TSA	28°C	13/9/16	948	99,79	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus sp</i>

<b>VC28</b>	WC	Cruce	TSA	28°C	13/9/16	1475	99,59	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Psychrobacillus; Psychrobacillus lasiicapitis</i>
<b>VC29</b>	WC	Cruce	TSA	28°C	13/9/16	1439	99,86	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC30</b>	WC	Cruce	TSA	28°C	13/9/16	1426	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC31</b>	WC	Cruce	TSA	28°C	13/9/16	1441	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces exfoliatus</i>
<b>VC32</b>	WC	Cruce	TSA	28°C	13/9/16	1442	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC33</b>	WC	Cruce	TSA	28°C	13/9/16	1438	100,00	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Rhodococcus; Rhodococcus wratislaviensis</i>
<b>VC34</b>	GC	Cruce	TSA	28°C	13/9/16	1441	99,65	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC35</b>	GC	Cruce	TSA	28°C	13/9/16	1467	99,58	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus licheniformis</i>
<b>VC36</b>	GC	Vestibulo	TSA	28°C	13/9/16	1475	98,96	<i>Firmicutes; Bacilli; Bacillales; Planococcaceae; Sporosarcina; Sporosarcina globispora</i>
<b>VC37</b>	GC	Vestibulo	TSA	28°C	13/9/16	1431	99,00	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia alba</i>
<b>VC38</b>	GC	Vestibulo	TSA	28°C	13/9/16	1441	99,58	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC39</b>	YC	Vestibulo	TSA	28°C	13/9/16	1442	99,65	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>

<b>VC40</b>	GC	Vestibulo	TSA	28°C	13/9/16	1471	99,86	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus idriensis</i>
<b>VC41</b>	GC	Cruce	TSA 0,1X	28°C	27/9/16	1440	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC42</b>	GC	Cruce	TSA 0,1X	28°C	27/9/16	1472	100,00	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus mycoides</i>
<b>VC43</b>	GC	Cruce	TSA 0,1X	28°C	27/9/16	1435	100,00	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Rhodococcus; Rhodococcus wratislaviensis</i>
<b>VC44</b>	GC	Cruce	TSA 0,1X	28°C	27/9/16	1444	99,58	<i>Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; Micrococcus; Micrococcus luteus</i>
<b>VC45</b>	GC	Cruce	TSA 0,1X	28°C	27/9/16	1433	99,93	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia asteroides</i>
<b>VC46</b>	GC	Cruce	TSA 0,1X	28°C	27/9/16	1471	99,73	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus zhangzhouensis</i>
<b>VC47</b>	YC	Vestibulo	AIA	28°C	31/1/17	1442	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces exfoliatus</i>
<b>VC49</b>	YC	Vestibulo	AIA	28°C	31/1/17	1433	99,79	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia asteroides</i>
<b>VC50</b>	YC	Vestibulo	AIA	28°C	31/1/17	1443	99,93	<i>Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; Kocuria; Kocuria palustris</i>
<b>VC51</b>	YC	Vestibulo	AIA	28°C	31/1/17	1432	98,99	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia alba</i>
<b>VC52</b>	WC	Vestibulo	AIA	28°C	31/1/17	1448	99,79	<i>Actinobacteria; Actinobacteria; Micrococcales; Microbacteriaceae; Agromyces; Agromyces fucosus</i>

<b>VC53</b>	WC	Vestibulo	AIA	28°C	31/1/17	1472	99,94	<i>Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus mycoides</i>
<b>VC54</b>	WC	Vestibulo	AIA	28°C	31/1/17	1273	100,00	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia asteroides</i>
<b>VC55</b>	WC	Cruce	AIA	28°C	31/1/17	1434	100,00	<i>Actinobacteria; Actinobacteria; Pseudonocardiales; Pseudonocardaceae; Crossiella; Crossiella cryophila</i>
<b>VC56</b>	WC	Cruce	AIA	28°C	31/1/17	1450	98,96	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces spiroverticillatus</i>
<b>VC57</b>	WC	Cruce	AIA	28°C	31/1/17	1433	99,00	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia asteroides</i>
<b>VC58</b>	WC	Cruce	AIA	28°C	31/1/17	1442	99,79	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces exfoliatus</i>
<b>VC59</b>	WC	Cruce	AIA	28°C	31/1/17	1433	99,86	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia asteroides</i>
<b>VC60</b>	WC	Vestibulo	AIA	28°C	31/1/17	1442	100,00	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC61</b>	WC	Cruce	AIA	28°C	31/1/17	1515	99,00	<i>Actinobacteria; Actinobacteria; Pseudonocardiales; Pseudonocardaceae; Crossiella; Crossiella cryophila</i>
<b>VC62</b>	WC	Cruce	AIA	28°C	31/1/17	1402	99,05	<i>Proteobacteria; Alphaproteobacteria; Rhodospirillales; Rhodospirillaceae; Inquilinus; Inquilinus ginsengisoli</i>
<b>VC63</b>	WC	Cruce	AIA	28°C	31/1/17	1454	99,72	<i>Actinobacteria; Actinobacteria; Micrococcales; Microbacteriaceae; Agromyces; Agromyces fucosus</i>
<b>VC64</b>	WC	Cruce	AIA	28°C	31/1/17	1442	99,00	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces mauvecolor</i>

<b>VC65</b>	WC	Cruce	AIA	28°C	31/1/17	1438	99,79	<i>Actinobacteria;Actinobacteria;Corynebacteriales;Nocardiaceae;Nocardia;Nocardia asteroides</i>
<b>VC66</b>	WC	Cruce	AIA	28°C	31/1/17	1440	99,72	<i>Actinobacteria;Actinobacteria;Streptomycetales;Streptomycetaceae;Streptomyces;Streptomyces sp</i>
<b>VC67</b>	GC	Cruce	AIA	28°C	31/1/17	1430	99,02	<i>Actinobacteria;Actinobacteria;Pseudonocardiales;PseudoNocardiaceae;Crossiella;Crossiella cryophila</i>
<b>VC68</b>	GC	Cruce	AIA	28°C	31/1/17	1449	99,58	<i>Actinobacteria;Actinobacteria;Streptomycetales;Streptomycetaceae;Streptomyces;Streptomyces exfoliatus</i>
<b>VC69</b>	GC	Cruce	AIA	28°C	31/1/17	1448	98,62	<i>Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacillus idriensis</i>
<b>VC70</b>	GC	Cruce	AIA	28°C	31/1/17	1432	99,08	<i>Actinobacteria;Actinobacteria;Corynebacteriales;Nocardiaceae;Rhodococcus;Rhodococcus maanshanensis</i>
<b>VC71</b>	GC	Vestibulo	AIA	28°C	31/1/17	1442	100,00	<i>Actinobacteria;Actinobacteria;Streptomycetales;Streptomycetaceae;Streptomyces;Streptomyces sp</i>
<b>VC72</b>	GC	Vestibulo	AIA	28°C	31/1/17	1472	96,65	<i>Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacillus idriensis</i>
<b>VC73</b>	GC	Vestibulo	AIA	28°C	31/1/17	1448	99,72	<i>Actinobacteria;Actinobacteria;Micrococcales;Microbacteriaceae;Agromyces;Agromyces fucosus</i>
<b>VC74</b>	GC	Vestibulo	AIA	28°C	31/1/17	1497	99,29	<i>Actinobacteria;Actinobacteria;Pseudonocardiales;Pseudonocardiaceae;Pseudonocardia;Pseudonocardia eucalypti</i>
<b>VC76</b>	GC	Vestibulo	AIA	28°C	31/1/17	1430	99,14	<i>Actinobacteria;Actinobacteria;Corynebacteriales;Nocardiaceae;Rhodococcus;Rhodococcus maanshanensis</i>
<b>VC77</b>	GC	Vestibulo	AIA	28°C	31/1/17	1437	99,79	<i>Actinobacteria;Actinobacteria;Micromonosporales;Micromonosporaceae;Micromonospora;Micromonospora matsumotoense</i>

<b>VC78</b>	GC	Vestibulo	AIA	28°C	31/1/17	1437	98,95	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia alba</i>
<b>VC79</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1441	99,79	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sporoclivatus</i>
<b>VC80</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1439	99,86	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC82</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1439	99,58	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC84</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1448	99,93	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC85</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1447	99,79	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC86</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1447	99,86	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC87</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1449	99,93	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces olivochromogenes</i>
<b>VC88</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1440	98,68	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC90</b>	YC	Vestibulo	TSA 0,1X	28°C	29/02/17	1439	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC91</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1442	98,00	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia salmonicida</i>

<b>VC92</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1442	99,51	<i>Actinobacteria;Actinobacteria;Corynebacteriales;Nocardiaceae;Rhodococcus;Rhodococcus wratislaviensis</i>
<b>VC93</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1441	99,58	<i>Actinobacteria;Actinobacteria;Corynebacteriales;Nocardiaceae;Rhodococcus;Rhodococcus koreensis</i>
<b>VC94</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1443	99,65	<i>Actinobacteria;Actinobacteria;Streptomycetales;Streptomycetaceae;Streptomyces;Streptomyces sp</i>
<b>VC95</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1447	98,62	<i>Actinobacteria;Actinobacteria;Micrococcales;Microbacteriaceae;Microbacterium;Microbacterium foliorum</i>
<b>VC96</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1447	99,93	<i>Actinobacteria;Actinobacteria;Micrococcales;Microbacteriaceae;Microbacterium;Microbacterium oxydans</i>
<b>VC97</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1451	98,61	<i>Bacteroidetes;Sphingobacteriia;Sphingobacteriales;Sphingobacteriaceae;Sphingobacterium;Sphingobacterium faecium</i>
<b>VC99</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1454	99,93	<i>Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;Pseudomonas chlororaphis</i>
<b>VC100</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1403	99,06	<i>Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Phyllobacterium;Phyllobacterium myrsinacearum</i>
<b>VC101</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1473	99,86	<i>Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacillus mobilis</i>
<b>VC102</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1512	99,00	<i>Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;Pseudomonas chlororaphis</i>
<b>VC104</b>	YC	Vestibulo	AIA	16°C, 80% H	31/04/17	1446	99,58	<i>Actinobacteria;Actinobacteria;Streptomycetales;Streptomycetaceae;Streptomyces;Streptomyces sp</i>

<b>VC107</b>	YC	Vestibulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	14/11/17	1441	99,65	<i>Actinobacteria; Actinobacteria; Streptomyetales; Streptomyetaceae; Streptomyces; Streptomyces sp</i>
<b>VC109</b>	WC	Vestibulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	14/11/17	1401	99,78	<i>Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Sphingobium; Sphingobium hydrophobicum</i>
<b>VC110</b>	GC	Vestibulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	24/11/17	1380	99,42	<i>Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae;; Aminobacter; Aminobacter anthyllidis</i>
<b>VC111</b>	YC	Vestibulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	24/11/17	1388	99,78	<i>Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Phyllobacterium;; Phyllobacterium trifolii</i>
<b>VC112</b>	WC	Vestibulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	24/11/17	1402	99,43	<i>Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Sphingomonadaceae; Sphingomonadaceae sp</i>
<b>VC113</b>	YC	Vestibulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	24/11/17	1400	99,72	<i>Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Phyllobacterium; Phyllobacterium trifolii</i>
<b>VC115</b>	GC	Vestibulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	15/12/17	1430	99,44	<i>Actinobacteria; Actinobacteria; Micrococcales; Microbacteriaceae; Frigoribacterium Frigoribacterium faeni</i>
<b>VC116</b>	YC	Vestibulo	AIA	28°C	2/11/18	1408	99,22	<i>Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae Agrobacterium Agrobacterium tumefaciens</i>
<b>VC117</b>	YC	Vestibulo	AIA	28°C	2/11/18	1461	99,79	<i>Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas; Pseudomonas chlororaphis</i>
<b>VC118</b>	YC	Vestibulo	AIA	28°C	2/11/18	1474	99,80	<i>Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus mobilis</i>
<b>VC121</b>	YC	Vestibulo	AIA	28°C	2/11/18	1436	99,00	<i>Actinobacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia alba</i>

<b>VC122</b>	YC	Vestibulo	AIA	28°C	2/11/18	1452	99,72	<i>Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas; Pseudomonas chlororaphis</i>
<b>VC123</b>	GC	Vestibulo	AIA	28°C	22/2/19	1423	99,36	<i>Actinobacteria; Actinobacteria; Pseudonocardiales; PseudoNocardiaceae; PseudoNocardia; Pseudonocardia eucalypti</i>
<b>VC124</b>	GC	Vestibulo	AIA	28°C	22/2/19	1441	99,45	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC125</b>	GC	Vestibulo	AIA	28°C	22/2/19	1407	99,05	<i>Proteobacteria; Alphaproteobacteria; Rhodospirillales; Rhodospirillaceae; Inquilinus; Inquilinus ginsengisoli</i>
<b>VC126</b>	GC	Vestibulo	AIA	28°C	22/2/19	1427	99,10	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces spiroverticillatus</i>
<b>VC127</b>	GC	Vestibulo	AIA	28°C	22/2/19	1444	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC128</b>	YC	Vestibulo	AIA	28°C	22/2/19	1444	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC129</b>	YC	Vestibulo	AIA	28°C	22/2/19	1444	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC130</b>	YC	Vestibulo	AIA	28°C	22/2/19	1441	99,72	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC131</b>	YC	Vestibulo	AIA	28°C	22/2/19	1442	99,65	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces finlayi</i>
<b>VC132</b>	YC	Vestibulo	AIA	28°C	22/2/19	1442	99,93	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>

<b>VC133</b>	YC	Vestíbulo	AIA	28°C	22/2/19	1442	100,00	<i>Actinobacteria; Actinobacteria; Streptomyetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC134</b>	YC	Vestíbulo	AIA	28°C	22/2/19	1439	99,79	<i>Actinobacteria; Actinobacteria; Micrococcales; Microbacteriaceae; Microbacterium; Microbacterium profundum</i>
<b>VC135</b>	YC	Vestíbulo	AIA	28°C	22/2/19	1451	99,86	<i>Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; Kocuria; Kocuria marina</i>
<b>VC136</b>	BC	Vestíbulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	24/11/20	1437	99,00	<i>Actinobacteria; Actinobacteria; Streptomyetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC137</b>	BC	Vestíbulo	AIA	16°C	24/11/20	1439	99,00	<i>Actinobacteria; Actinobacteria; Streptomyetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC138</b>	BC	Vestíbulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	24/11/20	1368	100,00	<i>Bacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Rhodococcus; Rhodococcus kronopolitis</i>
<b>VC139</b>	BC	Vestíbulo	TSA 0,1X +NaHCO <sub>3</sub>	28°C	24/11/20	1429	99,00	<i>Bacteria; Actinobacteria; Corynebacteriales; Nocardiaceae; Nocardia; Nocardia alba</i>
<b>VC146</b>	BC	Vestíbulo	TSA 0,1X	28°C	24/11/20	1439	99,00	<i>Bacteria; Actinobacteria; Streptomyetales; Streptomycetaceae; Streptomyces; Streptomyces mauvecolor</i>
<b>VC147</b>	BC	Vestíbulo	TSA 0,1X	28°C	24/11/20	1438	99,00	<i>Bacteria; Actinobacteria; Micrococcales; Microbacteriaceae; Agromyces; Agromyces cerinus</i>
<b>VC148</b>	BC	Vestíbulo	TSA 0,1X	16°C	24/11/20	1434	99,00	<i>Bacteria; Actinobacteria; Micrococcales; Microbacteriaceae; Microbacterium; Microbacterium oxydans</i>
<b>VC149</b>	BC	Vestíbulo	TSA 0,1X	16°C	24/11/20	1470	98,00	<i>Bacteria; Firmicutes; Bacilli; Bacillales; Paenibacillaceae; Paenibacillus; Paenibacillus sinopodophylli</i>

<b>VC150</b>	BC	Vestíbulo	TSA 0,1X	16°C	24/11/20	1438	99,00	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC151</b>	BC	Vestíbulo	TSA 0,1X +NaHCO <sub>3</sub>	16°C	24/11/20	1438	99,00	<i>Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces sp</i>
<b>VC152</b>	BC	Vestíbulo	TSA 0,1X +NaHCO <sub>3</sub>	16°C	24/11/20	1395	99,00	<i>Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Phyllobacterium; Phyllobacterium trifolii</i>
<b>VC157</b>	BC	Vestíbulo	TSA	16°C	24/11/20	1433	99,00	<i>Bacteria; Actinobacteria; Micrococcales; Micrococcaceae; Micrococcus; Micrococcus yunnanensis</i>
<b>VC161</b>	BC	Vestíbulo	TSA 0,1X	16°C	24/11/20	1436	99,00	<i>Bacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces mauvecolor</i>
<b>VC162</b>	BC	Vestíbulo	TSA 0,1X	28°C	24/11/20	1398	97,00	<i>Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales; Rhodospirillaceae sp</i>
<b>VC163</b>	BC	Vestíbulo	AIA	28°C	24/11/20	1386	100,00	<i>Bacteria; Actinobacteria; Pseudonocardiales; Pseudonocardiaceae; Pseudonocardia; Pseudonocardia eucalypti</i>
<b>VC164</b>	BC	Vestíbulo	AIA	28°C	24/11/20	1440	99,00	<i>Bacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces mauvecolor</i>
<b>VC165</b>	BC	Vestíbulo	AIA	28°C	24/11/20	1461	96,00	<i>Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus idriensis</i>
<b>VC166</b>	BC	Vestíbulo	AIA	28°C	24/11/20	1457	96,00	<i>Bacteria; Firmicutes; Bacilli; Bacillales; Paenibacillaceae; Paenibacillus; Paenibacillus endophyticus</i>
<b>VC172</b>	BC	Vestíbulo	TSA 0,1X	16°C	24/11/20	1438	99,00	<i>Bacteria; Actinobacteria; Micrococcales; Microbacteriaceae; Microbacterium; Microbacterium profundi</i>

Tabla S11. Lista de las bacterias de referencias y su número de acceso a GenBank más cercanas a los aislados cultivables de las colonias visibles utilizados en la construcción del árbol filogenético de los aislados.

<b>Taxonomía</b>	<b>Nº Acceso en GenBank</b>	<b>Bacteria de referencia</b>
<i>Firmicutes</i>	NR_157609.1	<i>Bacillus haynesii</i>
	NR_043268.1	<i>Bacillus idriensis</i>
	NR_116023.1	<i>Bacillus licheniformis</i>
	NR_157731.1	<i>Bacillus mobilis</i>
	NR_115993.1	<i>Bacillus mycoides</i>
	NR_114919.1	<i>Bacillus simplex</i>
	NR_116017.1	<i>Bacillus subtilis</i>
	NR_137421.1	<i>Bacillus paralicheniformis</i>
	NR_148786.1	<i>Bacillus zhangzhouensis</i>
	NR_026482.1	<i>Facklamia tabacinasalis</i>
	NR_159144.1	<i>Psychrobacillus lasiicapitis</i>
NR_029233.1	<i>Sporosarcina globispora</i>	
<i>Alphaproteobacteria</i>	NR_115516.1	<i>Agrobacterium tumefaciens</i>
	NR_108530.1	<i>Aminobacter anthyllidis</i>
	NR_040824.1	<i>Blastomonas natatoria</i>
	NR_029046.1	<i>Inquilinus limosus</i>
	NR_112560.1	<i>Inquilinus ginsengisoli</i>
	NR_113874.1	<i>Phyllobacterium myrsinacearum</i>
	NR_043193.1	<i>Phyllobacterium trifolii</i>
	NR_152071.1	<i>Sphingobium hydrophobicum</i>
NR_040825.1	<i>Sphingomonas ursincola</i>	
<i>Gammaproteobacteria</i>	NR_116763.1	<i>Pseudomonas chlororaphis</i>
<i>Actinobacteria</i>	NR_104982.1	<i>Agromyces fucosus</i>
	NR_024964.1	<i>Crossiella cryophila</i>
	NR_026511.1	<i>Frigoribacterium faeni</i>
	NR_025723.1	<i>Kocuria marina</i>
	NR_026451.1	<i>Kocuria palustris</i>
	NR_025368.1	<i>Microbacterium foliorum</i>
	NR_044931.1	<i>Microbacterium oxydans</i>
NR_044321.1	<i>Microbacterium profundum</i>	

	NR_075062.2	<i>Micrococcus luteus</i>
	NR_025015.1	<i>Micromonospora matsumotoense</i>
	NR_025726.1	<i>Nocardia alba</i>
	NR_041856.1	<i>Nocardia asteroides</i>
	NR_114571.1	<i>Nocardia salmonicida</i>
	NR_116816.1	<i>Pseudonocardia eucalypti</i>
	NR_114500.1	<i>Rhodococcus koreensis</i>
	NR_025190.1	<i>Rhodococcus maanshanensis</i>
	NR_026524.1	<i>Rhodococcus wratislaviensis</i>
	NR_112527.1	<i>Streptomyces anulatus</i>
	NR_112255.1	<i>Streptomyces atrolivaceus</i>
	NR_041132.1	<i>Streptomyces avidinii</i>
	NR_115396.1	<i>Streptomyces exfoliatus</i>
	NR_043354.1	<i>Streptomyces finlayi</i>
	NR_041154.1	<i>Streptomyces mauvecolor</i>
	NR_043854.1	<i>Streptomyces microflavus</i>
	NR_042303.1	<i>Streptomyces nojiriensis</i>
	NR_112483.1	<i>Streptomyces olivochromogenes</i>
	NR_041214.1	<i>Streptomyces spiroverticillatus</i>
	NR_041413.1	<i>Streptomyces sporoclivatus</i>
	NR_042306.1	<i>Streptomyces spororaveus</i>
	NR_026203.1	<i>Streptomyces subutilus</i>
<i>Bacteroidetes</i>	NR_025537.1	<i>Sphingobacterium faecium</i>



