

# **Estudio de la microbiota subgingival en sujetos sanos y pacientes con periodontitis de distintos orígenes geográficos**

**Alexandre Arredondo Campaña**  
Màster de Bioinformàtica y Bioestadística  
Genómica comparativa

**Yolanda Guillén Montalbán**  
**Javier Luis Cánovas Izquierdo**  
**Marc Maceira Duch**

Junio 2020



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Estudio de la microbiota subgingival en sujetos sanos y pacientes con periodontitis de distintos orígenes geográficos</i>
<b>Nombre del autor:</b>	<i>Alexandre Arredondo Campaña</i>
<b>Nombre del consultor/a:</b>	<i>Yolanda Guillén Montalbán</i>
<b>Nombre del PRA:</b>	<i>Javier Luis Cánovas Izquierdo Marc Maceira Duch</i>
<b>Fecha de entrega (mm/aaaa):</b>	06/2020
<b>Titulación:</b>	Máster de Bioinformática y Bioestadística
<b>Área del Trabajo Final:</b>	<i>Genómica comparativa</i>
<b>Idioma del trabajo:</b>	<i>Castellano</i>
<b>Palabras clave</b>	<i>Metagenómica, microbioma, periodontitis.</i>
<b>Resumen del Trabajo:</b>	
<p>La periodontitis es una inflamación de los tejidos periodontales que puede derivar en la pérdida de piezas dentales. Se ha descrito que un biofilm subgingival disbiótico puede causar esta inflamación, y por lo tanto se han dedicado muchos esfuerzos a la caracterización de la microbiota causante de la periodontitis. Estos estudios de caracterización han construido perfiles microbianos del biofilm subgingival en estado de salud y de enfermedad periodontal, aunque centrándose en poblaciones locales y sin tener en cuenta la variabilidad que el origen geográfico de las muestras puede aportar, creando un sesgo al universalizar dichos perfiles. Para determinar si existen diferencias significativas entre la microbiota subgingival de individuos sanos y de pacientes con periodontitis de 4 orígenes geográficos distintos (Bélgica, Perú, Chile y España), en este estudio se analizaron muestras subgingivales mediante secuenciación masiva, determinando la diversidad, riqueza y composición bacteriana en las muestras según su diagnóstico y origen geográfico. Las secuencias fueron procesadas mediante el programa mothur y el análisis estadístico se realizó con los paquetes estadísticos PhyloSeq, Vegan y DESeq2. Los resultados mostraron que existen diferencias significativas tanto por diagnóstico como por origen geográfico de las muestras, lo cual abre la puerta a realizar más estudios para analizar las causas y consecuencias de esta variabilidad.</p>	

**Abstract:**

Periodontitis is an inflammatory disease of the periodontal tissues, which can lead to tooth loss. It has been described that a dysbiotic subgingival biofilm can be the cause of such inflammation, and therefore many efforts have focused in the characterization of the microbiota that might cause periodontitis. These studies have built microbial profiles of the healthy and periodontitis-related subgingival biofilm, although based in local populations and without taking into account the variability that the geographical origin can add to the samples, thus creating a bias when universalizing such profiles. In order to assess if there are significant differences among the subgingival microbiota of healthy subjects and patients with periodontitis of 4 different geographical origins (Belgium, Peru, Chile and Spain), subgingival samples were analyzed using next generation sequencing, evaluating their diversity, richness and bacterial composition according to their diagnostic and geographical origin. Sequences were processed through the software mothur and the statistical analyses were performed using the R statistical packages PhyloSeq, Vegan and DESeq2. Results showed that there are significant differences in the samples, both by diagnostic and by geographical origin, which opens the door to further studies in order to analyze the causes and consequences of such variability.

# Índice

1. Introducción	1
1.1 Contexto y justificación del trabajo	1
1.2 Objetivos del trabajo	3
1.2.1 Objetivos generales	3
1.2.2 Objetivos específicos	4
1.3 Enfoque y método elegido	4
1.4 Planificación del trabajo	5
1.4.1 Tareas	5
1.4.2 Calendario	6
1.4.3 Hitos	7
1.4.4 Análisis de riesgos	7
1.5 Resultados esperados	7
1.6 Breve descripción de los otros capítulos de la memoria	8
2. Material y métodos	9
2.1 Origen de las muestras	9
2.2 Criterios de inclusión	9
2.3 Criterios de exclusión	9
2.4 Toma de muestras	9
2.5 Extracción de DNA	10
2.6 Secuenciación masiva	10
2.7 Análisis de las secuencias	11
2.8 Análisis estadístico	12
3. Resultados	13
3.1 Características de las muestras	13
3.2 Análisis del microbioma	13
4. Discusión	22
5. Conclusiones	26
6. Glosario	27
7. Bibliografía	28
8. Anexos	32
8.1 Anexo 1 – Procotolo Mothur	32
8.2 Anexo 2 – Código R para el análisis estadístico	35

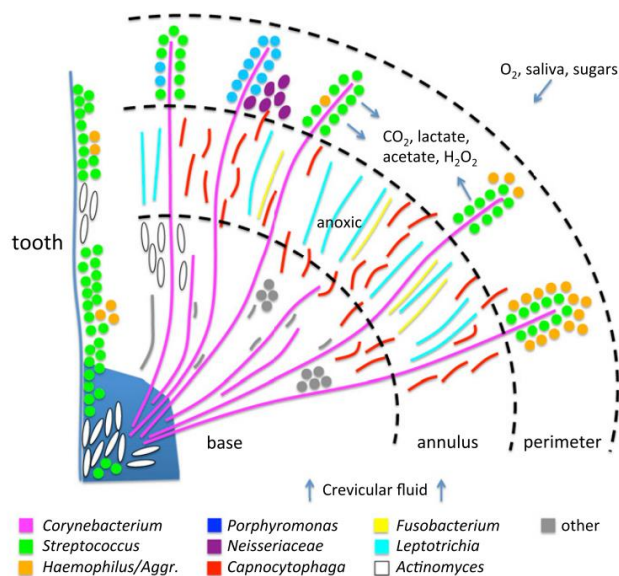
## Lista de figuras

<b>Figura 1.</b> Esquema de la estructura del biofilm subgingival	1
<b>Figura 2.</b> Esquema de clústers bacterianos y su asociación a la enfermedad periodontal	2
<b>Figura 3.</b> Calendarización del TFM	6
<b>Figura 4.</b> Esquemmatización del flujo de trabajo usado en mothur	12
<b>Figura 5.</b> Gráficos de riqueza y diversidad bacteriana	14
<b>Figura 6.</b> Gráfico de diversidad basada en los 10 géneros más detectados	15
<b>Figura 7.</b> Gráficos NMDS	16
<b>Figura 8.</b> Gráfico de diferencias de OTUs según la geografía	19
<b>Figura 9.</b> Gráfico de diferencias de OTUs entre países de sujetos sanos	20
<b>Figura 10.</b> Gráfico de diferencias de OTU entre países de pacientes con periodontitis	21

# 1. Introducción

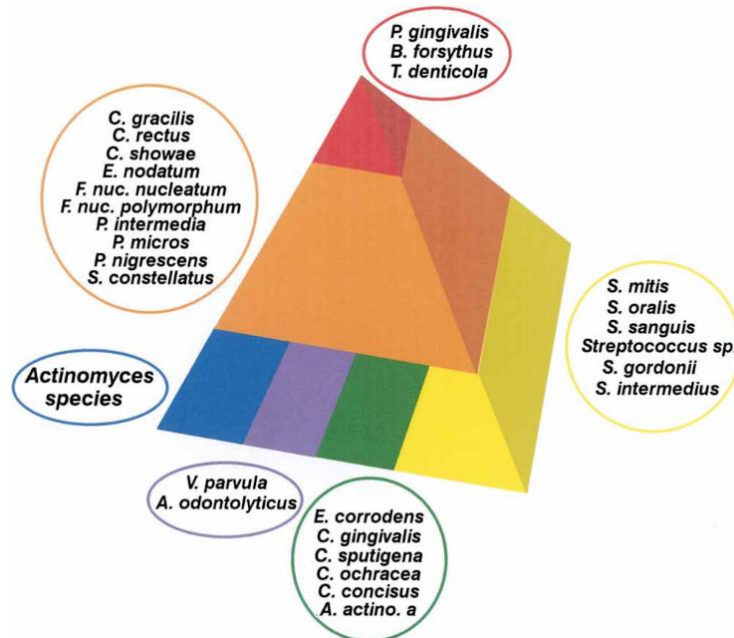
## 1.1 Contexto y justificación del trabajo

La cavidad oral es un nicho ampliamente colonizado por microorganismos debido a la alta disponibilidad de nutrientes, la variedad de superficies y a la facilidad de acceso desde el exterior<sup>1</sup>. Estos microorganismos suelen crecer en las superficies formando un biofilm, que se inicia al establecerse una interacción entre la película adquirida, conjunto de proteínas y glicoproteínas que se deposita sobre el esmalte dental, y los colonizadores primarios, los cuales emplean adhesinas para unirse a dicha película. Estos colonizadores primarios suelen pertenecer a géneros bacterianos como *Streptococcus*, *Actinomyces*, *Capnocytophaga*, *Veillonella* y *Campylobacter*<sup>2-5</sup>. Estos microorganismos tienen la capacidad de coagregar con otros géneros bacterianos, los considerados colonizadores tardíos, como por ejemplo *Prevotella*, *Fusobacterium* y *Corynebacterium* (Figura 1). Estos dos últimos géneros son de especial importancia puesto que se ha descrito que son pilares fundamentales en la formación del biofilm patogénico, favoreciendo la coagregación de periodontopatógenos como *Porphyromonas gingivalis*, *Tannerella forsythia* y *Treponema denticola*, los cuales han sido asociados con enfermedad periodontal<sup>3,6-9</sup>.



**Figura 1.** Representación esquemática de un biofilm en el que se muestra como las bacterias del género *Corynebacterium* (en morado) tienen un importante papel estructural (imagen extraída de la referencia Welch et al. 2016).

Esta asociación con la enfermedad periodontal fue descrita, entre otros, por Socransky y Haffajje<sup>8</sup>, quienes realizaron un modelo piramidal en el que se establecía, mediante un código de colores, la progresión hacia la enfermedad periodontal dependiendo de los microorganismos presentes en las bolsas subgingivales (figura 2).



**Figura 2.** Complejos de Socransky y Haffajje. En la base de la pirámide se observan los 4 grupos de colonizadores primarios. Encima de estos colonizadores se asienta el complejo naranja y sobre este el rojo. La estructuración del biofilm suele seguir el orden mostrado en el esquema, aunque en menor medida se pueden dar casos en los que los complejos naranja o rojo asuman el rol de colonizadores primarios, o que el complejo rojo se establezca encima de los colonizadores primarios (imagen extraída de la referencia Socransky and Haffajee 2002).

La periodontitis es una enfermedad infecciosa crónica no transmisible que cursa con inflamación de los tejidos periodontales y que puede derivar en la pérdida de piezas dentales debido a la destrucción del hueso alveolar que sirve de soporte al diente<sup>10,11</sup>. El diagnóstico de la periodontitis se basa en el análisis de las variables clínicas como, por ejemplo: profundidad de sondaje y sangrado, presencia o ausencia de signos de inflamación, alcance y patrón de pérdida de hueso y nivel clínico de inserción, etc.<sup>12-14</sup> Algunas clínicas añaden a su diagnóstico la detección de microorganismos, ya sea por medios microbiológicos clásicos o por medios moleculares tales como la detección de genes específicos<sup>15-17</sup>. Estos análisis buscan confirmar la presencia o ausencia de bacterias como *P. gingivalis*, *T. forsythia*, *P. denticola*, *Fusobacterium nucleatum*, etc. con el objetivo de determinar la progresión de la enfermedad



periodontal desde un punto de vista microbiológico, y para determinar la mejor terapia antibiótica basada en los perfiles de susceptibilidad publicados.

La asociación de perfiles microbianos a la enfermedad periodontal no está carente de sesgos, puesto que o bien se basa en modelos antiguos que solamente tenían en cuenta la microbiota cultivable<sup>3</sup> o bien se basan en estudios de secuenciación masiva que se han limitado a poblaciones pequeñas y restringidas a ciertas zonas geográficas<sup>18-21</sup>. Teniendo en cuenta que estos estudios pueden ser de utilidad para desarrollar futuras terapias para el tratamiento de la periodontitis, es esencial que se tenga una fotografía real de la asociación entre la microbiota presente en las bolsas periodontales y el desarrollo de la enfermedad. Bajo esta premisa, en este estudio se plantearon los objetivos detallados en el siguiente apartado.

## **1.2 Objetivos del Trabajo**

### **1.2.1 *Objetivos generales***

1. Caracterizar la microbiota oral en individuos sanos e identificar los microorganismos diferencialmente representados entre varias localizaciones geográficas.
2. Explorar los microorganismos asociados al estado de periodontitis en base a la comparación con individuos sanos, teniendo en cuenta los microorganismos diferencialmente representados en cada zona geográfica.

### **1.2.2 Objetivos específicos**

Siguiendo la numeración de los objetivos principales:

3. Analizar las secuencias obtenidas mediante secuenciación masiva para obtener un perfil microbiano de la microbiota subgingival de cada uno de los sujetos del estudio, tantos sanos como afectados por la patología.
  4. Realizar un análisis estadístico comparativo para determinar las posibles diferencias entre los perfiles microbianos previamente analizados, buscando diferencias entre los sujetos sanos por un lado y entre los pacientes con periodontitis por el otro, teniendo en cuenta su origen geográfico.
  5. Discutir las causas de las diferencias, o falta de ellas, entre los perfiles microbianos.
- 2.2 Realizar un análisis estadístico que determine si los perfiles microbianos obtenidos presentan diferencias entre si al comparar la microbiota de sujetos sanos con la de pacientes con periodontitis.
  - 2.3 Comparar las diferencias anteriores, si las hubiese, entre todos los orígenes geográficos de las muestras.
  - 2.4 Discutir las causas de las diferencias si las hubiese.

### **1.3 Enfoque y método elegido**

De entre las distintas plataformas de secuenciación que existen, en este estudio se empleó el secuenciador Illumina MiSeq. Con este secuenciador se obtienen secuencias correspondientes a las regiones V3 – V4 del gen ribosomal 16S de un tamaño medio de 300 pares de bases.

Para el análisis de las secuencias se puede utilizar software de pago y con interfaz gráfica como *Geneious* o *CLC Genomic Workbench*, o software libre como *Galaxy* o *Mothur*. Pese a que en el centro donde se desarrollará el estudio se dispone de *Geneious*, para el TFM se empleará *Mothur*, tanto por su filosofía Open Source como por ser multiplataforma y por lo tanto extrapolable a otros entornos de trabajo.

Para el análisis bioestadístico se utilizará Bioconductor debido a sus múltiples paquetes diseñados para el análisis de muestras biológicas en entorno R. Para trabajar con R se utilizará el software RStudio, el cual, tras una breve introducción a su funcionamiento, facilita enormemente el flujo de trabajo con R.

## **1.4 Planificación del Trabajo**

### **1.4.1 Tareas**

Las tareas y su duración se presentan ordenadas por cada objetivo específico y con un porcentaje aproximado de la duración relativa a la duración total de la realización del TFM:

1. *Analizar las secuencias obtenidas mediante secuenciación masiva para obtener un perfil microbiano de la microbiota subgingival de cada uno de los sujetos del estudio.*
  - a. Utilizar el software *Mothur* para procesar las secuencias que hemos obtenido mediante secuenciación masiva.
2. *Realizar un análisis estadístico para determinar las posibles diferencias entre los perfiles microbianos previamente analizados, e identificar qué grupos de microorganismos contribuyen a las diferencias entre localizaciones geográficas, tanto en pacientes sanos como afectados..*
  - a. Usando RStudio y Bioconductor, realizar el análisis estadístico.

3. *Discutir las causas de las diferencias, o falta de ellas, entre los perfiles microbianos.*

a. Apoyándonos en la literatura existente, relacionar los datos obtenidos con los de otros estudios y obtener conclusiones.

4. *Realizar un análisis estadístico que determine si los perfiles microbianos obtenidos presentan diferencias entre si al comparar la microbiota de sujetos sanos con la de pacientes con periodontitis.*

a. Usando RStudio y Bioconductor, realizar el análisis estadístico.

5. *Comparar las diferencias anteriores, si las hubiese, entre todos los orígenes geográficos de las muestras.*

a. Usando RStudio y Bioconductor, realizar el análisis estadístico.

6. *Discutir las causas de las diferencias si las hubiese.*

a. Apoyándonos en la literatura existente, relacionar los datos obtenidos con los de otros estudios y obtener conclusiones.

Redacción y revisión del TFM.

### 1.4.2 Calendario

En la siguiente figura (Figura 3) se muestra el calendario proyectado para la realización de las tareas de este Trabajo.



Figura 3. Calendario de cumplimiento de las tareas del TFM dispuesto gráficamente.

### **1.4.3 Hitos**

16/03/2020: Entrega de la PEC1.

26/03/2020: Fecha límite para tener las secuencias analizadas.

20/04/2020: Fecha límite para tener el análisis estadístico finalizado.

22/04/2020: Entrega de la PEC2.

18/05/2020: Entrega de la PEC3.

22/05/2020: Los resultados han sido analizados y discutidos.

24/06/2020: Fecha límite para entregar la memoria del TFM.

28/06/2020: Fecha límite para preparar la presentación del TFM.

### **1.4.4 Análisis de riesgos**

1. La falta de familiaridad con el software *Mothur* puede introducir complicaciones en el análisis de las secuencias.
2. Elegir un pipeline de tests estadísticos erróneo puede llevar a retrasos debido a la necesidad de repetir el proceso.
3. Defectos en las secuencias a analizar pueden dar lugar a tener que repensar las expectativas del estudio, si por ejemplo todas las secuencias de un país han resultado ser defectuosas debido a una mala toma de muestras, a una mala conservación de las mismas o a un mal procesado durante el trabajo de secuenciación.
4. Retrasos en los análisis estadísticos y de las secuencias pueden provocar que la fase de discusión y de redacción de la memoria se vean más comprimidas y se vea comprometida su calidad.

## **1.5 Resultados esperados**

En este estudio esperamos encontrar diferencias entre la microbiota subgingival de sujetos sanos y la de pacientes con periodontitis, lo que nos permitiría confirmar la teoría de que existen distintos perfiles microbianos dependiendo del estado de salud periodontal. Además, es posible que también encontremos diferencias significativas entre dicha microbiota dependiendo del origen geográfico de las muestras, lo cual añadiría complejidad al análisis de

los datos y a la posible descripción de un microbioma asociado a la enfermedad periodontal.

Por último, la información obtenida de este análisis podría ser compilada para realizar un artículo científico o una comunicación en algún congreso especializado.

## **1.6 Breve descripción de los otros capítulos de la memoria**

- **Material y métodos:** conjunto de técnicas empleadas para la realización del estudio, con el detalle suficiente para asegurar su reproducibilidad. Los anexos complementan la información aportada en este capítulo.
- **Resultados:** exposición en forma de texto, figuras y tablas de los principales hallazgos obtenidos en este estudio.
- **Discusión:** análisis de los resultados obtenidos, comparándolos con otros estudios y proponiendo las propias deducciones.
- **Conclusiones:** análisis final del trabajo, tanto a nivel de aprendizaje como a nivel de aportación al campo de conocimiento.

## **2. Material y métodos**

### **2.1 Origen de las muestras**

En este estudio se emplearon 240 muestras subgingivales obtenidas de 240 voluntarios de cuatro centros ubicados en cuatro estados distintos: Universitat Internacional de Catalunya (Barcelona, España), Pontificia Universidad Católica de Chile (Santiago, Chile), Universidad Científica del Sur (Lima, Perú) y Katholieke Universiteit Leuven (Lovaina, Bélgica). Cada centro universitario recolectó 30 muestras procedentes de 30 individuos sanos y 30 muestras de 30 pacientes con periodontitis (ver criterios de inclusión y exclusión a continuación). El estudio fue aprobado por un comité ético y cumplía con los principios de la Declaración de Helsinki.

### **2.2 Criterios de inclusión**

El grupo de pacientes con periodontitis lo formaron aquellos sujetos diagnosticados con periodontitis generalizada de estadio III o IV<sup>22</sup>. El grupo de sujetos sanos debían presentar un buen estado de salud general sin historia médica relacionada con enfermedad periodontal (gingivitis moderada o severa incluida).

### **2.3 Criterios de exclusión**

Ninguno de los pacientes tomó antibióticos en los últimos 6 meses antes del estudio. El embarazo y la lactancia fueron criterios de exclusión, así como la presencia de ortodoncias o prótesis fijas o removibles. Fumar más de 5 cigarrillos por día fue otro criterio de exclusión aplicado a los individuos sanos.

### **2.4 Toma de muestras**

Las muestras subgingivales fueron tomadas de la bolsa más profunda de cada cuadrante. Cada área fue aislada mediante torundas de algodón y se eliminó la

acumulación de placa supragingival mediante el uso de curetas. Tras este procedimiento se insertaron durante 20 segundos dos puntas de papel estériles en la bolsa periodontal escogida. Todas las puntas de un mismo sujeto se juntaron en un mismo tubo conteniendo fluido de transporte reducido (FTR) sin EDTA<sup>23</sup> y se mantuvieron congeladas hasta su transporte al laboratorio de microbiología (Dentaid Research Center, Cerdanyola del Vallès, España) en condiciones de frío.

## 2.5 Extracción de DNA:

La extracción de DNA genómico de las muestras subgingivales se realizó mediante el kit QiAamp DNA Mini Kit (Qiagen, Valencia, CA, EEUU) siguiendo las instrucciones del fabricante. El DNA fue cuantificado mediante fluorometría, utilizando el kit Qubit® dsDNA HS Assay Kits (Life Technologies, Eugene, OR, EEUU).

## 2.6 Secuenciación masiva

El protocolo de secuenciación masiva fue el indicado por el fabricante y se describe brevemente: en primer lugar se amplificó el fragmento V3-V4 del gen 16S rRNA del DNA extraído utilizando un termociclador T3000 (Biomebtra, Goettingen, Alemania) y empleando la mix de PCR y el programa de termociclador de la tabla 1.

**Tabla 1.** Reactivos y condiciones de la PCR empleada para amplificar las regiones V3 y V4.

Mix PCR	Programa termociclador
<ul style="list-style-type: none"> <li>• 12,5 ng de DNA.</li> <li>• 0,2 µM primer Forward.</li> <li>• 0,2 µM primer Reverse.</li> <li>• 12,5 µl 2x KAPA HiFi HotStart ReadyMix.</li> <li>• Ajustar con agua ultrapura grado Mili-Q hasta 25 µl.</li> </ul>	<ol style="list-style-type: none"> <li>1. 95°C 3 min.</li> <li>2. 25 ciclos de:               <ol style="list-style-type: none"> <li>a. 95°C 30 s</li> <li>b. 55°C 30 s</li> <li>c. 72°C 30 s</li> </ol> </li> <li>3. 72°C 5 min.</li> <li>4. Hold a 4°C</li> </ol>

Los amplicones se comprobaron en geles de agarosa al 3%, observándose una banda de alrededor de 550 pares de bases. En aquellos casos donde no se observara banda o ésta fuera muy tenue se repitió la PCR. Posteriormente, se

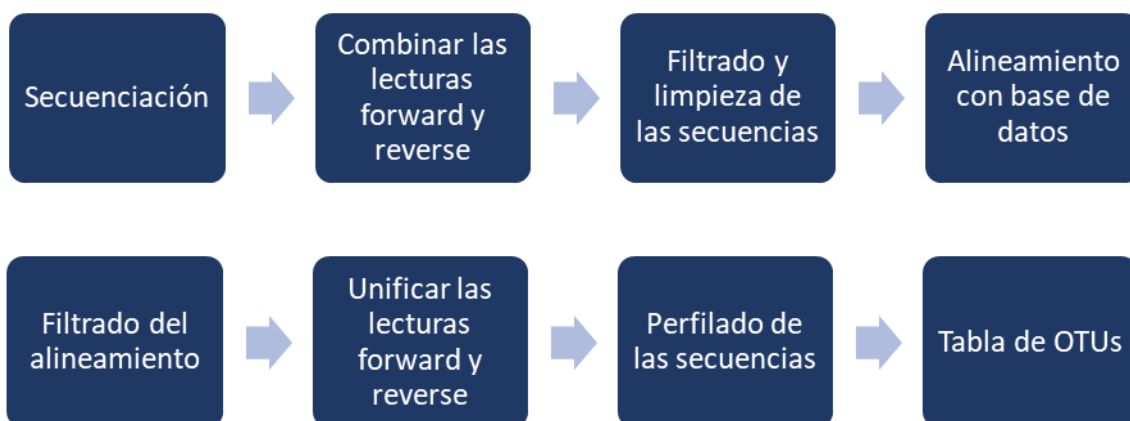


limpió el DNA amplificado de los primers y dímeros presentes usando perlas magnéticas suministradas en el kit del fabricante y siguiendo sus instrucciones. Una vez purificado, se añadieron los índices mediante PCR, elegidos mediante la herramienta Illumina Experiment Manager. En este estudio se utilizaron los sets A (ref. FC-131-2001, índices N701 - N707, N710 - N712, N714, N715; S502, S503, S505 - S508, S510, S511) y D (ref. FC-131-2004, índices N716, N718 - N724, N726 - N729; S513, S515 - S518, S520 - S522) del kit Nextera XT Index Kit v2. Se comprobaron los amplicones en un gel de agarosa al 3% observándose bandas de unos 630 pares de bases y se repitió la PCR en el caso de observarse poco DNA o de no ver banda. Tal y como se hizo con la anterior PCR, se volvió a limpiar el DNA para eliminar residuos indeseados usando el mismo protocolo. Una vez limpio, se cuantificó el DNA con el kit Qubit® dsDNA HS Assay Kits y se calculó la molaridad para lograr una concentración de DNA final de 4 nM. Finalmente, se preparó el cartucho de secuenciación MiSeq Reagent v3 siguiendo las instrucciones del fabricante.

## 2.7 Análisis de las secuencias

Las secuencias obtenidas fueron analizadas (Figura 4) utilizando el programa mothur<sup>24</sup> y la base de datos SILVA<sup>25</sup> para identificar las secuencias a nivel de género. El *output* del secuenciador fueron 10,86 millones de secuencias totales. Al filtrar por tamaño (470 pb) y eliminar ambigüedades, eliminamos el 16.74% de las secuencias, quedándonos con  $9.045259 \times 10^6$ . Posteriormente, y para aligerar la carga computacional, se eliminaron los duplicados, quedándonos con el 25.74% de las secuencias. Estas secuencias fueron alineadas con la base de datos SILVA, restringiendo el alineamiento a las regiones 3000 y 27000, que corresponden a las regiones V3-V4 del gen 16S. Tras alinear, se observó que la mayoría de las secuencias se colocaron entre las posiciones 3388 y 22316, por lo que realizamos otro filtrado en el que eliminamos el 5.15% de las secuencias, quedándonos con un total de  $2.208659 \times 10^6$ . El último paso del procesamiento de las secuencias consistió en perfilarlas, eliminando  $2.2672 \times 10^4$  caracteres en el proceso. Una vez alineadas y filtradas todas las secuencias, se creó el listado de unidades taxonómicas operativas (UTOs u OTUs en inglés), aplicando un filtro de 4

nucleótidos de diferencia por secuencia siguiendo las recomendaciones de los desarrolladores de Mothur. Durante este proceso se buscaron y eliminaron las quimeras (31.3% de las secuencias), dando lugar a un total de  $3.24819 \times 10^5$  OTUs. En el anexo 1 se puede consultar el listado de órdenes ejecutadas en Mothur para realizar este análisis.



**Figura 4.** Protocolo esquematizado de tratamiento de las secuencias con el programa Mothur, con el objetivo de lograr una lista de OTUs asociados a taxonomía.

## 2.8 Análisis estadístico

Las secuencias fueron posteriormente analizadas mediante los paquetes estadísticos de R PhyloSeq<sup>26</sup>, Vegan<sup>27</sup> y DESeq2<sup>28</sup> utilizando el programa Rstudio (el código empleado está disponible en el anexo 2). La diversidad y la riqueza de las muestras fue estudiada mediante los estimadores Chao1 y Shannon y Simpson respectivamente, del paquete PhyloSeq. Para determinar si existían diferencias significativas entre la diversidad y riqueza de las muestras se utilizó el test no paramétrico Mann-Whitney.

Para agrupar los OTUs según las variables, se utilizó escalamiento multidimensional no métrico (non-metric multidimensional scaling, o NMDS en inglés), y se utilizó el test ADONIS, disponible en el paquete Vegan, para determinar si las variables de este estudio tenían un impacto estadísticamente significativo en la composición de los microbiomas subgingivales.

### 3. Resultados

#### 3.1 Características de las muestras

En este estudio se utilizaron 240 muestras subgingivales, clasificadas según su origen geográfico y su estado de salud periodontal (tabla 2).

**Tabla 2.** Listado de muestras utilizadas en este estudio.

Número de muestras	Origen geográfico	Diagnóstico
30	Bélgica	Sanos
30	Bélgica	Periodontitis
30	España	Sanos
30	España	Periodontitis
30	Perú	Sanos
30	Perú	Periodontitis
30	Chile	Sanos
30	Chile	Periodontitis

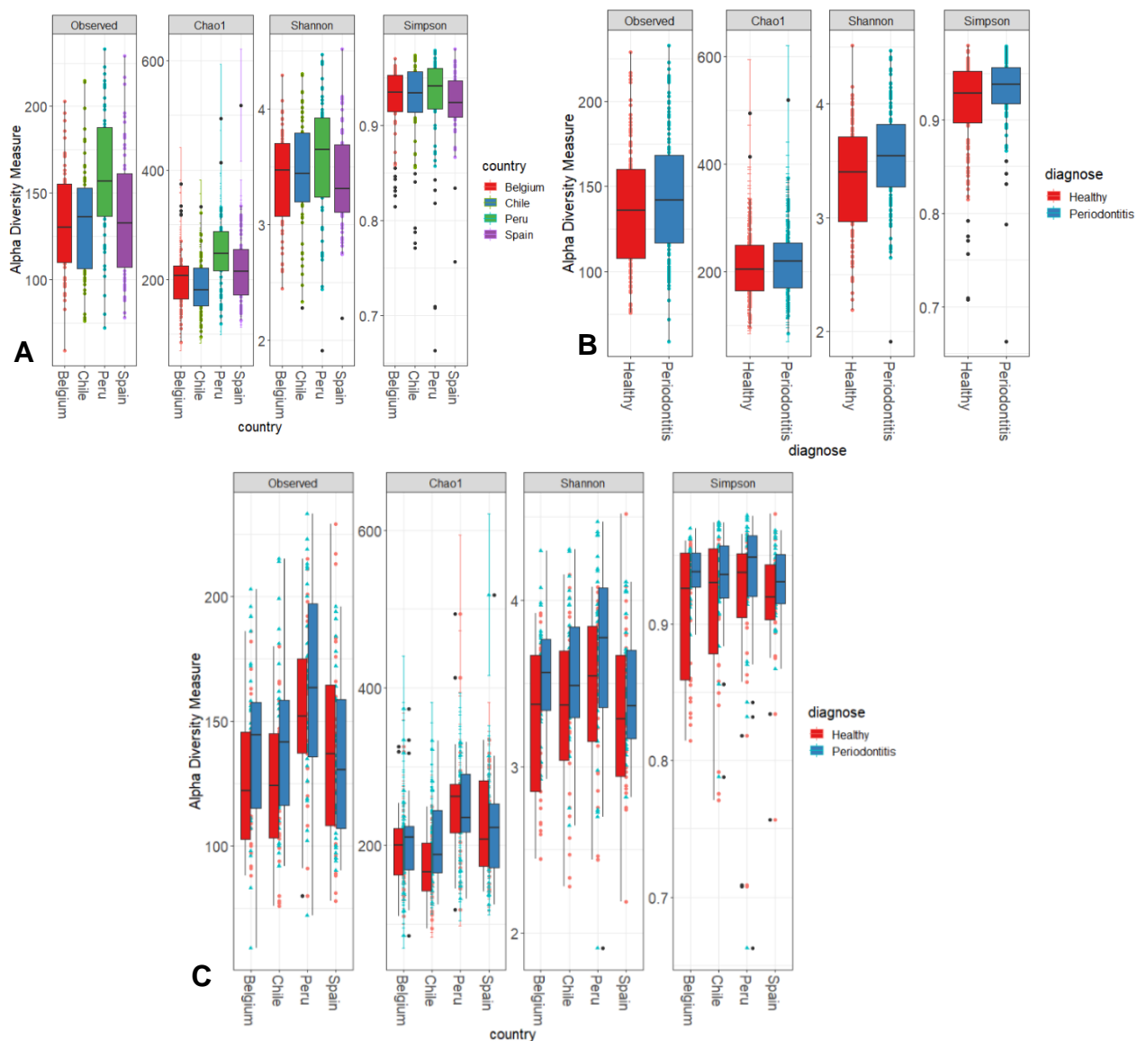
#### 3.2 Análisis del microbioma

Al analizar la diversidad y la riqueza entre las distintas zonas geográficas y/o condición (Figura 5), pudimos observar que la riqueza era significativamente distinta al comparar la microbiota subgingival de Bélgica y Perú ( $p = 1.8 \times 10^{-4}$ ), Perú y Chile ( $p = 1.0 \times 10^{-5}$ ) y de Chile y España ( $p = 1.95 \times 10^{-2}$ ). No se observaron diferencias significativas al comparar la riqueza de microbiota subgingival de todos los pacientes con periodontitis con la de todos los sujetos sanos, aunque al analizar país por país, se observaron diferencias significativas en la población chilena ( $p = 1.1 \times 10^{-2}$ ). La diversidad no presentó diferencias significativas entre países pero sí entre diagnósticos ( $p = 4.0 \times 10^{-3}$  y  $p = 3.8 \times 10^{-3}$ , según los índices de Shannon y Simpson, respectivamente). Al comparar los sujetos sanos con los pacientes con periodontitis en cada país, tan solo se observaron diferencias significativas en la población belga y únicamente tomando como medida el índice de Shannon ( $p = 4.1 \times 10^{-2}$ ). La diversidad de los 9 géneros más representados en las muestras analizadas se muestra en la figura 6.

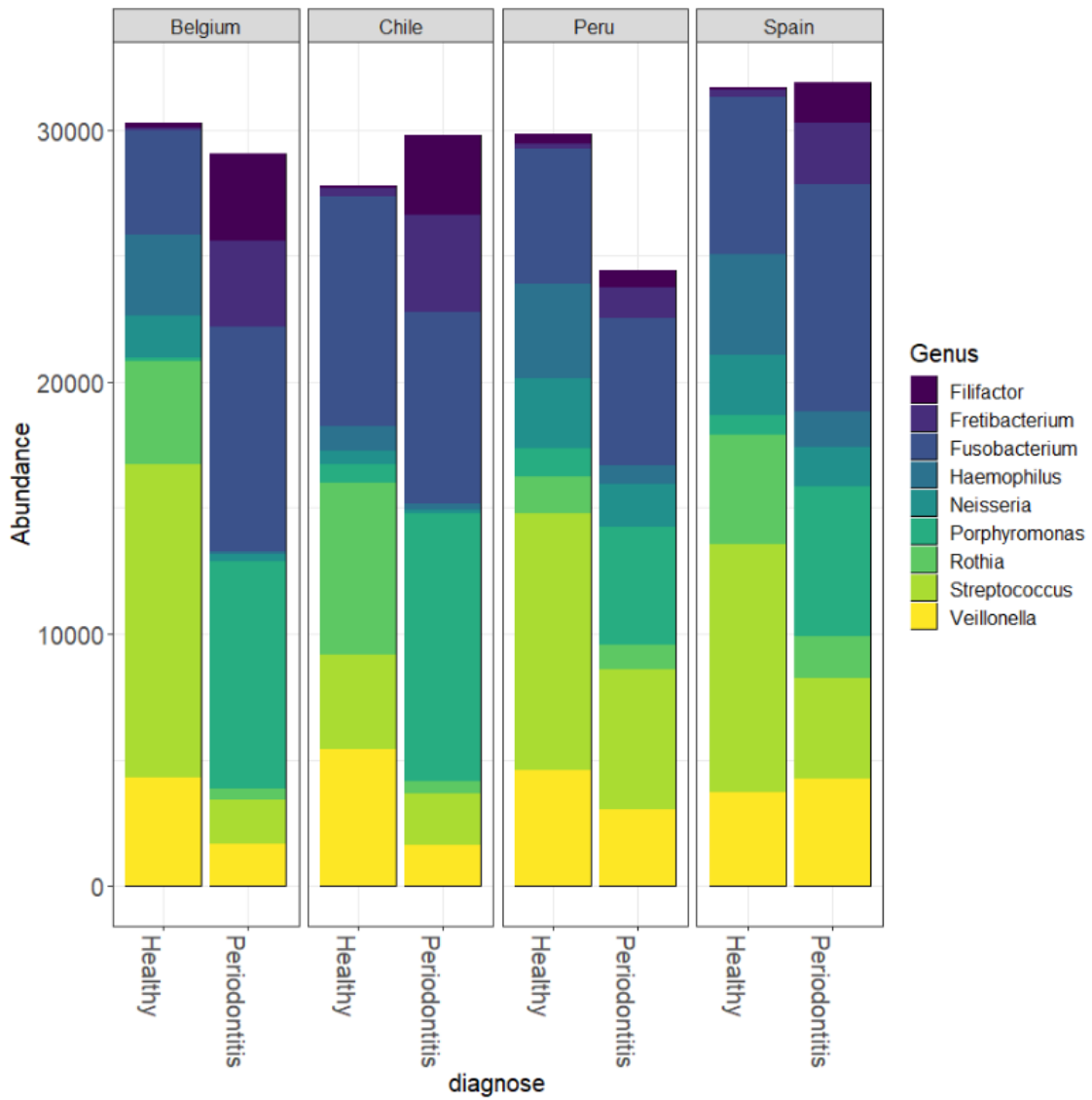
Para ver si los OTUs se agrupaban según las variables de las muestras, se empleó NMDS (Figura 7), el cual reportó un valor de estrés de 0.187, indicando que la ordenación no fue arbitraria. Los agrupamientos demostraron ser

estadísticamente significativos según el análisis multivariante ADONIS, que dio unos resultados de  $R^2$  de 0.048 y 0.139 ( $p = 0.001$ ) al analizar la variable de países y diagnóstico respectivamente, lo cual indica que el diagnóstico tiene un mayor peso a la hora de determinar la estructura microbiana que el origen geográfico en las muestras analizadas.

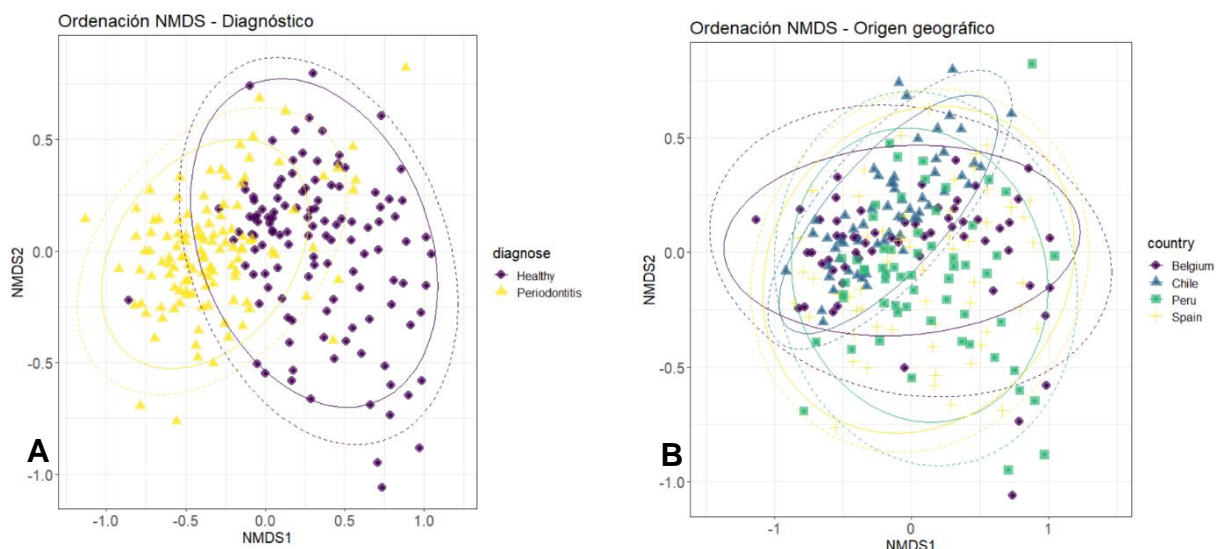
Para determinar las diferencias en la distribución de los géneros bacterianos entre las distintas poblaciones analizadas, se empleó el paquete DESeq2. Los resultados de este análisis se muestran en las tabla 3, 4 y 5 y las figura 8, 9 y 10. Para facilitar la lectura de los datos únicamente se muestran aquellos 10 OTUs con el  $P_{adj}$  menor.



**Figura 5.** Riqueza y diversidad expresadas según los índices de Chao1 y Sannon y Simpson respectivamente agrupadas según su origen geográfico (A), según su diagnóstico (B) y según ambas variables a la vez (C).



**Figura 6.** Diversidad de los 9 géneros más representados en las muestras analizadas, separadas por diagnóstico y origen geográfico. Para la realización de este gráfico se igualaron aleatoriamente el número de reads por muestra, por lo que los valores son relativos y no absolutos.



**Figura 7.** Gráficos NMDS de distancias Bray-Curtis. En el gráfico A se disponen las muestras y se resaltan según su diagnóstico, y en el gráfico B se resalta el origen geográfico.

**Tabla 3.** Lista de los 10 géneros con diferencias más significativas entre pacientes con periodontitis y sujetos sanos según el origen geográfico de las muestras, ordenados según su nivel de significancia estadística.

Pacientes con periodontitis vs sujetos sanos							
Bélgica				Perú			
Genus	Base Mean	Log 2 FC	$P_{adj}$	Genus	Base Mean	Log 2 FC	$P_{adj}$
<i>Leptotrichia</i>	143,53	28,76	2,77E-73	<i>Anaerovorax</i>	21,93	24,58	1,06E-24
<i>Prevotella</i>	123,54	28,19	1,22E-48	<i>Treponema</i>	627,55	7,23	4,40E-12
<i>Oribacterium</i>	34,47	26,76	4,92E-42	<i>Eubacterium</i>	160,09	7,85	1,01E-11
<i>Lachnospiraceae</i> unclassified	106,46	28,34	7,54E-40	<i>Lachnospiraceae</i> unclassified	53,98	7,74	7,17E-10
<i>Porphyromonas</i>	6933,83	13,26	9,77E-33	<i>Prevotella</i>	175,81	7,20	9,30E-10
<i>Mycoplasma</i>	320,47	12,93	9,02E-32	<i>Stomatobaculum</i>	40,80	8,62	1,76E-09
<i>Anaerolineaceae</i> unclassified	152,21	11,86	2,18E-31	<i>Desulfobulbus</i>	49,67	6,87	3,18E-08
<i>Odoribacter</i>	122,58	28,54	5,08E-29	<i>Streptococcus</i>	1522,40	3,97	4,39E-08
<i>Treponema</i>	1025,95	9,72	9,55E-28	<i>Prevotella</i>	505,11	5,06	4,39E-08
<i>Lachnospiraceae</i> unclassified	317,27	13,57	1,53E-27	<i>Eubacterium</i>	5,32	5,76	6,54E-08
España				Chile			
Genus	Base Mean	Log 2 FC	$P_{adj}$	Genus	Base Mean	Log 2 FC	$P_{adj}$
<i>Lachnospiraceae</i> unclassified	9,31	24,20	1,05E-22	Other	25,52	25,37	2,43E-27
<i>Aggregatibacter</i>	23,44	25,49	1,23E-18	<i>Porphyromonas</i>	7374,56	9,46	2,48E-20
<i>Haemophilus</i>	133,13	-7,40	2,82E-13	<i>Tannerella</i>	1749,41	6,73	1,96E-16
<i>Prevotella</i>	638,34	9,83	2,49E-10	<i>Filifactor</i>	2788,13	7,61	2,19E-15
<i>Anaerovorax</i>	108,12	6,56	3,29E-09	<i>Treponema</i>	11,72	24,35	2,12E-14
<i>Acinetobacter</i>	40,17	-9,98	3,29E-09	<i>Fretibacterium</i>	3078,52	6,16	1,26E-13
<i>Treponema</i>	1183,23	7,64	7,91E-09	<i>Anaerolineaceae</i> unclassified	154,37	7,50	1,26E-13
<i>Filifactor</i>	1800,70	6,53	4,26E-08	<i>Phocaecicola</i>	156,78	7,60	2,46E-13
<i>Stomatobaculum</i>	40,80	8,34	8,35E-08	<i>Sphingobium</i>	77,51	-8,33	4,54E-13
<i>Cellulomonadaceae</i> unclassified	72,38	-4,62	1,13E-07	<i>Pseudoramibacter</i>	39,17	7,00	1,45E-12

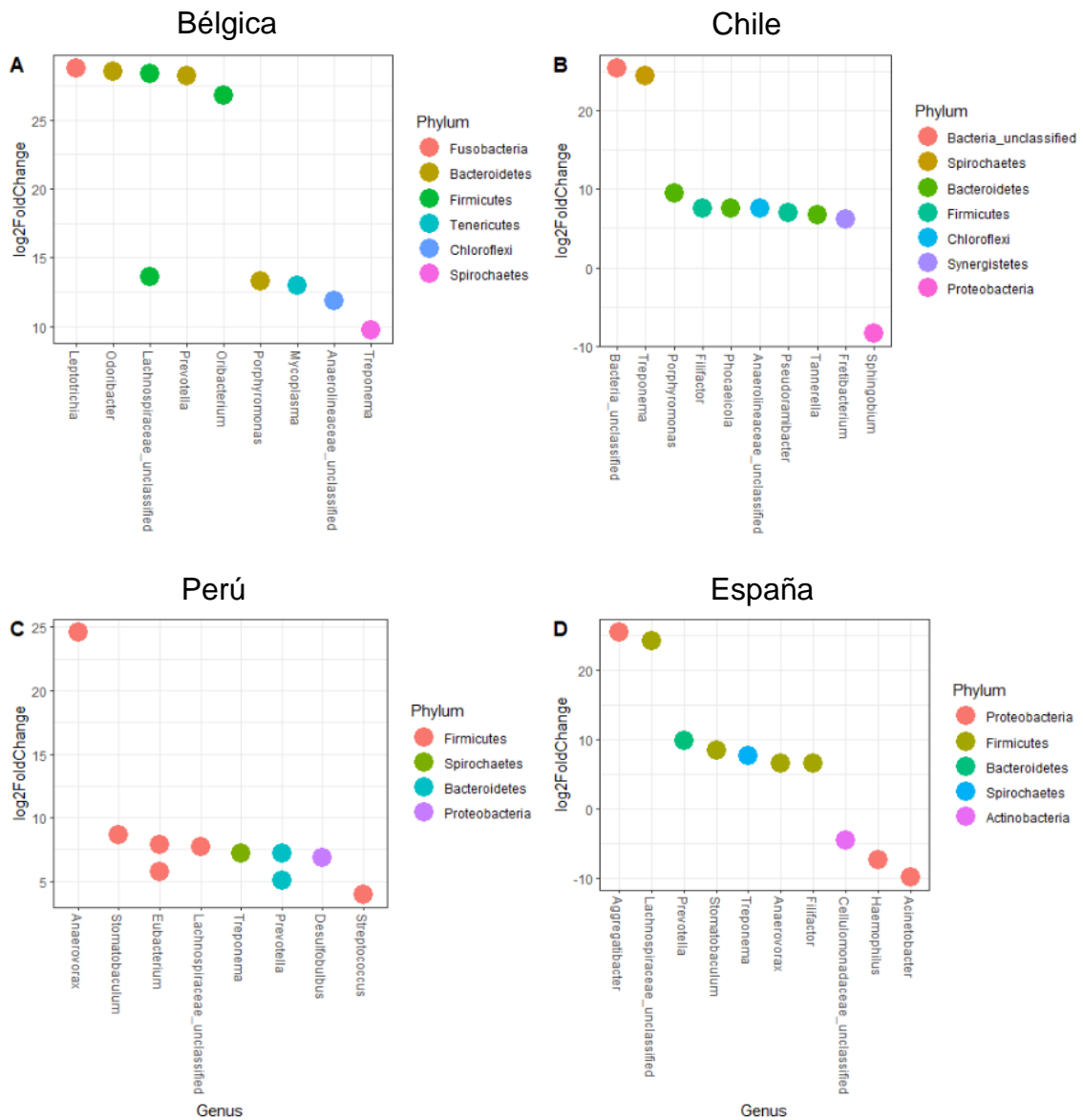
**Tabla 4.** Lista de los 10 OTUs asociados a géneros bacterianos estadísticamente más diferenciados entre sujetos sanos, comparando los distintos orígenes geográficos contemplados en este estudio y ordenados según su nivel de significancia estadística.

Sujetos sanos							
Chile vs Bélgica				España vs Bélgica			
Genus	Base Mean	Log 2 FC	P <sub>adj</sub>	Genus	Base Mean	Log 2 FC	P <sub>adj</sub>
<i>Alkalibacterium</i>	132,24	-10,08	4,79E-32	<i>Anaerobacillus</i>	19,94	8,05	5,35E-16
<i>Propioniceclava</i>	20,68	-24,98	9,64E-30	<i>Cellulomonadaceae</i> unclassified	135,76	6,06	9,63E-12
<i>Bacillaceae</i> unclassified	29,28	-7,90	1,72E-27	<i>Rhodobacteraceae</i> unclassified	11,14	6,70	9,63E-12
<i>Enterococcus</i>	52,06	-8,10	7,80E-19	<i>Sphingobium</i>	36,91	-8,18	1,51E-11
<i>Clostridiales</i> unclassified	22,34	-7,50	2,74E-18	<i>Lachnospiraceae</i> unclassified	42,82	9,54	2,42E-07
<i>Opitutus</i>	34,45	-8,13	5,11E-18	<i>Xanthomonas</i>	42,20	-5,49	7,42E-06
<i>Bacillales</i> unclassified	28,74	-7,87	1,15E-14	<i>Prevotella</i>	64,62	7,39	7,89E-06
<i>Xanthomonas</i>	55,16	-8,50	1,36E-14	<i>Acetoanaerobium</i>	6,47	-5,15	1,15E-05
<i>Tessaracoccus</i>	16,36	-7,05	1,36E-14	<i>Pseudomonas</i>	52,08	7,58	1,16E-04
<i>Gemella</i>	413,93	-4,98	1,37E-14	<i>Streptobacillus</i>	17,45	10,46	1,68E-04
Perú vs Bélgica				España vs Chile			
Genus	Base Mean	Log 2 FC	P <sub>adj</sub>	Genus	Base Mean	Log 2 FC	P <sub>adj</sub>
<i>Dysgonomonas</i>	80,28	10,24	3,58E-19	<i>Alkalibacterium</i>	259,60	11,21	3,75E-29
<i>Anaerobacillus</i>	33,15	8,87	7,37E-18	<i>Enterococcus</i>	127,61	9,60	2,43E-28
<i>Actinomycetales</i> unclassified	35,00	8,33	3,76E-13	<i>Sphingobium</i>	145,63	-10,68	5,98E-20
<i>Xanthomonas</i>	43,74	-8,21	5,58E-13	<i>Bacillaceae</i> unclassified	53,40	8,93	2,21E-18
<i>Alishewanella</i>	48,82	8,33	1,75E-12	<i>Bacillales</i> unclassified	49,66	8,82	4,12E-18
<i>Erysipelotrichaceae</i> unclassified	13,79	7,69	2,01E-12	<i>Cellulomonadaceae</i> unclassified	56,24	9,01	7,58E-18
<i>Flavobacterium</i>	18,50	7,40	4,52E-12	<i>Exiguobacterium</i>	112,94	9,63	2,01E-17
<i>Alishewanella</i>	11,13	7,09	7,30E-11	<i>Paracoccus</i>	44,46	7,94	1,24E-15
<i>Stenotrophomonas</i>	46,88	8,98	7,43E-11	<i>Gemella</i>	226,66	4,30	2,88E-13
<i>Propionibacteriaceae</i> unclassified	35,45	5,55	1,21E-10	<i>Erysipelotrichaceae</i> unclassified	21,11	7,58	1,15E-12
Perú vs Chile				Perú vs España			
Genus	Base Mean	Log 2 FC	P <sub>adj</sub>	Genus	Base Mean	Log 2 FC	P <sub>adj</sub>
<i>Alkalibacterium</i>	580,65	12,58	1,15E-35	<i>Pseudomonas</i>	42,94	11,27	3,79E-09
<i>Paracoccus</i>	278,28	10,78	1,88E-30	<i>Ralstonia</i>	9,06	-6,28	3,79E-09
<i>Bacillales</i> unclassified	138,57	10,52	4,82E-25	<i>Acinetobacter</i>	38,67	6,86	2,00E-06
<i>Enterococcus</i>	89,04	9,27	2,30E-24	<i>Actinomycetales</i> unclassified	36,23	-5,63	5,27E-06
<i>Bacillales</i> unclassified	87,19	9,85	7,18E-23	<i>Stenotrophomonas</i>	52,79	-6,73	3,48E-05
<i>Sphingobium</i>	100,70	-9,78	1,70E-20	<i>Alishewanella</i>	51,21	-4,59	5,87E-05
<i>Erysipelotrichaceae</i> unclassified	42,15	8,80	1,19E-19	<i>Bacillales</i> _unclassified	156,91	-4,18	8,47E-05
<i>Exiguobacterium</i>	48,30	8,59	5,20E-18	<i>Propionibacteriaceae</i> unclassified	34,21	-4,01	1,41E-04
<i>Dysgonomonas</i>	68,53	9,50	7,73E-18	<i>Parabacteroides</i>	8,19	-5,67	4,12E-04
<i>Rhodobacteraceae</i> unclassified	40,75	8,75	7,70E-17	<i>Lachnospiraceae</i> unclassified	6,87	-5,07	5,19E-04

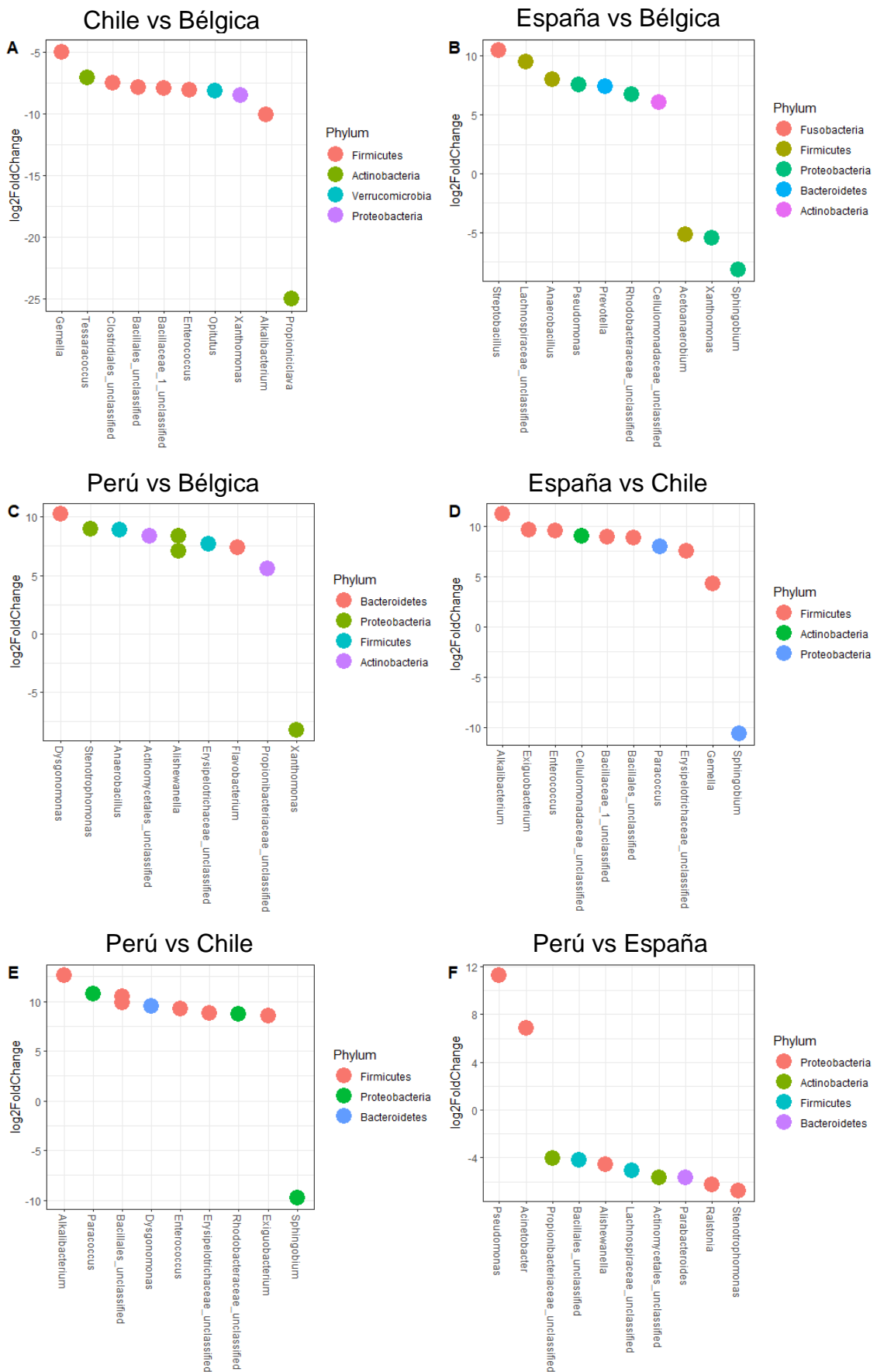
**Tabla 5.** Lista de los 10 OTUs asociados a géneros bacterianos estadísticamente más diferenciados entre pacientes con periodontitis, comparando los distintos orígenes geográficos contemplados en este estudio y ordenados según su nivel de significancia estadística.

Pacientes con periodontitis									
Chile vs Bélgica				España vs Bélgica					
Genus	Base Mean	Log 2 FC	P <sub>adj</sub>	Genus	Base Mean	Log 2 FC	P <sub>adj</sub>		
<i>Alkalibacterium</i>	4,28	-5,53	2,76E-11	<i>Paracoccus</i>	6,79	4,96	1,58E-08		
<i>Bacillaceae</i> unclassified	2,74	-4,31	1,11E-05	<i>Bacillales</i> unclassified	10,41	4,68	1,43E-06		
<i>Leptotrichia</i>	97,69	3,12	5,88E-03	<i>Haemophilus</i>	263,89	3,59	1,67E-04		
<i>Streptococcus</i>	451,55	2,54	7,52E-03	<i>Rothia</i>	32,63	2,54	7,06E-03		
<i>Streptococcus</i>	13,55	3,20	7,52E-03	<i>Puniceicoccaceae</i> unclassified	2,45	-4,92	7,06E-03		
				<i>Alkalibacterium</i>	36,10	2,10	7,45E-03		
				<i>Actinomyces</i>	8,87	-2,56	8,86E-03		
Perú vs Bélgica				España vs Chile					
Genus	Base Mean	Log 2 FC	P <sub>adj</sub>	Genus	Base Mean	Log 2 FC	P <sub>adj</sub>		
<i>Streptococcus</i>	895,89	3,52	3,78E-09	<i>Alkalibacterium</i>	27,87	8,10	6,21E-26		
<i>Staphylococcus</i>	70,79	9,33	3,78E-09	<i>Bacillales</i> unclassified	10,67	6,59	1,98E-09		
<i>Rothia</i>	65,94	3,91	1,96E-07	<i>Paracoccus</i>	6,38	5,96	2,11E-09		
<i>Haemophilus</i>	294,65	4,18	7,88E-07	<i>Bacillales</i> unclassified	3,25	4,87	1,41E-04		
<i>Neisseria</i>	133,32	4,64	7,88E-07	<i>Bacillaceae</i> unclassified	3,40	4,54	1,37E-03		
<i>Bulleidia</i>	12,12	-5,86	1,72E-06	<i>Erysipelotrichaceae</i> unclassified	2,98	4,86	1,61E-03		
<i>Streptococcus</i>	3677,14	3,30	3,60E-06						
<i>Treponema</i>	11,02	10,40	9,69E-06						
<i>Aggregatibacter</i>	142,20	5,21	2,11E-05						
<i>Fretibacterium</i>	501,12	-2,59	3,14E-05						
Perú vs Chile				Perú vs España					
Genus	Base Mean	Log 2 FC	P <sub>adj</sub>	Genus	Base Mean	Log 2 FC	P <sub>adj</sub>		
<i>Staphylococcus</i>	88,13	7,50	9,36E-06	<i>Staphylococcus</i>	213,23	-9,76	1,82E-09		
<i>Granulicatella</i>	27,17	5,86	1,01E-05	<i>Porphyromonas</i>	14,02	10,56	1,27E-03		
<i>Neisseria</i>	438,71	3,59	1,33E-04	<i>Leptotrichia</i>	13,03	-6,55	1,27E-03		
<i>Porphyromonas</i>	87,09	3,84	1,33E-04	<i>Prevotella</i>	41,67	6,53	3,90E-03		
<i>Gemella</i>	94,24	3,27	1,62E-04	<i>Treponema</i>	36,29	-7,07	5,91E-03		
<i>Streptococcus</i>	14,17	2,87	1,22E-03	<i>Lachnospiraceae</i> unclassified	17,85	-4,51	9,47E-03		
<i>Granulicatella</i>	96,55	2,59	5,72E-03	<i>Bacteroidetes</i> unclassified	8,95	5,94	9,47E-03		

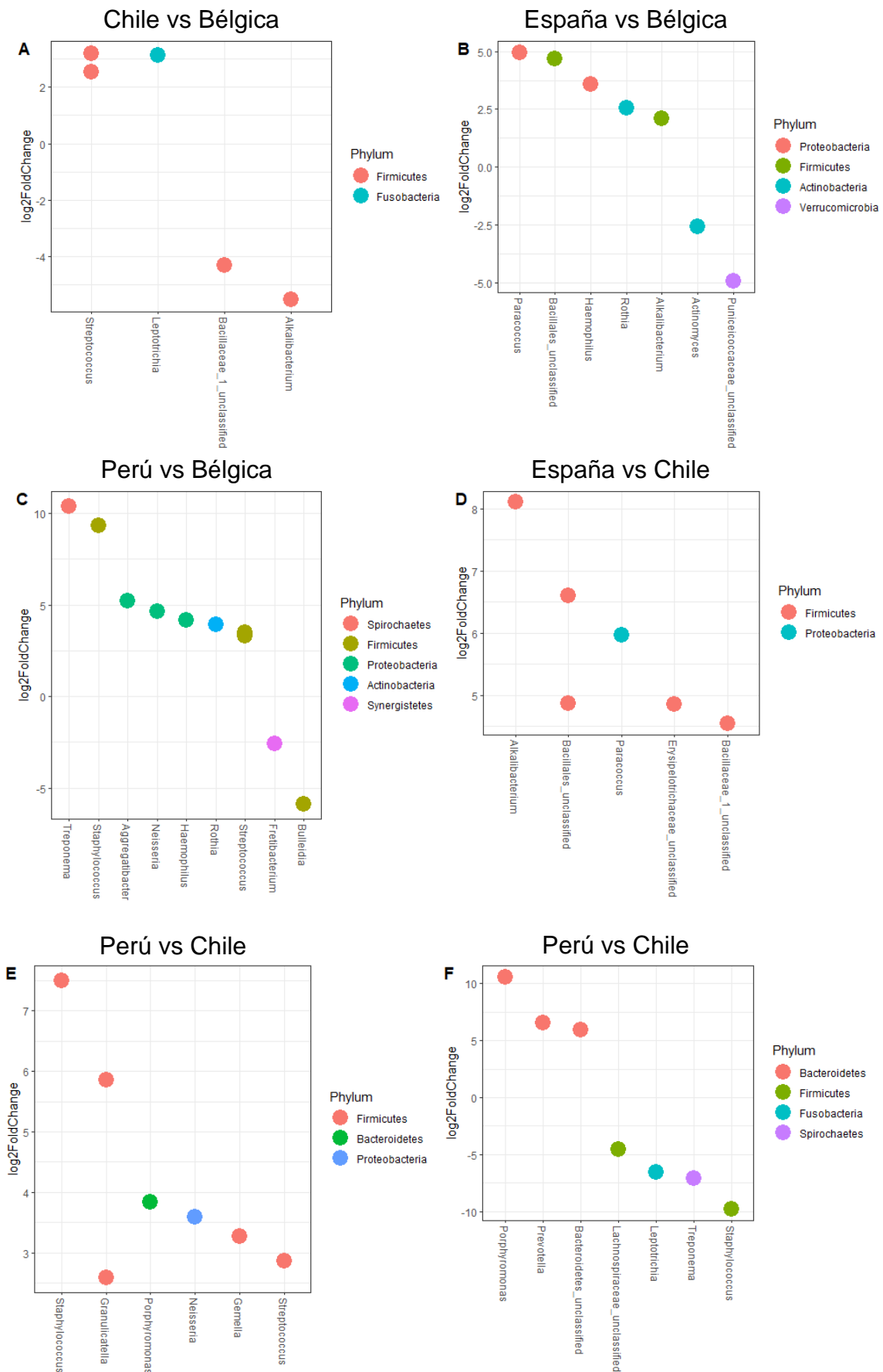




**Figura 8.** Gráfico con los 10 OTUs asociados a géneros y filos bacterianos estadísticamente más diferenciados entre pacientes con periodontitis y sujetos sanos, según el origen geográfico de las muestras y ordenados según su nivel de significancia estadística.



**Figura 9.** Gráfico con los 10 OTUs asociados a géneros y fillos bacterianos estadísticamente más diferenciados entre sujetos sanos, según el origen geográfico de las muestras y ordenados según su nivel de significancia estadística.



**Figura 10.** Gráfico con los 10 OTUs asociados a géneros y filos bacterianos estadísticamente más diferenciados entre pacientes con periodontitis, según el origen geográfico de las muestras y ordenados según su nivel de significancia estadística.

## 4. Discusión

El estudio de las comunidades microbianas que colonizan nuestro cuerpo puede dar lugar a estrategias para su manipulación y control en distintas condiciones de salud. La irrupción de la tecnología de secuenciación masiva permitió al campo de la ecología microbiana describir con un mayor alcance la diversidad presente en diferentes tipos de muestras. En el caso del microbioma oral, estos estudios se han realizado a partir de tamaños muestrales relativamente pequeños y con ciertas restricciones geográficas, proviniendo las muestras de zonas geográficas muy limitadas<sup>18-21</sup>. El aún todavía elevado coste de realizar estos estudios implica que no todos los centros de investigación pueden permitírselos y por lo tanto se crea un sesgo importante, al añadirse gran parte del conocimiento del microbioma oral procedente de muestras de países o regiones más ricas o con capacidad de realizar dichos análisis. Con tal de comprobar si este sesgo tiene una repercusión en cuanto al conocimiento del microbioma subgingival “estándar” asociado a salud periodontal o a periodontitis, en este estudio se analizó la composición de la microbiota subgingival de individuos sanos y de pacientes con periodontitis con distintos orígenes geográficos (Perú, Chile, Bélgica y España) mediante la secuenciación masiva del gen 16S de las bacterias que habitaban el surco subgingival de 240 individuos.

Los resultados de este estudio indican que existen diferencias significativas entre la microbiota subgingival de individuos con distinto origen geográfico. En cuanto a la riqueza de especies, se observó que la microbiota de individuos peruanos era la más rica, aunque presentaba unos niveles de diversidad muy similares a la de los otros países, indicando que, pese a que existía un mayor número de especies, su distribución era equitativa<sup>29,30</sup>. Sin embargo, no se observaron diferencias significativas al analizar la riqueza entre muestras sanas y muestras de periodontitis. Al comparar la diversidad entre pacientes con periodontitis y sujetos sanos sí que se detectaron diferencias significativas, observándose una mayor diversidad en las muestras de periodontitis. Al analizar estas diferencias país por país, únicamente se pudo detectar un bajo nivel de significación estadística en las muestras belgas. Estos datos sugieren

que los surcos subgingivales de pacientes con periodontitis tienen un número similar de especies bacterianas que el de sujetos sanos, aunque dichas especies se distribuyen de manera ligeramente más equitativa lo cual concuerda con lo descrito previamente<sup>31</sup>.

Al analizar los géneros más representados en las muestras, se observó el patrón previamente descrito en el que comensales como *Veillonella*, *Streptococcus*, *Neisseria* o *Rothia* están menos presentes en las muestras de periodontitis en favor de otros géneros como *Porphyromonas*, *Fusobacterium* o *Filifactor*. Estos cambios se dieron, aunque a diferente escala, en los 4 países analizados. Los resultados de este estudio podrían confirmar la capacidad del género *Fretibacterium* de ser un indicador de la enfermedad periodontal, tal y como ha sido sugerido recientemente<sup>32</sup>.

Los gráficos NMDs de distancias Bray-Curtis revelaron que las muestras se agrupaban según su diagnóstico, aunque existe una ligera contribución del origen geográfico en la composición microbiana. Esto era de esperar debido a la fuerte correlación entre la disbiosis del biofilm subgingival y la aparición de la enfermedad periodontal<sup>33</sup>.

Las diferencias de la composición de la microbiota entre las distintas variables también fue analizada, observándose numerosas diferencias significativas. En todos los países se detectaron géneros sobrerrepresentados y subrepresentados de acuerdo al diagnóstico. Por ejemplo, el género *Treponema* estuvo sobrerrepresentado en todas las muestras de pacientes con periodontitis, especialmente en las muestras chilenas, y los géneros *Porphyromonas* y *Prevotella* estaban sobrerrepresentados en la mayoría de los países. Estos géneros han sido previamente asociados a la enfermedad periodontal<sup>3</sup> y por lo tanto su mayor prevalencia en muestras de pacientes con periodontitis era de esperar. El género *Aggregatibacter*, al cual pertenece *Aggregatibacter actinomycetemcomitans*, fue uno de los que más diferenciación mostró en las muestras españolas, lo cual es de destacar teniendo en cuenta la asociación que se ha descrito de este microorganismo con la anteriormente clasificada como periodontitis agresiva<sup>34</sup>, hoy periodontitis estadio III o IV y grado C<sup>22,35</sup>. Por otro lado, al analizar las muestras de un mismo cuadro clínico pero de origen geográfico distinto, se observaron numerosas especies diferentes, especialmente en el grupo de individuos

sanos. Por ejemplo, las diferencias más significativas entre Chile y Bélgica implicaban en las muestras chilenas una reducción en la proporción de secuencias pertenecientes a géneros extensivamente estudiados que se pueden aislar del microbioma oral como *Gemella*, *Enterococcus* y bacterias del orden *Clostridiales*<sup>29,36,37</sup>. Por otro lado, las muestras peruanas mostraron un aumento en la proporción de las especies diferencialmente más significativas al compararlo con Bélgica. De especial interés resultó ser el aumento de proporción de bacterias del género *Prevotella* en muestras españolas respecto a las belgas, puesto que las bacterias de este género han sido asociadas a la enfermedad periodontal y dos de sus especies (*Prevotella nigrescens* y *Prevotella intermedia*) forman parte del complejo naranja de la pirámide de Socransky y Haffajje<sup>3</sup>. Diferencias similares a las mencionadas se observaron en todas las comparaciones entre los países, reforzando la idea de que la microbiota subgingival de individuos sanos está sometida a una alta variabilidad si se tiene en cuenta el origen geográfico como criterio. Las diferencias fueron menos numerosas al analizar la microbiota de pacientes con periodontitis entre países, encontrándose en algunos casos menos de 10 OTUs por comparación. Además, cuantitativamente los cambios observados fueron menores que los detectados en las muestras de sujetos sanos. Aún así cabe destacar la mayor proporción en muestras peruanas de géneros con conocidos periodontopatógenos como *Treponema*, *Porphyromonas*, *Prevotella* y *Aggregatibacter*<sup>30,38</sup> respecto a las muestras de los otros países. Por otro lado también se observaron diferencias significativas en géneros asociados a la microbiota comensal como *Rothia*, *Haemophilus*, *Neisseria*, *Streptococcus* y *Gemella* entre otros<sup>39</sup>. Las distintas proporciones de estos géneros bacterianos podrían indicar interacciones entre microorganismos que aún no han sido descritas y que podrían suponer un avance en el conocimiento de las dinámicas poblacionales que dan lugar a la disbiosis del biofilm subgingival. Por otro lado, y pese a que los periodontopatógenos clásicos parecen estar presentes en mayor medida en estas muestras, se encontraron suficientes diferencias significativas como para afirmar que la microbiota subgingival de pacientes con periodontitis es variable según su origen geográfico.

Las limitaciones de este estudio son varias: en primer lugar el número de localizaciones geográficas es escaso y limita el alcance de las conclusiones. La

obtención de muestras periodontales por parte de profesionales no es una tarea sencilla, lo cual limita el acceso a las mismas. Además, ampliar el número de continentes incluídos en el estudio reforzaría su calidad. Del mismo modo, tener en cuenta variables socioeconómicas de los sujetos, sus hábitos alimenticios<sup>40</sup>, así como parámetros clínicos exhaustivos complementarían la información arrojada por la secuenciación masiva. Dicho esto, las diferencias observadas en la distribución de la microbiota subgingival de pacientes con periodontitis y de sujetos sanos de 4 países distintos abre la puerta a realizar análisis más exhaustivos que incluyan más variables y orígenes geográficos con el objetivo de profundizar en la hipótesis iniciada en este estudio.

## 5. Conclusiones

La conclusión de este trabajo es que efectivamente existen diferencias significativas en la composición de la microbiota subgingival de los grupos estudiados, especialmente entre la microbiota asociada a salud periodontal, por lo que habría que ser cautelosos a la hora de “universalizar” modelos de poblaciones bacterianas asociadas a ciertos diagnósticos. Los resultados obtenidos abren la puerta a realizar más estudios para analizar las causas de estas diferencias, así como las consecuencias que se puedan derivar de ellas. Pueden ciertas comunidades asociadas a salud periodontal prevenir mejor la disbiosis? Hay poblaciones geográficamente localizadas más predispuestas a la colonización por periodontopatógenos? Estas son algunas de las preguntas que pueden derivar de este estudio y que requerirían de una mayor profundización en las variables clínicas y socioeconómicas de los sujetos muestreados, así como incrementar el número de localizaciones geográficas para aumentar la validez de las conclusiones obtenidas. Estos aspectos han sido probablemente las mayores limitaciones del estudio, puesto que el bagaje socioeconómico está asociado a la higiene dental y por lo tanto a la composición del biofilm subgingival. Además, si bien encontrar diferencias en muestras de 4 países ya puede constituir en sí mismo un descubrimiento a reportar, aportar más muestras de otros países que incluyan otros continentes o subcontinentes habría aumentado la calidad del estudio. Por otro lado, utilizar otras técnicas tales como el shotgun sequencing, permitiría realizar este análisis con una mayor profundidad, sin las limitaciones que implican focalizarse en el gen 16S rRNA.

A lo largo del estudio se han cumplido todos los objetivos marcados dentro del plazo fijado, con la ligera excepción del análisis de las secuencias, cuyo retraso vino dado por la crisis generada por el SARS-CoV-2 y la imposibilidad de acceder a las mismas. Sin embargo se pudo compensar el retraso adelantando otros objetivos del TFM. Los conocimientos adquiridos a lo largo de la realización de este TFM me han permitido conocer y aplicar de primera mano software abierto para el análisis de secuencias obtenidas a partir de



secuenciadores masivos, así como su posterior análisis estadístico mediante flujos de trabajo estandarizados en esta área de conocimiento.

## **6. Glosario**

**DNA:** Ácido desoxiribonucleico

**EDTA:** Ácido etilendiaminotetraacético

**FTR:** Fluido de transporte reducido

**NMDS:** escalamiento multidimensional no métrico

**OTU:** Unidad taxonómica operativa

**PCR:** Reacción en cadena de la polimerasa

**rRNA:** Ácido ribonucleico ribosomal

## 7. Bibliografía

1. Manna, A., *et al.* Supragingival plaque microbial analysis in reflection to caries experience. *BMC Oral Health* **13**, 5 (2013). doi: 10.1186/1472-6831-13-5
2. Kolenbrander, P.E. *et al.* Oral multispecies biofilm development and the key role of cell–cell distance. *Nat. Rev. Microbiol.* **8**, 471–480 (2010). doi: 10.1038/nrmicro2381
3. Socransky, S. S. & Haffajee, A. D. Dental biofilms: difficult therapeutic targets. *Periodontol. 2000* **28**, 12–55 (2002). doi: 10.1034/j.1600-0757.2002.280102.x
4. Hannig, M. & Joiner, A. The Structure, Function and Properties of the Acquired Pellicle in *The Teeth and Their Environment* vol 19, pages 29–64 (2005). doi:10.1159/000090585.
5. Zijng, V. *et al.* Oral biofilm architecture on natural teeth. *PLoS One* **5**, e9321 (2010). doi: 10.1371/journal.pone.0009321
6. Kolenbrander, P.E. *et al.* Coaggregation of *Fusobacterium nucleatum*, *Selenomonas flueggei*, *Selenomonas infelix*, *Selenomonas noxia*, and *Selenomonas sputigena* with strains from 11 genera of oral bacteria. *Infect. Immun.* **57**, 3194–203 (1989).
7. Mark Welch, J. L., *et al.* Biogeography of a human oral microbiome at the micron scale. *Proc. Natl. Acad. Sci.* **113**, E791–E800 (2016). doi: 10.1073/pnas.1522149113
8. Socransky, S. S., *et al.* Microbial complexes in subgingival plaque. *J Clin Periodontol* **25**, 134–144 (1998). doi: 10.1111/j.1600-051x.1998.tb02419.x
9. Jakubovics, N. S. & Kolenbrander, P. E. The road to ruin: the formation of disease-associated oral biofilms. *Oral Dis* **16**, 729–739 (2010). doi: 10.1111/j.1601-0825.2010.01701.x
10. Chace, R. & Low, S. B. Survival Characteristics of Periodontally-Involved Teeth: A 40-Year Study. *J. Periodontol.* **64**, 701–705 (1993). doi: 10.1902/jop.1993.64.8.701
11. McLeod, D. E., *et al.* Tooth loss due to periodontal abscess: a retrospective study. *J. Periodontol.* **68**, 963–6 (1997). doi: 10.1902/jop.1997.68.10.963
12. Armitage, G. C. *et al.* Position Paper: Diagnosis of Periodontal Diseases. *J. Periodontol.* **74**, 1237–1247 (2003). doi: 10.1902/jop.2003.74.8.1237
13. Lang, N. P., *et al.* Monitoring disease during supportive periodontal

- treatment by bleeding on probing. *Periodontol. 2000* **12**, 44–8 (1996). doi: 10.1111/j.1600-0757.1996.tb00080.x
14. Greenstein, G. Contemporary Interpretation of Probing Depth Assessments: Diagnostic and Therapeutic Implications. A Literature Review. *J. Periodontol.* **68**, 1194–1205 (1997). doi: 10.1902/jop.1997.68.12.1194
  15. Haffajee, A. D., *et al.* The use of DNA probes to examine the distribution of subgingival species in subjects with different levels of periodontal destruction. *J. Clin. Periodontol.* **19**, 84–91 (1992). doi: 10.1111/j.1600-051x.1992.tb00445.x
  16. Tay, F., *et al.* Evaluation of a non-radioactive DNA probe for detecting *Porphyromonas gingivalis* in subgingival specimens. *Oral Microbiol. Immunol.* **7**, 344–8 (1992). doi: 10.1111/j.1399-302x.1992.tb00634.x
  17. Söder, P. O., *et al.* DNA probe detection of periodontopathogens in advanced periodontitis. *Scand. J. Dent. Res.* **101**, 363–70 (1993). doi: 10.1111/j.1600-0722.1993.tb01133.x
  18. Belda-Ferre, P. *et al.* The oral metagenome in health and disease. *ISME J.* **6**, 46–56 (2012). doi: 10.1038/ismej.2011.85
  19. Bostanci, N., *et al.* Periodontal disease: From the lenses of light microscopy to the specs of proteomics and next-generation sequencing in *Advances in Clinical Chemistry*, volume 93, pages 263–290 (2019). doi:10.1016/bs.acc.2019.07.006.
  20. Griffen, A. L. *et al.* Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J.* **6**, 1176–1185 (2012). doi: 10.1038/ismej.2011.191
  21. Kirst, M. E. *et al.* Dysbiosis and Alterations in Predicted Functions of the Subgingival Microbiome in Chronic Periodontitis. *Appl. Environ. Microbiol.* **81**, 783–793 (2015). doi: 10.1128/AEM.02712-14
  22. Caton, J. G. *et al.* A new classification scheme for periodontal and peri-implant diseases and conditions - Introduction and key changes from the 1999 classification. *J. Periodontol.* **89**, S1–S8 (2018). doi: 10.1002/JPER.18-0157
  23. Loesche, W. J., *et al.* Trypsin-Like Activity in Subgingival Plaque. *J. Periodontol.* **58**, 266–273 (1987). doi: 10.1902/jop.1987.58.4.266
  24. Kozich, J. J., *et al.* Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).

doi: 10.1128/AEM.01043-13

25. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012). doi: 10.1093/nar/gks1219
26. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217 (2013). doi: 10.1371/journal.pone.0061217
27. Oksanen, J., *et al.* vegan: Community Ecology Package. <https://cran.r-project.org/package=vegan> (2016).
28. Love, M. I., *et al.* Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8
29. Aas, J. A., *et al.* Defining the normal bacterial flora of the oral cavity. *J. Clin. Microbiol.* **43**, 5721–5732 (2005). doi: 10.1128/JCM.43.11.5721-5732.2005
30. Marsh, P. D., *et al.* Dental plaque biofilms: communities, conflict and control. *Periodontol. 2000* **55**, 16–35 (2011). doi: 10.1111/j.1600-0757.2009.00339.x
31. Abusleme, L. *et al.* The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *ISME J.* **7**, 1016–1025 (2013). doi: 10.1038/ismej.2012.174
32. Na, H. S. *et al.* Identification of Potential Oral Microbial Biomarkers for the Diagnosis of Periodontitis. *J. Clin. Med.* **9**, 1549 (2020). doi: 10.3390/jcm9051549
33. Hajishengallis, G. & Lamont, R. J. Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. *Mol. Oral Microbiol.* **27**, 409–419 (2012). doi: 10.1111/j.2041-1014.2012.00663.x
34. Albandar, J. M., *et al.* Associations Between Serum Antibody Levels to Periodontal Pathogens and Early-Onset Periodontitis. *J. Periodontol.* **72**, 1463–1469 (2001). doi: 10.1902/jop.2001.72.11.1463
35. Armitage, G. C. Development of a Classification System for Periodontal Diseases and Conditions. *Ann. Periodontol.* **4**, 1–6 (1999). doi: 10.1902/annals.1999.4.1.1
36. Keijsers, B. J. F. *et al.* Pyrosequencing analysis of the Oral Microflora of healthy adults. *J. Dent. Res.* **87**, 1016–1020 (2008). doi: 10.1177/154405910808701104

37. Souto, R. & Colombo, A. P. V. Prevalence of *Enterococcus faecalis* in subgingival biofilm and saliva of subjects with chronic periodontal infection. *Arch. Oral Biol.* **53**, 155–160 (2008). doi: 10.1016/j.archoralbio.2007.08.004
38. Jakubovics, N. S. Saliva as the Sole Nutritional Source in the Development of Multispecies Communities in Dental Plaque. *Microbiol. Spectr.* **3**, (2015). doi: 10.1128/microbiolspec.MBP-0013-2014
39. Heller, D. *et al.* Microbial Diversity in the Early In Vivo -Formed Dental Biofilm. *Appl. Environ. Microbiol.* **82**, 1881–1888 (2016). doi: 10.1128/AEM.03984-15
40. Jia, G. *et al.* The oral microbiota - a mechanistic role for systemic diseases. *Br Dent J.* **224**, 447-455 (2018). doi:10.1038/sj.bdj.2018.217

## 8. Anexos

### 8.1 Anexo 1 – Protocolo Mothur

1. Generamos un índice de los archivos .fastq:

```
mothur > make.file(inputdir=Dataset, type=fastq, prefix=stability)
```

2. Creamos los contigs, especificando el número de procesadores con los que cuenta nuestro ordenador (una vez establecidos, mothur lo recordará y usará el mismo número a no ser que le indiquemos lo contrario).

```
mothur > make.contigs(file=Dataset\stability.files, processors=12)
```

3. Comprobamos las secuencias:

```
mothur > summary.seqs(fasta=Dataset\stability.trim.contigs.fasta)
```

4. Filtramos las secuencias según su tamaño (en este caso un máximo de 470 pares de bases) y el número de ambigüedades (en este caso 0).

```
mothur > screen.seqs(fasta=Dataset\stability.trim.contigs.fasta,  
group=Dataset\stability.contigs.groups, maxambig=0, maxlength=470)
```

5. A continuación volvemos a comprobar las secuencias.

```
mothur > summary.seqs()
```

6. A continuación eliminamos duplicados, rebajando aún más el número de secuencias y aligerando trabajo computacional:

```
mothur > unique.seqs(fasta=Dataset\stability.trim.contigs.good.fasta)
```

7. Generamos una tabla que indicará cuantas veces se ve representada cada secuencia:

```
mothur > count.seqs(name=Dataset\stability.trim.contigs.good.names,  
group=Dataset\stability.contigs.good.groups)
```

8. Ahora tenemos que realizar el alineamiento con la base de datos elegida, en este caso SILVADB. Sabemos que la posición de los V3 y los V4 en la base de datos está entre las posiciones 3000 y 27000, por lo que preparamos la base de datos para utilizar únicamente esa zona:

```
mothur > pcr.seqs(fasta=silva.v3v4.fasta, start=3000, end=27000, keepd  
ots=F, processors=8)
```

Y ahora alineamos con la base de datos reducida:

```
mothur >  
align.seqs(fasta=Dataset\stability.trim.contigs.good.unique.fasta,  
reference=silva.v3v4.pcr.fasta)
```

**9. Comprobamos como ha ido el alineamiento:**

```
mothur > summary.seqs()
```

**10. Seleccionamos las secuencias que han alineado en las posiciones que nos interesan:**

```
mothur >
screen.seqs(fasta=Dataset\stability.trim.contigs.good.unique.align,
count=Dataset\stability.trim.contigs.good.count_table,
summary=Dataset\stability.trim.contigs.good.unique.summary,
start=3388, end=22316, maxhomop=8)
```

**11. Eliminamos los sobrantes de las secuencias y aquellas columnas que contengan gaps:**

```
mothur >
filter.seqs(fasta=Dataset\stability.trim.contigs.good.unique.good.alig
n, vertical=T, trump=.)
```

**12. Eliminamos redundancias:**

```
mothur >
unique.seqs(fasta=Dataset\stability.trim.contigs.good.unique.good.filt
er.fasta, count=Dataset\stability.trim.contigs.good.good.count_table)
```

**13. Agrupamos las secuencias con la función pre-cluster, la cual separará las secuencias por grupos y las ordenará de más a menos abundantes y caracterizándolas según un cut-off de 4 nucleótidos de diferencia, de acuerdo con la recomendación de los desarrolladores de mothur consistente en aplicar un cut-off de 1 nucleótido por cada 100 pares de bases.**

```
mothur >
pre.cluster(fasta=Dataset\stability.trim.contigs.good.unique.good.filt
er.unique.fasta,
count=Dataset\stability.trim.contigs.good.unique.good.filter.count_tab
le, diffs=4)
```

**14. Buscamos las quimeras:**

```
mothur >
chimera.vsearch(fasta=Dataset\stability.trim.contigs.good.unique.good.
filter.unique.precluster.fasta,
count=Dataset\stability.trim.contigs.good.unique.good.filter.unique.pr
ecluster.count_table, dereplicate=t)
```

**15. Y las eliminamos del archivo .fasta:**

```
mothur >
remove.seqs(fasta=Dataset\stability.trim.contigs.good.unique.good.filt
er.unique.precluster.fasta,
accnos=Dataset\stability.trim.contigs.good.unique.good.filter.unique.p
recluster.denovo.vsearch.accnos)
```

16. Es posible que nuestra muestra incluya DNA humano, vírico, fúngico o de otros organismos que no resultan de interés para nuestro estudio, basado en la diversidad bacteriana. Para ello eliminamos todas aquellas secuencias que el programa nos identifique como no-bacterianas:

```
mothur >
classify.seqs(fasta=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta,
count=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.count_table,
reference=trainset16_022016.pds.fasta,
taxonomy=trainset16_022016.pds.tax, cutoff=80)
```

```
mothur >
remove.lineage(fasta=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta,
count=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.count_table,
taxonomy=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-Archaea-Eukaryota)
```

17. Ahora agruparemos las secuencias en OTUs. Mediante el argumento `taxlevel` podemos seleccionar hasta donde queremos ahondar en el análisis taxonómico. Como por defecto ya está al máximo (nivel de género), no añadimos el argumento.

```
mothur >
dist.seqs(fasta=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta, cutoff=0.03)
```

18. Agrupamos los OTUs en clústeres:

```
mothur >
cluster(column=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.dist,
count=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.count_table)
```

19. Contamos cuantas secuencias hay en cada OTU

```
mothur >
make.shared(list=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.opti_mcc.list,
count=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.count_table, label=0.03)
```

20. Finalmente asociamos la taxonomía a los OTUs:

```
mothur >
classify.otu(list=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.opti_mcc.list,
count=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.count_table,
taxonomy=Dataset\stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy, label=0.03)
```



Con esta lista de OTUs que incluye el número de secuencias y el género bacteriano al cual pertenecen podemos iniciar nuestro análisis estadístico.

## 8.2 Anexo 2 – Código R para el análisis estadístico

```
library(ggplot2)
library(vegan)
library(dplyr)
library(scales)
library(grid)
library(reshape2)
library(cowplot)
library(phyloseq)
library(DESeq2)
library(colorspace)
library(viridis)

# Importar Los archivos generados en mothur a phyloseq y realizar algunos
cambios:
setwd("./Resultados Mothur")
theme_set(theme_bw())

taxfile <- "./dades/taxonomy2.taxonomy"
sharedfile <- "./dades/shared2.shared"
mapfile <- "./dades/metadata2.metadata"

mothur_data <- import_mothur(mothur_shared_file = sharedfile,
mothur_constaxonomy_file = taxfile)
map <- read.delim(mapfile)
map <- sample_data(map)
rownames(map) <- map$group
mothur_data <- merge_phyloseq(mothur_data, map)
colnames(tax_table(mothur_data)) <- c("Kingdom", "Phylum", "Class", "Order",
"Family", "Genus")

# Análisis de diversidad y riqueza:

set.seed(1)
scaled <- rarefy_even_depth(mothur_data, sample.size=2400, replace=FALSE,
rngseed=1)
barplotgeneres <- sort(taxa_sums(scaled), TRUE)[1:10]
barplotgeneres2 <- prune_taxa(names(barplotgeneres), scaled)

alpha_diversity <- estimate_richness(scaled, split = TRUE, measures = NULL)

pairwise.wilcox.test(alpha_diversity$Observed, sample_data(scaled)$country)
pairwise.wilcox.test(alpha_diversity$Chao1, sample_data(scaled)$country)
pairwise.wilcox.test(alpha_diversity$Shannon, sample_data(scaled)$country)
pairwise.wilcox.test(alpha_diversity$Simpson, sample_data(scaled)$country)

pairwise.wilcox.test(alpha_diversity$Observed, sample_data(scaled)$diagnose)
pairwise.wilcox.test(alpha_diversity$Chao1, sample_data(scaled)$diagnose)
pairwise.wilcox.test(alpha_diversity$Shannon, sample_data(scaled)$diagnose)
pairwise.wilcox.test(alpha_diversity$Simpson, sample_data(scaled)$diagnose)
```

```

Peru <- subset_samples(scaled, country == "Peru")
alpha_diversity_Peru <- estimate_richness(Peru, split = TRUE, measures =
NULL)
Belgium <- subset_samples(scaled, country == "Belgium")
alpha_diversity_Belgium <- estimate_richness(Belgium, split = TRUE, measures
= NULL)
Chile <- subset_samples(scaled, country == "Chile")
alpha_diversity_Chile <- estimate_richness(Chile, split = TRUE, measures =
NULL)
Spain <- subset_samples(scaled, country == "Spain")
alpha_diversity_Spain <- estimate_richness(Spain, split = TRUE, measures =
NULL)

pairwise.wilcox.test(alpha_diversity_Peru$Observed,
sample_data(Peru)$diagnose)
pairwise.wilcox.test(alpha_diversity_Peru$Chao1, sample_data(Peru)$diagnose)
pairwise.wilcox.test(alpha_diversity_Peru$Shannon,
sample_data(Peru)$diagnose)
pairwise.wilcox.test(alpha_diversity_Peru$Simpson,
sample_data(Peru)$diagnose)

pairwise.wilcox.test(alpha_diversity_Belgium$Observed,
sample_data(Belgium)$diagnose)
pairwise.wilcox.test(alpha_diversity_Belgium$Chao1,
sample_data(Belgium)$diagnose)
pairwise.wilcox.test(alpha_diversity_Belgium$Shannon,
sample_data(Belgium)$diagnose)
pairwise.wilcox.test(alpha_diversity_Belgium$Simpson,
sample_data(Belgium)$diagnose)

pairwise.wilcox.test(alpha_diversity_Chile$Observed,
sample_data(Chile)$diagnose)
pairwise.wilcox.test(alpha_diversity_Chile$Chao1,
sample_data(Chile)$diagnose)
pairwise.wilcox.test(alpha_diversity_Chile$Shannon,
sample_data(Chile)$diagnose)
pairwise.wilcox.test(alpha_diversity_Chile$Simpson,
sample_data(Chile)$diagnose)

pairwise.wilcox.test(alpha_diversity_Spain$Observed,
sample_data(Spain)$diagnose)
pairwise.wilcox.test(alpha_diversity_Spain$Chao1,
sample_data(Spain)$diagnose)
pairwise.wilcox.test(alpha_diversity_Spain$Shannon,
sample_data(Spain)$diagnose)
pairwise.wilcox.test(alpha_diversity_Spain$Simpson,
sample_data(Spain)$diagnose)

# Boxplots de diversidad y riqueza (figura 5):
box_country <- plot_richness(scaled, "country", measures=c("Observed",
"Chao1", "Shannon", "Simpson"), color="country")
box_country + geom_boxplot(data=box_country$data, aes(x=country, color=NULL,
fill=country)) +
  theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

box_diagnose <- plot_richness(scaled, "diagnose", measures=c("Observed",
"Chao1", "Shannon", "Simpson"), color="diagnose")
box_diagnose + geom_boxplot(data=box_diagnose$data, aes(x=diagnose,

```

```

color=NULL, fill=diagnose)) +
  theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

box_diagnose_country <- plot_richness(scaled, "country", "diagnose",
measures=c("Observed", "Chao1", "Shannon", "Simpson"), color="diagnose")
box_diagnose_country + geom_boxplot(data=box_diagnose_country$data,
aes(x=country, color=NULL, fill=diagnose)) +
  theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

# Gráfico de diversidad bacteriana (figura 6):
plot_bar(barplotgeneres2, x="diagnose", fill="Genus", facet_grid=~country) +
  scale_fill_manual(values = viridis(9)) +
  geom_bar(stat="identity") +
  theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

# Gráficos NMDS de distancias Bray-Curtis (figura 7):
mothur_otus_nmnds <- ordinate(scaled, "NMDS", "bray")

variable_country <- get_variable(scaled, "country")
variable_diagnose <- get_variable(scaled, "diagnose")

anosim_country <- anosim(phyloseq::distance(scaled, "bray"),
variable_country)
anosim_diagnose <- anosim(phyloseq::distance(scaled, "bray"),
variable_diagnose)
df_adonis <- as(sample_data(scaled), "data.frame")
resultat_adonis <- adonis((phyloseq::distance(scaled, "bray")) ~ country +
diagnose, df_adonis)
plot(resultat_adonis$aov.tab)

plot_ordination(scaled, mothur_otus_nmnds, type = "taxa", color = "Phylum",
title = "NMDS ordination - OTUs")
plot_ordination(scaled, mothur_otus_nmnds, type = "taxa", color = "Phylum",
title = "NMDS ordination - OTUs") +
  facet_wrap(~Phylum, 5) +
  theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=12))

plot_ordination(scaled, mothur_otus_nmnds, type = "samples", color =
"country", title = "NMDS ordination - OTUs")
plot_ordination(scaled, mothur_otus_nmnds, type = "samples", color =
"country", title = "NMDS ordination - OTUs") +
  facet_wrap(~country, 5) +
  theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

plot_ordination(scaled, mothur_otus_nmnds, type = "samples", color =
"diagnose", title = "NMDS ordination - OTUs")
plot_ordination(scaled, mothur_otus_nmnds, type = "samples", color =
"diagnose", title = "NMDS ordination - OTUs") +
  facet_wrap(~diagnose, 5) +
  theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

plot_nmnds_diagnose <- plot_ordination(scaled, mothur_otus_nmnds,
type = "samples",

```

```

color = "diagnose",
shape = "diagnose",
title = "Ordenación NMDS - Diagnóstico") +
scale_color_manual(values = viridis(2)) +
geom_point(aes(color = diagnose), alpha = 0.7, size = 4)

plot_nmds_diagnose +
stat_ellipse(type = "norm", linetype = 2) +
stat_ellipse(type = "t") +
theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

plot_nmds_country <- plot_ordination(scaled, mothur_otus_nmds,
type = "samples",
color = "country",
shape = "country",
title = "Ordenación NMDS - Origen geográfico") +
scale_color_manual(values = viridis(4)) +
geom_point(aes(color = country), alpha = 0.7, size = 4)

plot_nmds_country +
stat_ellipse(type = "norm", linetype = 2) +
stat_ellipse(type = "t") +
theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

plot_nmds_country_diagnose <- plot_ordination(scaled, mothur_otus_nmds,
type = "samples",
color = "diagnose",
shape = "country",
title = "Ordenación NMDS - Origen geográfico vs diagnóstico") +
scale_color_manual(values = viridis(2)) +
geom_point(aes(color = diagnose), alpha = 0.7, size = 4)

plot_nmds_country_diagnose +
stat_ellipse(type = "norm", linetype = 2) +
stat_ellipse(type = "t") +
theme(axis.text = element_text(size=15), axis.title =
element_text(size=15), text = element_text(size=15))

# Análisis de la distribución de Los géneros bacterianos usando DESeq2
(tablas 3-5 y figuras 8-10):

deseq_belgium <- subset_samples(mothur_data, country == "Belgium")

deseq_belgium_deseq <- phyloseq_to_deseq2(deseq_belgium, ~ diagnose)
deseq_belgium_deseq <- DESeq(deseq_belgium_deseq, test="Wald", fitType =
"parametric")

results_deseq_belgium_deseq <- results(deseq_belgium_deseq)
alpha <- 0.01
sigtab_belgium <-
results_deseq_belgium_deseq[which(results_deseq_belgium_deseq$padj < alpha),
]
sigtab_belgium <- cbind(as(sigtab_belgium, "data.frame"),
as(tax_table(deseq_belgium)[rownames(sigtab_belgium), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
scale_fill_brewer(palette = palname, ...)
}

```

```

}

x <- tapply(sigtab_belgium_sub_10$log2FoldChange,
sigtab_belgium_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_belgium_sub_10$Phylum <-
factor(as.character(sigtab_belgium_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_belgium_sub_10$log2FoldChange,
sigtab_belgium_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_belgium_sub_10$Genus <-
factor(as.character(sigtab_belgium_sub_10$Genus), levels=names(x))
belgica_plot <- ggplot(sigtab_belgium_sub_10, aes(x=Genus, y=log2FoldChange,
color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

deseq_chile <- subset_samples(mothur_data, country == "Chile")

deseq_chile_deseq <- phyloseq_to_deseq2(deseq_chile, ~ diagnose)
deseq_chile_deseq <- DESeq(deseq_chile_deseq, test="Wald", fitType =
"parametric")

results_deseq_chile_deseq <- results(deseq_chile_deseq)
alpha <- 0.01
sigtab_chile <-
results_deseq_chile_deseq[which(results_deseq_chile_deseq$padj < alpha), ]
sigtab_chile <- cbind(as(sigtab_chile, "data.frame"),
as(tax_table(deseq_chile)[rownames(sigtab_chile), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_chile_sub_10$log2FoldChange, sigtab_chile_sub_10$Phylum,
function(x) max(x))
x <- sort(x, TRUE)
sigtab_chile_sub_10$Phylum <-
factor(as.character(sigtab_chile_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_chile_sub_10$log2FoldChange, sigtab_chile_sub_10$Genus,
function(x) max(x))
x <- sort(x, TRUE)
sigtab_chile_sub_10$Genus <- factor(as.character(sigtab_chile_sub_10$Genus),
levels=names(x))
chile_plot <- ggplot(sigtab_chile_sub_10, aes(x=Genus, y=log2FoldChange,
color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

deseq_spain <- subset_samples(mothur_data, country == "Peru")

deseq_peru_deseq <- phyloseq_to_deseq2(deseq_peru, ~ diagnose)
deseq_peru_deseq <- DESeq(deseq_peru_deseq, test="Wald", fitType =
"parametric")

results_deseq_peru_deseq <- results(deseq_peru_deseq)
alpha <- 0.01
sigtab_peru <- results_deseq_peru_deseq[which(results_deseq_peru_deseq$padj <
alpha), ]

```

```

sigtab_peru <- cbind(as(sigtab_peru, "data.frame"),
as(tax_table(deseq_peru)[rownames(sigtab_peru), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_peru_sub_10$log2FoldChange, sigtab_peru_sub_10$Phylum,
function(x) max(x))
x <- sort(x, TRUE)
sigtab_peru_sub_10$Phylum <- factor(as.character(sigtab_peru_sub_10$Phylum),
levels=names(x))

x <- tapply(sigtab_peru_sub_10$log2FoldChange, sigtab_peru_sub_10$Genus,
function(x) max(x))
x <- sort(x, TRUE)
sigtab_peru_sub_10$Genus <- factor(as.character(sigtab_peru_sub_10$Genus),
levels=names(x))
peru_plot <- ggplot(sigtab_peru_sub_10, aes(x=Genus, y=log2FoldChange,
color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

deseq_spain <- subset_samples(mothur_data, country == "Spain")

deseq_spain_deseq <- phyloseq_to_deseq2(deseq_spain, ~ diagnose)
deseq_spain_deseq <- DESeq(deseq_spain_deseq, test="Wald", fitType =
"parametric")

results_deseq_spain_deseq <- results(deseq_spain_deseq)
alpha <- 0.01
sigtab_spain <-
results_deseq_spain_deseq[which(results_deseq_spain_deseq$padj < alpha), ]
sigtab_spain <- cbind(as(sigtab_spain, "data.frame"),
as(tax_table(deseq_spain)[rownames(sigtab_spain), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_spain_sub_10$log2FoldChange, sigtab_spain_sub_10$Phylum,
function(x) max(x))
x <- sort(x, TRUE)
sigtab_spain_sub_10$Phylum <-
factor(as.character(sigtab_spain_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_spain_sub_10$log2FoldChange, sigtab_spain_sub_10$Genus,
function(x) max(x))
x <- sort(x, TRUE)
sigtab_spain_sub_10$Genus <- factor(as.character(sigtab_spain_sub_10$Genus),
levels=names(x))
spain_plot <- ggplot(sigtab_spain_sub_10, aes(x=Genus, y=log2FoldChange,
color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_perio <- subset_samples(mothur_data, diagnose == "Periodontitis")
all_healthy <- subset_samples(mothur_data, diagnose == "Healthy")

all_perio_be_cl <- subset_samples(all_perio, country == "Belgium" | country
== "Chile")

```

```

all_perio_be_cl_deseq2 <- phyloseq_to_deseq2(all_perio_be_cl, ~ country)
all_perio_be_cl_deseq2 <- DESeq(all_perio_be_cl_deseq2, test="Wald", fitType
= "parametric")

results_all_perio_be_cl_deseq2 <- results(all_perio_be_cl_deseq2)
alpha <- 0.01
sigtab_perio_be_cl <-
results_all_perio_be_cl_deseq2[which(results_all_perio_be_cl_deseq2$padj <
alpha), ]
sigtab_perio_be_cl <- cbind(as(sigtab_perio_be_cl, "data.frame"),
as(tax_table(all_perio_be_cl)[rownames(sigtab_perio_be_cl), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_perio_be_cl_sub_10$log2FoldChange,
sigtab_perio_be_cl_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_be_cl_sub_10$Phylum <-
factor(as.character(sigtab_perio_be_cl_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_perio_be_cl_sub_10$log2FoldChange,
sigtab_perio_be_cl_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_be_cl_sub_10$Genus <-
factor(as.character(sigtab_perio_be_cl_sub_10$Genus), levels=names(x))
perio_be_cl_plot <- ggplot(sigtab_perio_be_cl_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_perio_be_es <- subset_samples(all_perio, country == "Belgium" | country
== "Spain")

all_perio_be_es_deseq2 <- phyloseq_to_deseq2(all_perio_be_es, ~ country)
all_perio_be_es_deseq2 <- DESeq(all_perio_be_es_deseq2, test="Wald", fitType
= "parametric")

results_all_perio_be_es_deseq2 <- results(all_perio_be_es_deseq2)
alpha <- 0.01
sigtab_perio_be_es <-
results_all_perio_be_es_deseq2[which(results_all_perio_be_es_deseq2$padj <
alpha), ]
sigtab_perio_be_es <- cbind(as(sigtab_perio_be_es, "data.frame"),
as(tax_table(all_perio_be_es)[rownames(sigtab_perio_be_es), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_perio_be_es_sub_10$log2FoldChange,
sigtab_perio_be_es_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_be_es_sub_10$Phylum <-
factor(as.character(sigtab_perio_be_es_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_perio_be_es_sub_10$log2FoldChange,
sigtab_perio_be_es_sub_10$Genus, function(x) max(x))

```

```

x <- sort(x, TRUE)
sigtab_perio_be_es_sub_10$Genus <-
factor(as.character(sigtab_perio_be_es_sub_10$Genus), levels=names(x))
perio_be_es_plot <- ggplot(sigtab_perio_be_es_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_perio_be_pe <- subset_samples(all_perio, country == "Belgium" | country
== "Peru")

all_perio_be_pe_deseq2 <- phyloseq_to_deseq2(all_perio_be_pe, ~ country)
all_perio_be_pe_deseq2 <- DESeq(all_perio_be_pe_deseq2, test="Wald", fitType
= "parametric")

results_all_perio_be_pe_deseq2 <- results(all_perio_be_pe_deseq2)
alpha <- 0.01
sigtab_perio_be_pe <-
results_all_perio_be_pe_deseq2[which(results_all_perio_be_pe_deseq2$padj <
alpha), ]
sigtab_perio_be_pe <- cbind(as(sigtab_perio_be_pe, "data.frame"),
as(tax_table(all_perio_be_pe)[rownames(sigtab_perio_be_pe), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_perio_be_pe_sub_10$log2FoldChange,
sigtab_perio_be_pe_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_be_pe_sub_10$Phylum <-
factor(as.character(sigtab_perio_be_pe_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_perio_be_pe_sub_10$log2FoldChange,
sigtab_perio_be_pe_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_be_pe_sub_10$Genus <-
factor(as.character(sigtab_perio_be_pe_sub_10$Genus), levels=names(x))
perio_be_pe_plot <- ggplot(sigtab_perio_be_pe_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_perio_cl_es <- subset_samples(all_perio, country == "Chile" | country ==
"Spain")

all_perio_cl_es_deseq2 <- phyloseq_to_deseq2(all_perio_cl_es, ~ country)
all_perio_cl_es_deseq2 <- DESeq(all_perio_cl_es_deseq2, test="Wald", fitType
= "parametric")

results_all_perio_cl_es_deseq2 <- results(all_perio_cl_es_deseq2)
alpha <- 0.01
sigtab_perio_cl_es <-
results_all_perio_cl_es_deseq2[which(results_all_perio_cl_es_deseq2$padj <
alpha), ]
sigtab_perio_cl_es <- cbind(as(sigtab_perio_cl_es, "data.frame"),
as(tax_table(all_perio_cl_es)[rownames(sigtab_perio_cl_es), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

```



```

x <- tapply(sigtab_perio_cl_es_sub_10$log2FoldChange,
sigtab_perio_cl_es_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_cl_es_sub_10$Phylum <-
factor(as.character(sigtab_perio_cl_es_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_perio_cl_es_sub_10$log2FoldChange,
sigtab_perio_cl_es_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_cl_es_sub_10$Genus <-
factor(as.character(sigtab_perio_cl_es_sub_10$Genus), levels=names(x))
perio_cl_es_plot <- ggplot(sigtab_perio_cl_es_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_perio_cl_pe <- subset_samples(all_perio, country == "Chile" | country ==
"Peru")

all_perio_cl_pe_deseq2 <- phyloseq_to_deseq2(all_perio_cl_pe, ~ country)
all_perio_cl_pe_deseq2 <- DESeq(all_perio_cl_pe_deseq2, test="Wald", fitType
= "parametric")

results_all_perio_cl_pe_deseq2 <- results(all_perio_cl_pe_deseq2)
alpha <- 0.01
sigtab_perio_cl_pe <-
results_all_perio_cl_pe_deseq2[which(results_all_perio_cl_pe_deseq2$padj <
alpha), ]
sigtab_perio_cl_pe <- cbind(as(sigtab_perio_cl_pe, "data.frame"),
as(tax_table(all_perio_cl_pe)[rownames(sigtab_perio_cl_pe), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_perio_cl_pe_sub_10$log2FoldChange,
sigtab_perio_cl_pe_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_cl_pe_sub_10$Phylum <-
factor(as.character(sigtab_perio_cl_pe_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_perio_cl_pe_sub_10$log2FoldChange,
sigtab_perio_cl_pe_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_cl_pe_sub_10$Genus <-
factor(as.character(sigtab_perio_cl_pe_sub_10$Genus), levels=names(x))
perio_cl_pe_plot <- ggplot(sigtab_perio_cl_pe_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_perio_es_pe <- subset_samples(all_perio, country == "Spain" | country ==
"Peru")

all_perio_es_pe_deseq2 <- phyloseq_to_deseq2(all_perio_es_pe, ~ country)
all_perio_es_pe_deseq2 <- DESeq(all_perio_es_pe_deseq2, test="Wald", fitType
= "parametric")

results_all_perio_es_pe_deseq2 <- results(all_perio_es_pe_deseq2)
alpha <- 0.01

```

```

sigtab_perio_es_pe <-
results_all_perio_es_pe_deseq2[which(results_all_perio_es_pe_deseq2$padj <
alpha), ]
sigtab_perio_es_pe <- cbind(as(sigtab_perio_es_pe, "data.frame"),
as(tax_table(all_perio_es_pe)[rownames(sigtab_perio_es_pe), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_perio_es_pe_sub_10$log2FoldChange,
sigtab_perio_es_pe_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_es_pe_sub_10$Phylum <-
factor(as.character(sigtab_perio_es_pe_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_perio_es_pe_sub_10$log2FoldChange,
sigtab_perio_es_pe_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_perio_es_pe_sub_10$Genus <-
factor(as.character(sigtab_perio_es_pe_sub_10$Genus), levels=names(x))
perio_es_pe_plot <- ggplot(sigtab_perio_es_pe_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_healthy_be_cl <- subset_samples(all_healthy, country == "Belgium" |
country == "Chile")

all_healthy_be_cl_deseq2 <- phyloseq_to_deseq2(all_healthy_be_cl, ~ country)
all_healthy_be_cl_deseq2 <- DESeq(all_healthy_be_cl_deseq2, test="Wald",
fitType = "parametric")

results_all_healthy_be_cl_deseq2 <- results(all_healthy_be_cl_deseq2)
alpha <- 0.01
sigtab_healthy_be_cl <-
results_all_healthy_be_cl_deseq2[which(results_all_healthy_be_cl_deseq2$padj
< alpha), ]
sigtab_healthy_be_cl <- cbind(as(sigtab_healthy_be_cl, "data.frame"),
as(tax_table(all_healthy_be_cl)[rownames(sigtab_healthy_be_cl), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_healthy_be_cl_sub_10$log2FoldChange,
sigtab_healthy_be_cl_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_be_cl_sub_10$Phylum <-
factor(as.character(sigtab_healthy_be_cl_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_healthy_be_cl_sub_10$log2FoldChange,
sigtab_healthy_be_cl_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_be_cl_sub_10$Genus <-
factor(as.character(sigtab_healthy_be_cl_sub_10$Genus), levels=names(x))
healthy_be_cl_plot <- ggplot(sigtab_healthy_be_cl_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

```

```

all_healthy_be_es <- subset_samples(all_healthy, country == "Belgium" |
country == "Spain")

all_healthy_be_es_deseq2 <- phyloseq_to_deseq2(all_healthy_be_es, ~ country)
all_healthy_be_es_deseq2 <- DESeq(all_healthy_be_es_deseq2, test="Wald",
fitType = "parametric")

results_all_healthy_be_es_deseq2 <- results(all_healthy_be_es_deseq2)
alpha <- 0.01
sigtab_healthy_be_es <-
results_all_healthy_be_es_deseq2[which(results_all_healthy_be_es_deseq2$padj
< alpha), ]
sigtab_healthy_be_es <- cbind(as(sigtab_healthy_be_es, "data.frame"),
as(tax_table(all_healthy_be_es)[rownames(sigtab_healthy_be_es), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_healthy_be_es_sub_10$log2FoldChange,
sigtab_healthy_be_es_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_be_es_sub_10$Phylum <-
factor(as.character(sigtab_healthy_be_es_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_healthy_be_es_sub_10$log2FoldChange,
sigtab_healthy_be_es_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_be_es_sub_10$Genus <-
factor(as.character(sigtab_healthy_be_es_sub_10$Genus), levels=names(x))
healthy_be_es_plot <- ggplot(sigtab_healthy_be_es_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_healthy_be_pe <- subset_samples(all_healthy, country == "Belgium" |
country == "Peru")

all_healthy_be_pe_deseq2 <- phyloseq_to_deseq2(all_healthy_be_pe, ~ country)
all_healthy_be_pe_deseq2 <- DESeq(all_healthy_be_pe_deseq2, test="Wald",
fitType = "parametric")

results_all_healthy_be_pe_deseq2 <- results(all_healthy_be_pe_deseq2)
alpha <- 0.01
sigtab_healthy_be_pe <-
results_all_healthy_be_pe_deseq2[which(results_all_healthy_be_pe_deseq2$padj
< alpha), ]
sigtab_healthy_be_pe <- cbind(as(sigtab_healthy_be_pe, "data.frame"),
as(tax_table(all_healthy_be_pe)[rownames(sigtab_healthy_be_pe), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_healthy_be_pe_sub_10$log2FoldChange,
sigtab_healthy_be_pe_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_be_pe_sub_10$Phylum <-
factor(as.character(sigtab_healthy_be_pe_sub_10$Phylum), levels=names(x))

```

```

x <- tapply(sigtab_healthy_be_pe_sub_10$log2FoldChange,
sigtab_healthy_be_pe_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_be_pe_sub_10$Genus <-
factor(as.character(sigtab_healthy_be_pe_sub_10$Genus), levels=names(x))
healthy_be_pe_plot <- ggplot(sigtab_healthy_be_pe_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_healthy_cl_es <- subset_samples(all_healthy, country == "Chile" | country
== "Spain")

all_healthy_cl_es_deseq2 <- phyloseq_to_deseq2(all_healthy_cl_es, ~ country)
all_healthy_cl_es_deseq2 <- DESeq(all_healthy_cl_es_deseq2, test="Wald",
fitType = "parametric")

results_all_healthy_cl_es_deseq2 <- results(all_healthy_cl_es_deseq2)
alpha <- 0.01
sigtab_healthy_cl_es <-
results_all_healthy_cl_es_deseq2[which(results_all_healthy_cl_es_deseq2$padj
< alpha), ]
sigtab_healthy_cl_es <- cbind(as(sigtab_healthy_cl_es, "data.frame"),
as(tax_table(all_healthy_cl_es)[rownames(sigtab_healthy_cl_es), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_healthy_cl_es_sub_10$log2FoldChange,
sigtab_healthy_cl_es_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_cl_es_sub_10$Phylum <-
factor(as.character(sigtab_healthy_cl_es_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_healthy_cl_es_sub_10$log2FoldChange,
sigtab_healthy_cl_es_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_cl_es_sub_10$Genus <-
factor(as.character(sigtab_healthy_cl_es_sub_10$Genus), levels=names(x))
healthy_cl_es_plot <- ggplot(sigtab_healthy_cl_es_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_healthy_cl_pe <- subset_samples(all_healthy, country == "Chile" | country
== "Peru")

all_healthy_cl_pe_deseq2 <- phyloseq_to_deseq2(all_healthy_cl_pe, ~ country)
all_healthy_cl_pe_deseq2 <- DESeq(all_healthy_cl_pe_deseq2, test="Wald",
fitType = "parametric")

results_all_healthy_cl_pe_deseq2 <- results(all_healthy_cl_pe_deseq2)
alpha <- 0.01
sigtab_healthy_cl_pe <-
results_all_healthy_cl_pe_deseq2[which(results_all_healthy_cl_pe_deseq2$padj
< alpha), ]
sigtab_healthy_cl_pe <- cbind(as(sigtab_healthy_cl_pe, "data.frame"),
as(tax_table(all_healthy_cl_pe)[rownames(sigtab_healthy_cl_pe), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {

```

```

    scale_fill_brewer(palette = palname, ...)
  }

x <- tapply(sigtab_healthy_cl_pe_sub_10$log2FoldChange,
sigtab_healthy_cl_pe_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_cl_pe_sub_10$Phylum <-
factor(as.character(sigtab_healthy_cl_pe_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_healthy_cl_pe_sub_10$log2FoldChange,
sigtab_healthy_cl_pe_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_cl_pe_sub_10$Genus <-
factor(as.character(sigtab_healthy_cl_pe_sub_10$Genus), levels=names(x))
healthy_cl_pe_plot <- ggplot(sigtab_healthy_cl_pe_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

all_healthy_es_pe <- subset_samples(all_healthy, country == "Spain" | country
== "Peru")

all_healthy_es_pe_deseq2 <- phyloseq_to_deseq2(all_healthy_es_pe, ~ country)
all_healthy_es_pe_deseq2 <- DESeq(all_healthy_es_pe_deseq2, test="Wald",
fitType = "parametric")

results_all_healthy_es_pe_deseq2 <- results(all_healthy_es_pe_deseq2)
alpha <- 0.01
sigtab_healthy_es_pe <-
results_all_healthy_es_pe_deseq2[which(results_all_healthy_es_pe_deseq2$padj
< alpha), ]
sigtab_healthy_es_pe <- cbind(as(sigtab_healthy_es_pe, "data.frame"),
as(tax_table(all_healthy_es_pe)[rownames(sigtab_healthy_es_pe), ], "matrix"))

scale_fill_discrete <- function(palname = "Set1", ...) {
  scale_fill_brewer(palette = palname, ...)
}

x <- tapply(sigtab_healthy_es_pe_sub_10$log2FoldChange,
sigtab_healthy_es_pe_sub_10$Phylum, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_es_pe_sub_10$Phylum <-
factor(as.character(sigtab_healthy_es_pe_sub_10$Phylum), levels=names(x))

x <- tapply(sigtab_healthy_es_pe_sub_10$log2FoldChange,
sigtab_healthy_es_pe_sub_10$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab_healthy_es_pe_sub_10$Genus <-
factor(as.character(sigtab_healthy_es_pe_sub_10$Genus), levels=names(x))
healthy_es_pe_plot <- ggplot(sigtab_healthy_es_pe_sub_10, aes(x=Genus,
y=log2FoldChange, color=Phylum)) + geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5))

sigtab_healthy_es_pe_sub <- subset(sigtab_healthy_es_pe, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_healthy_es_pe_sub_10 <-
head(sigtab_healthy_es_pe_sub[order(sigtab_healthy_es_pe_sub$padj),],10)
write.csv(sigtab_healthy_es_pe_sub_10, "./healthy_es_pe.csv")

sigtab_healthy_cl_pe_sub <- subset(sigtab_healthy_cl_pe, padj <= 0.1,

```

```

select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_healthy_cl_pe_sub_10 <-
head(sigtab_healthy_cl_pe_sub[order(sigtab_healthy_cl_pe_sub$padj),],10)
write.csv(sigtab_healthy_cl_pe_sub_10, "./healthy_cl_pe.csv")

sigtab_healthy_cl_es_sub <- subset(sigtab_healthy_cl_es, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_healthy_cl_es_sub_10 <-
head(sigtab_healthy_cl_es_sub[order(sigtab_healthy_cl_es_sub$padj),],10)
write.csv(sigtab_healthy_cl_es_sub_10, "./healthy_cl_es.csv")

sigtab_healthy_be_pe_sub <- subset(sigtab_healthy_be_pe, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_healthy_be_pe_sub_10 <-
head(sigtab_healthy_be_pe_sub[order(sigtab_healthy_be_pe_sub$padj),],10)
write.csv(sigtab_healthy_be_pe_sub_10, "./healthy_be_pe.csv")

sigtab_healthy_be_es_sub <- subset(sigtab_healthy_be_es, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_healthy_be_es_sub_10 <-
head(sigtab_healthy_be_es_sub[order(sigtab_healthy_be_es_sub$padj),],10)
write.csv(sigtab_healthy_be_es_sub_10, "./healthy_be_es.csv")

sigtab_healthy_be_cl_sub <- subset(sigtab_healthy_be_cl, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_healthy_be_cl_sub_10 <-
head(sigtab_healthy_be_cl_sub[order(sigtab_healthy_be_cl_sub$padj),],10)
write.csv(sigtab_healthy_be_cl_sub_10, "./healthy_be_cl.csv")

sigtab_perio_es_pe_sub <- subset(sigtab_perio_es_pe, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_perio_es_pe_sub_10 <-
head(sigtab_perio_es_pe_sub[order(sigtab_perio_es_pe_sub$padj),],10)
write.csv(sigtab_perio_es_pe_sub_10, "./perio_es_pe.csv")

sigtab_perio_cl_pe_sub <- subset(sigtab_perio_cl_pe, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_perio_cl_pe_sub_10 <-
head(sigtab_perio_cl_pe_sub[order(sigtab_perio_cl_pe_sub$padj),],10)
write.csv(sigtab_perio_cl_pe_sub_10, "./perio_cl_pe.csv")

sigtab_perio_cl_es_sub <- subset(sigtab_perio_cl_es, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_perio_cl_es_sub_10 <-
head(sigtab_perio_cl_es_sub[order(sigtab_perio_cl_es_sub$padj),],10)
write.csv(sigtab_perio_cl_es_sub_10, "./perio_cl_es.csv")

sigtab_perio_be_pe_sub <- subset(sigtab_perio_be_pe, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_perio_be_pe_sub_10 <-
head(sigtab_perio_be_pe_sub[order(sigtab_perio_be_pe_sub$padj),],10)
write.csv(sigtab_perio_be_pe_sub_10, "./perio_be_pe.csv")

sigtab_perio_be_es_sub <- subset(sigtab_perio_be_es, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_perio_be_es_sub_10 <-
head(sigtab_perio_be_es_sub[order(sigtab_perio_be_es_sub$padj),],10)
write.csv(sigtab_perio_be_es_sub_10, "./perio_be_es.csv")

```

```

sigtab_perio_be_cl_sub <- subset(sigtab_perio_be_cl, padj <= 0.1,
select=c(Genus, baseMean, log2FoldChange, padj, Phylum))
sigtab_perio_be_cl_sub_10 <-
head(sigtab_perio_be_cl_sub[order(sigtab_perio_be_cl_sub$padj),],10)
write.csv(sigtab_perio_be_cl_sub_10, "./perio_be_cl.csv")

sigtab_belgium_sub <- subset(sigtab_belgium, padj <= 0.1, select=c(Genus,
baseMean, log2FoldChange, padj, Phylum))
sigtab_belgium_sub_10 <-
head(sigtab_belgium_sub[order(sigtab_belgium_sub$padj),],10)
write.csv(sigtab_belgium_sub_10, "./belgium.csv")

sigtab_chile_sub <- subset(sigtab_chile, padj <= 0.1, select=c(Genus,
baseMean, log2FoldChange, padj, Phylum))
sigtab_chile_sub_10 <-
head(sigtab_chile_sub[order(sigtab_chile_sub$padj),],10)
write.csv(sigtab_chile_sub_10, "./chile.csv")

sigtab_peru_sub <- subset(sigtab_peru, padj <= 0.1, select=c(Genus, baseMean,
log2FoldChange, padj, Phylum))
sigtab_peru_sub_10 <- head(sigtab_peru_sub[order(sigtab_peru_sub$padj),],10)
write.csv(sigtab_peru_sub_10, "./peru.csv")

sigtab_spain_sub <- subset(sigtab_spain, padj <= 0.1, select=c(Genus,
baseMean, log2FoldChange, padj, Phylum))
sigtab_spain_sub_10 <-
head(sigtab_spain_sub[order(sigtab_spain_sub$padj),],10)
write.csv(sigtab_spain_sub_10, "./spain.csv")

```