

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS



TESIS DOCTORAL

**Estudio bioinformático de muestras clínicas en la infección
de Helicobacter pylori: importancia del microbioma**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Claudio Alba Rubio

DIRECTORA

Teresa Alarcón Caveró

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Biológicas

Departamento de Microbiología



TESIS DOCTORAL

**Estudio bioinformático de muestras clínicas en la
infección de *Helicobacter pylori*: Importancia del
microbioma**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

CLAUDIO ALBA RUBIO

Directora

Teresa Alarcón Caveró

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS

**Estudio bioinformático de muestras clínicas en la
infección de *Helicobacter pylori*: Importancia del
microbioma**

PRESENTADA POR:

Claudio Alba Rubio

VºBº DIRECTORA DE TESIS

Fdo. Dra. Teresa Alarcón Cavero

Fdo. Claudio Alba Rubio

MADRID, 2020

"Si quieres construir un barco, no empieces por buscar madera, cortar tablas o distribuir el trabajo, primero has de evocar en los hombres el anhelo del mar."

Antoine de Saint-Exupéry (1900-1944)

Agradecimientos

Agradecimientos

Para empezar, no se originaría el conocimiento sin curiosidad. Gracias, en primer lugar a mi directora, la doctora Teresa Alarcón. Has sido la guía que evocó en mí el anhelo del mar. Sin ti, esta tesis no habría empezado nunca. Fuiste capaz de despertar esa inquietud y ganas de aprender que solo surgen de presenciar la pasión de esas personas que sienten devoción por su investigación. Con ello, no solo me ayudaste a adentrarme en el mundo de la microbiología, sino también en el mundo de la bioinformática y todas sus posibilidades. Estoy seguro de que todo ello me acompañará a lo largo de toda la vida.

En segundo lugar, me gustaría dar las gracias a todo el Servicio de Microbiología y Parasitología del Hospital Universitario de La Princesa. A todos los técnicos y facultativos que, desde el primer día, me hicieron sentir como uno más, enseñándome de todas sus rutinas y su conocimiento acumulado a lo largo de los años. Especial mención tienen todos los residentes que coincidieron conmigo durante esos dos años: Belén, Silvia, Nelly, Tamara, Arturo, Sara, Sandra, Ana Miqueleiz, M^o Dolores y Luz... Me acogisteis como a uno más. Gracias por todo lo que me enseñasteis y todo el tiempo que pasamos juntos. De las residentes hay dos que no he mencionado antes por su gran papel en esta tesis: la Dra. Laura Llorca, que abrió camino en esta área de la bioinformática y ha sido una fuente de inspiración a lo largo de toda la tesis y la Dra. Ana Blanco, compañera de tesis con la que siempre ha sido un placer trabajar y aprender de ella.

De todo el servicio no puedo dejarme a dos personas muy especiales: el Dr. Diego Domingo, que siempre logró que cada día en el servicio fuera mejor que el anterior gracias a su alegría y consiguiendo que todo el mundo fuera un poco más feliz y, por supuesto, a Vero, los dos años en el hospital habrían valido la pena aunque solo fuera por lo que aprendimos y nos apoyamos juntos.

Tengo que agradecer, sin ninguna duda, al Dr. Juan Miguel Rodríguez por darme la oportunidad de seguir creciendo en su casa, la Facultad de Veterinaria, donde no ha dejado de enseñarme y darme facilidades para poder realizar esta tesis. Ahí he tenido la suerte de rodearme de gente maravillosa y muy trabajadora: Rebeca, Josué, Irma,

Agradecimientos

Marina, Aina, las dos Raqueles, Edu... Esas conversaciones y ratos de desconexión en el café contribuyeron más que las horas interminables frente a la pantalla. A la Dra. Leónides, contigo ha sido con quien más he aprendido a realizar los análisis estadísticos. Sin ti y tus ganas de encontrar la perfección, esta tesis sería mucho más pobre y poco práctica.

A mi familia, especialmente a mis padres, habéis sido consejeros, apoyo, fortaleza, emociones, risas, enfados, médicos, psicólogos, esfuerzo, confianza, amor... A mis tías, que siempre habéis sido mis "segunda madre" y se aplica, sin lugar a duda, todo lo anterior. A mi hermana Aurorita, que me ha ayudado a forjar la paciencia para esta tarea y que quiero tanto. A mis primas y primos con los que he crecido, en especial a Alicia, has llenado de color esta tesis.

A mis amigos Guille, Alex, Kike y Euge. Vosotros sois mis hermanos, así que entráis en el grupo de familia. Gracias por aguantarme.

Y, por último, el mayor de los agradecimientos es para Yolanda, mi bióloga favorita. Eres una continua fuente de risas y alegrías y la mejor compañera de vida. Gracias por quererme y cuidarme tanto. Te quiero mucho. ¿Nos imaginabas así hace 8 años?

Por ello, no puedo hacer otra cosa que agradecerles una y mil veces a cada uno de ellos por su gran papel en esta tesis: GRACIAS.

Índice

Índice

1. Resumen	3
2. Introducción	13
2.1 <i>Helicobacter pylori</i>	13
2.1.1 Antecedentes históricos y situación taxonómica actual	13
2.1.2 Morfología y características fisiológicas del género <i>Helicobacter</i>	14
2.1.3 Genoma de <i>Helicobacter pylori</i>	16
2.1.4 Factores de virulencia	17
2.1.5 Patogenicidad	19
2.1.6 Métodos de diagnóstico	21
2.1.7 Tratamiento	25
2.2 Estudio del microbioma y sus limitaciones	26
2.2.1 Revisión de definiciones	27
2.2.2 Microbioma humano	29
2.2.3 Funciones de la microbiota	29
2.2.4 Técnicas para el estudio de la microbiota	30
2.2.5 Clasificación taxonómica	31
2.2.6 Estudios de diversidad	34
2.2.7. Microbioma gástrico	37
2.2.8 Importancia del microbioma en las patologías gástricas	39
2.3 Herramientas bioinformáticas	41
2.3.1 Proteómica	41
2.3.2 Secuenciación masiva	42
2.3.3 Los repositorios de datos	45
2.3.4 Etapas en el estudio de la microbiota y sus limitaciones	46
3. Objetivos	57
4. Material y métodos	61

Índice

4.1 Pacientes y muestras	61
4.2 Identificación de <i>H. pylori</i> por MALDI-TOF	62
4.2.1 Selección de las cepas	62
4.2.2 Extracción proteica e introducción en la base de datos	63
4.2.3 Ensayo de mejora de la identificación por MALDI.....	64
4.2.4 Ensayo de factibilidad en clínica	64
4.3 Análisis bioinformático para estudio taxonómico	65
4.3.1 Ensamblaje, control de calidad y agrupación de las secuencias.....	65
4.3.2 Asignación taxonómica.....	66
4.3.3 Análisis metataxonómico en muestras gástricas.....	66
4.4 Selección de proyecto para el cálculo del tamaño muestral	72
4.5 Selección del resto de proyectos y muestras	73
5. Resultados	81
5.1 Aplicación de bioinformática a estudios proteómicos	81
5.2. Documento de instrucciones para el análisis metataxonómico	83
5.3 Cálculo del tamaño muestral	84
5.4 Determinación del efecto de la utilización de diferentes bases de datos de referencia en estudios metataxonómicos	86
5.4.1 Diversidad alfa	87
5.4.2 Diversidad beta	87
5.4.3 Asignaciones taxonómicas.....	88
5.5 Descripción taxonómica de los proyectos estudiados	92
5.5.1 Proyecto Llorca <i>et al.</i> (2017).....	93
5.5.2 Proyecto PRJNA313391.....	94
5.5.3 Proyecto PRJEB22107	95
5.5.4 Proyecto PRJEB21104	96
5.5.5 Proyecto PRJNA310127.....	97
5.5.6 Proyecto PRJNA313391.....	98

5.5.7 Estudio de idoneidad del proyecto PRJNA313391	99
5.6 Influencia de <i>Helicobacter pylori</i>	105
5.6.1 Diversidad alfa.	105
5.6.2 Diversidad beta	106
5.6.3 Diferencias taxonómicas	107
5.6.4 Relación de los géneros mayoritarios.	109
5.6.5 Microbiota <i>core</i>	111
5.6.6 Efecto de la sustracción de las secuencias del género <i>Helicobacter</i> de los análisis.	112
5.7 Influencia de la patología	118
5.7.1 Diversidad microbiana por patología en pacientes negativos	119
5.7.2 Diversidad microbiana por patología en pacientes positivos.....	123
5.7.3 Diversidad microbiana por patología eliminando las secuencias pertenecientes a bacterias del género <i>Helicobacter</i>	128
5.8 Influencia del país de residencia en la patología por <i>H. pylori</i>	134
5.8.1 Diversidad alfa	134
5.8.2 Diversidad beta	135
5.8.3 Taxonomía.....	136
6. Discusión	145
6.1 Aplicaciones bioinformáticas a estudios proteómicos	145
6.2 Desarrollo de un documento con instrucciones	147
6.3 Estudio de la microbiota gástrica	149
6.3.1 Descripción de las muestras y estudios seleccionados	149
6.3.2 Cálculo del tamaño muestral.	151
6.4 Utilización de diferentes bases de datos	152
6.5 Influencia de <i>H. pylori</i>	153
6.6 Influencia de la patología	158
6.7 Efecto de la región geográfica en la microbiota gástrica de pacientes <i>H. pylori</i> positivos	162

Índice

6.8 Limitaciones del estudio.....	165
6.9 Discusión final.....	167
7. Conclusiones	173
8. Bibliografía	177
Anexo I	205
Anexo II	217
Anexo III	253

Índice de tablas

TABLA 1: PRINCIPALES CARACTERÍSTICAS DE LOS PACIENTES Y SU DIAGNÓSTICO DE LAS MUESTRAS DE LLORCA ET AL. (2017).....	72
TABLA 2: PRINCIPALES CARACTERÍSTICAS DE LOS PACIENTES CUYAS MUESTRAS HAN SIDO INCLUIDAS Y SU DIAGNÓSTICO (PRJEB21497).....	74
TABLA 3: PRINCIPALES CARACTERÍSTICAS DE LOS PACIENTES CUYAS MUESTRAS HAN SIDO INCLUIDAS Y SU DIAGNÓSTICO (PRJEB22107).....	75
TABLA 4: PRINCIPALES CARACTERÍSTICAS DE LOS PACIENTES CUYAS MUESTRAS HAN SIDO INCLUIDAS Y SU DIAGNÓSTICO (PRJEB21104).....	76
TABLA 5: PRINCIPALES CARACTERÍSTICAS DE LOS PACIENTES CUYAS MUESTRAS HAN SIDO INCLUIDAS Y SU DIAGNÓSTICO (PRJNA310127).....	77
TABLA 6: PRINCIPALES CARACTERÍSTICAS DE LOS PACIENTES CUYAS MUESTRAS HAN SIDO INCLUIDAS Y SU DIAGNÓSTICO (PRJNA313391).....	78
TABLA 7: NÚMERO Y PORCENTAJE POR PUNTUACIÓN DE CEPAS IDENTIFICADAS COMO <i>H. PYLORI</i> EN FUNCIÓN DE LA BASE DE DATOS UTILIZADA.....	81
TABLA 8: NÚMERO Y PORCENTAJE DE CEPAS IDENTIFICADAS COMO <i>H. PYLORI</i> CON SUS VALORES POR MALDI CON LAS BASES DE DATOS COMBINADAS.....	82
TABLA 9: CÁLCULO DE LOS ÍNDICES DE DIVERSIDAD EN FUNCIÓN DEL ESTADO DE INFECCIÓN DE LOS PACIENTES POR <i>H. PYLORI</i> Y POTENCIA DE LA PRUEBA ESTADÍSTICA WILKONXONMANN-WHITNE	84
TABLA 10: COMPARATIVA DE LAS 20 ASIGNACIONES TAXONÓMICAS MÁS ABUNDANTES. EN GRIS POSIBLES REASIGNACIONES TAXONÓMICAS.....	89
TABLA 11: TAXONOMÍA DE LOS 20 GÉNEROS MAYORITARIOS ORDENADOS POR LA ABUNDANCIA RELATIVA DE LOS FILOS (GRISES) Y LOS GÉNEROS.....	91
TABLA 12: DATOS DE DIAGNÓSTICO DE LA INFECCIÓN POR <i>H. PYLORI</i> Y PATOLOGÍA ASOCIADA DE LAS MUESTRAS QUE CUMPLIERON LOS CRITERIOS DE INCLUSIÓN DE TODOS LOS PROYECTOS.....	92
TABLA 13: ABUNDANCIA RELATIVA (%), EXPRESADA COMO LA MEDIANA Y LOS RANGOS INTERCUARTÍlicos (IQR), DE LOS 20 GÉNEROS Y 4 FILOS MAYORITARIOS EN FUNCIÓN DEL PROYECTO ANALIZADO.....	102
TABLA 14: ABUNDANCIA RELATIVA (%), EXPRESADA COMO LA MEDIANA Y LOS RANGOS INTERCUARTÍlicos (IQR), DE LOS 20 GÉNEROS Y 4 FILOS MAYORITARIOS EN FUNCIÓN DIAGNÓSTICO PARA <i>H. PYLORI</i>	108
TABLA 15: ABUNDANCIA RELATIVA (%), EXPRESADA COMO LA MEDIANA Y LOS RANGOS INTERCUARTÍlicos (IQR), DE LOS 20 GÉNEROS Y 4 FILOS MAYORITARIOS EN FUNCIÓN DEL DIAGNÓSTICO PARA <i>H. PYLORI</i> TRAS ELIMINAR LAS SECUENCIAS DE <i>H. PYLORI</i> DE LOS ANÁLISIS.....	115
TABLA 16: NÚMERO DE MUESTRAS DE BIOPSIA GÁSTRICA POR DIAGNÓSTICO PARA LA INFECCIÓN POR <i>H. PYLORI</i> (NEGATIVOS= NEG; POSITIVOS = HpPOS) Y PATOLOGÍA GÁSTRICA ASOCIADA. SE MUESTRA EL NÚMERO DE SECUENCIAS E INFORMACIÓN SOBRE EL NÚMERO MÍNIMO (MIN), MEDIANA, CUARTIL 1 (Q1), CUARTIL 3 (Q3) Y EL NÚMERO MÁXIMO DE SECUENCIAS EN UNA MUESTRA DE ESE GRUPO (MAX).....	118
TABLA 17: BIODIVERSIDAD, MEDIDA POR EL ÍNDICE DE SHANNON Y SIMPSON, EN MUESTRAS DE BIOPSIA GÁSTRICA DE PACIENTES DIAGNOSTICADOS COMO NEGATIVOS A INFECCIÓN POR <i>H. PYLORI</i> AGRUPADOS POR PATOLOGÍA GÁSTRICA.....	119
TABLA 18: ABUNDANCIA RELATIVA (%), EXPRESADA COMO LA MEDIANA Y LOS RANGOS INTERCUARTÍlicos (IQR), DE LOS 20 GÉNEROS Y 4 FILOS MAYORITARIOS EN FUNCIÓN LA PATOLOGÍA EN PACIENTES <i>H. PYLORI</i> NEGATIVOS.....	122
TABLA 19: ABUNDANCIA RELATIVA (%), EXPRESADA COMO LA MEDIANA Y LOS RANGOS INTERCUARTÍlicos (IQR), DE LOS 20 GÉNEROS Y 4 FILOS MAYORITARIOS EN FUNCIÓN LA PATOLOGÍA EN PACIENTES <i>H. PYLORI</i> POSITIVOS.....	126
TABLA 20: NÚMERO DE MUESTRAS ACEPTADAS TRAS ELIMINAR LAS SECUENCIAS DE LAS BACTERIAS DEL GÉNERO <i>HELICOBACTER</i> DE BIOPSIAS GÁSTRICA POR DIAGNÓSTICO PARA LA INFECCIÓN POR <i>H. PYLORI</i> (NEGATIVOS= NEG; POSITIVOS = HpPOS) Y PATOLOGÍA GÁSTRICA ASOCIADA. SE MUESTRA EL NÚMERO DE SECUENCIAS E INFORMACIÓN SOBRE EL NÚMERO MÍNIMO (MIN),	

Índice

MEDIANA, CUARTIL 1 (Q1), CUARTIL 3 (Q3) Y EL NÚMERO MÁXIMO DE SECUENCIAS EN UNA MUESTRA DE ESE GRUPO (MAX).....	128
TABLA 21: GÉNEROS QUE CUMPLEN LA CONDICIÓN DE ESTAR PRESENTES EN MÁS DE UN 80% DE LAS MUESTRAS CON UNA ABUNDANCIA RELATIVA DEL 0,01% EN, AL MENOS, 1 DE LOS GRUPOS.	133
TABLA 22: GÉNEROS QUE CUMPLEN LA CONDICIÓN DE ESTAR PRESENTES EN MÁS DE UN 80% DE LAS MUESTRAS CON UNA ABUNDANCIA RELATIVA DEL 1% EN, AL MENOS, 1 DE LOS GRUPOS.	133
TABLA 23: NÚMERO DE MUESTRAS Y SECUENCIAS OBTENIDAS DE MUESTRAS DE BIOPSIA GÁSTRICA DE PACIENTES DIAGNOSTICADOS POSITIVOS PARA INFECCIÓN POR <i>H. PYLORI</i> CATEGORIZADAS EN FUNCIÓN DEL PAÍS DE RESIDENCIA DE LOS PACIENTES.....	134
TABLA 24: ABUNDANCIA RELATIVA (%), EXPRESADA COMO LA MEDIANA Y LOS RANGOS INTERCUARTÍlicos (IQR), DE LOS 20 GÉNEROS Y 4 FILOS MAYORITARIOS, DE MUESTRAS DE BIOPSIA GÁSTRICA AGRUPADAS EN FUNCIÓN DEL PAÍS DE RESIDENCIA DE LOS PACIENTES DIAGNOSTICADOS COMO <i>H. PYLORI</i> POSITIVOS.	138
TABLA 25: FILOS QUE CUMPLEN LA CONDICIÓN DE ESTAR PRESENTES EN MÁS DE UN 80% DE LAS MUESTRAS CON UNA ABUNDANCIA RELATIVA DEL 0,01% POR CADA GRUPO.....	141
TABLA 26: FILOS QUE CUMPLEN LA CONDICIÓN DE ESTAR PRESENTES EN MÁS DE UN 80% DE LAS MUESTRAS CON UNA ABUNDANCIA RELATIVA DEL 1% POR CADA GRUPO.....	141
TABLA 27: GÉNEROS QUE CUMPLEN LA CONDICIÓN DE ESTAR PRESENTES EN MÁS DE UN 80% DE LAS MUESTRAS CON UNA ABUNDANCIA RELATIVA DEL 0,01% POR CADA GRUPO.....	142

Índice de figuras

FIGURA 1: <i>H. PYLORI</i> CULTIVADO EN MEDIO DE COLUMBIA AGAR SANGRE (A) Y EN MEDIO SELECTIVO <i>PYLORI</i> AGAR (BIOMERIEUX) (B).....	22
FIGURA 2: DENDOGRAMA DE LAS CEPAS DE <i>H. PYLORI</i> PRESENTES EN LA BASE DE DATOS BRUKER (AZUL) Y LAS INCORPORADAS A LA BASE DE DATOS LOCAL (HLP, ROJO).....	82
FIGURA 3: POTENCIA DE LA PRUEBA ESTADÍSTICA WILKONXONMANN-WHITNE PARA LAS COMPARACIONES DE LOS ÍNDICES DE SIMPSON Y SHANNON DE ACUERDO CON EL TAMAÑO MUESTRAL.	85
FIGURA 4: VALORES DE SHANNON (A) Y SIMPSON (B) DE ACUERDO CON LA BASE DE DATOS UTILIZADA (GREEN= GREENGENES 13.8, EN VERDE Y SILVA= SILVA 132, MORADO). 87	
FIGURA 5: PCoAs DE LOS PERFILES MICROBIOLÓGICOS DE LAS MUESTRAS DE BIOPSIA GÁSTRICA EN FUNCIÓN DE LA ABUNDANCIA RELATIVA DE LAS DIFERENTES ASIGNACIONES TAXONÓMICAS POR MEDIO DE LA MATRIZ DE DISTANCIA DE BRAY-CURTIS (A) Y LA PRESENCIA/AUSENCIA DE DICHAS ASIGNACIONES POR MEDIO DE LA MATRIZ DE DISTANCIAS DE JACCARD BINARIA (B) OBTENIDAS POR MEDIO DE LA BASE DE DATOS SILVA 132 (SILVA, ROJO) Y GREENGENES 13,8 (GREEN, NEGRO).	88
FIGURA 6: ABUNDANCIA RELATIVA Y COMPOSICIÓN BACTERIANA A NIVEL DE GÉNERO EN FUNCIÓN DEL ESTADO DE INFECCIÓN POR <i>H. PYLORI</i> (NEGATIVO= NEG Y POSITIVO =HPPOS) Y LAS DIFERENTES PATOLOGÍAS GÁSTRICAS ASOCIADAS DE LAS MUESTRAS DEL PROYECTO LLORCA ET AL. (2017).	93
FIGURA 7: ABUNDANCIA RELATIVA Y COMPOSICIÓN BACTERIANA A NIVEL DE GÉNERO EN FUNCIÓN DEL ESTADO DE INFECCIÓN POR <i>H. PYLORI</i> (NEGATIVO= NEG Y POSITIVO =HPPOS) Y LAS DIFERENTES PATOLOGÍAS GÁSTRICAS ASOCIADAS DE LAS MUESTRAS PROCEDENTES DEL PROYECTO PRJNA313391.....	94
FIGURA 8: ABUNDANCIA RELATIVA Y COMPOSICIÓN BACTERIANA A NIVEL DE GÉNERO EN FUNCIÓN DEL ESTADO DE INFECCIÓN POR <i>H. PYLORI</i> (NEGATIVO= NEG Y POSITIVO =HPPOS) Y LAS DIFERENTES PATOLOGÍAS GÁSTRICAS ASOCIADAS DE LAS MUESTRAS PROCEDENTES DEL PROYECTO PRJEB22107.	95
FIGURA 9: ABUNDANCIA RELATIVA Y COMPOSICIÓN BACTERIANA A NIVEL DE GÉNERO EN FUNCIÓN DEL ESTADO DE INFECCIÓN POR <i>H. PYLORI</i> (NEGATIVO= NEG Y POSITIVO =HPPOS) Y LAS DIFERENTES PATOLOGÍAS GÁSTRICAS ASOCIADAS DE LAS MUESTRAS PROCEDENTES DEL PROYECTO PRJEB21104.	96
FIGURA 10: ABUNDANCIA RELATIVA Y COMPOSICIÓN BACTERIANA A NIVEL DE GÉNERO DE LAS MUESTRAS GÁSTRICAS DEL PROYECTO PRJEB21104. EN ESTE CASO EL GRUPO CORRESPONDE A PACIENTES CON CÁNCER E INFECCIÓN POR <i>H. PYLORI</i>	97
FIGURA 11: ABUNDANCIA RELATIVA Y COMPOSICIÓN BACTERIANA A NIVEL DE GÉNERO DE LAS MUESTRAS GÁSTRICAS DEL PROYECTO PRJEB21104. EN ESTE CASO EL GRUPO CORRESPONDE A PACIENTES CON POSITIVOS PARA INFECCIÓN POR <i>H. PYLORI</i> (HPPOS) Y DIFERENTES PATOLOGÍAS GÁSTRICAS.	98
FIGURA 12: ÍNDICES DE DIVERSIDAD DE SHANNON (A) Y SIMPSON (B) EN MUESTRAS DE LOS PROYECTOS PRJNA310127 (VERDE) Y PRJNA313391 (NARANJA).	99
FIGURA 13: PCoAs DE ASIGNACIONES TAXONÓMICAS (BRAY-CURTIS) (A) Y LA PRESENCIA/AUSENCIA DE DICHAS ASIGNACIONES (MATRIZ DE DISTANCIAS DE JACCARD BINARIA) (B) EN EL PROYECTO PRJNA310127 (TRIÁNGULO, ROJO) Y PRJNA313391(CÍRCULO, NEGRO) EL VALOR EN CADA EJE REPRESENTA EL PORCENTAJE DE VARIACIÓN EXPLICADO POR ESE EJE.	100
FIGURA 14: GRÁFICO DE CORRELACIONES MIXTO DE LOS 20 GÉNEROS CON MAYOR ABUNDANCIA RELATIVA MEDIA DEL PROYECTO PRJNA313391. LOS GÉNEROS SE HAN REORDENADO POR SIMILITUD DE CORRELACIONES MEDIANTE EL MÉTODO HCLUST. LAS	

CORRELACIONES POSITIVAS SE REPRESENTAN EN COLOR AZUL Y LAS CORRELACIONES NEGATIVAS EN COLOR ROJO.	103
FIGURA 15: MEDIANA DE LOS VALORES DE DIVERSIDAD DE SHANNON (A) Y SIMPSON (B) EN LAS MUESTRAS PROVENIENTES DE PACIENTES DIAGNOSTICADOS SIN INFECCIÓN DE <i>H. PYLORI</i> (NEG, AZUL) Y POSITIVOS A INFECCIÓN POR <i>H. PYLORI</i> (HPPOS, ROJO).	106
FIGURA 16: PCoAs DE LOS PERFILES BACTERIANOS A NIVEL ASV BASADO EN LA MATRIZ DE DISTANCIA BRAY-CURTIS (A) Y LA MATRIZ DE DISTANCIA JACCARD BINARIA PARA EL ESTUDIO POR PRESENCIA/AUSENCIA (B) DE MUESTRAS DE BIOPSIAS GÁSTRICAS. EL COLOR SIGNIFICA EL DIAGNOSTICO PARA LA INFECCIÓN POR <i>H. PYLORI</i> (NEG= NEGATIVO, CIRCULO NEGRO; HPPOS= POSITIVOS, TRIÁNGULO ROJO). EL VALOR EN CADA EJE REPRESENTA EL PORCENTAJE DE VARIACIÓN EXPLICADO POR ESE EJE.	107
FIGURA 17: ESTRUCTURA DE RELACIONES ESTIMADAS BASADA EN 223 MUESTRAS DE BIOPSIA GÁSTRICA: 159 MUESTRAS DE PACIENTES DIAGNOSTICADOS POSITIVOS (A) Y 64 MUESTRAS DE PACIENTES DIAGNOSTICADOS NEGATIVOS (B). HAN SIDO REPRESENTADOS LOS 19 GÉNEROS MAYORITARIOS. LAS LÍNEAS ROJAS INDICAN CORRELACIONES NEGATIVAS Y LAS LÍNEAS VERDES CORRELACIONES POSITIVAS.	110
FIGURA 18: DIAGRAMA DE VENN DE LOS GÉNEROS IDENTIFICADOS EN MUESTRAS DE BIOPSIA GÁSTRICA DE PACIENTES DIAGNOSTICADOS POSITIVOS PARA INFECCIÓN POR <i>H. PYLORI</i> (HPPOS) Y PACIENTES DIAGNOSTICADOS COMO NEGATIVOS (NEG). EL NÚMERO EXTERIOR INDICA EL NÚMERO TOTAL DE GÉNEROS DETECTADOS.	111
FIGURA 19: DIAGRAMA DE CAJAS MOSTRANDO LA BIODIVERSIDAD, MEDIDA POR EL ÍNDICE DE SHANNON EN MUESTRAS DE BIOPSIA GÁSTRICA DE PACIENTES DIAGNOSTICADOS COMO NEGATIVOS A INFECCIÓN POR <i>H. PYLORI</i> CON LAS SECUENCIAS DE <i>H. PYLORI</i> (CON_HP_NEG, AZUL), CON LAS SECUENCIAS DE <i>H. PYLORI</i> ELIMINADAS (SIN_HP_NEG, AZUL CON ENTRAMADO) Y DE PACIENTES DIAGNOSTICADOS COMO POSITIVOS CON LAS SECUENCIAS (CON_HP_HPPOS, ROJO) Y SIN LAS SECUENCIAS (SIN_HP_HPPOS, ROJO CON ENTRAMADO).	113
FIGURA 20: PCoAs DE LOS PERFILES BACTERIANOS A NIVEL ASV BASADO EN LA MATRIZ DE DISTANCIA BRAY-CURTIS (A) Y LA MATRIZ DE DISTANCIA JACCARD BINARIA PARA EL ESTUDIO POR PRESENCIA/AUSENCIA (B) DE MUESTRAS DE BIOPSIAS GÁSTRICAS A LAS QUE SE LES HA SUSTRADO LAS SECUENCIAS DE <i>H. PYLORI</i> . EL COLOR INDICA EL DIAGNOSTICO PARA LA INFECCIÓN POR <i>H. PYLORI</i> (NEG= NEGATIVO, CIRCULO NEGRO; HPPOS= POSITIVOS, TRIÁNGULO ROJO). EL VALOR EN CADA EJE REPRESENTA EL PORCENTAJE DE VARIACIÓN EXPLICADO POR ESE EJE.	114
FIGURA 21: COMPARACIÓN DE LA ABUNDANCIA RELATIVA DE LOS GÉNEROS MÁS ABUNDANTES DE LAS MUESTRAS DE BIOPSIAS EN FUNCIÓN DE LA EXTRACCIÓN DE LAS SECUENCIAS DE <i>H. PYLORI</i> (CON SECUENCIAS = CON_; SIN SECUENCIAS = SIN_) Y EL ESTADO DE INFECCIÓN POR <i>H. PYLORI</i> (NEGATIVO = HP_NEG; POSITIVOS = HP_HPPOS).....	116
FIGURA 22: ESTRUCTURA DE RELACIONES ESTIMADAS BASADA EN 185 MUESTRAS DE BIOPSIA GÁSTRICA: 122 MUESTRAS DE PACIENTES DIAGNOSTICADOS POSITIVOS (A) Y 63 MUESTRAS DE PACIENTES DIAGNOSTICADOS NEGATIVOS (B) TRAS ELIMINAR LAS SECUENCIAS DE LAS BACTERIAS DEL GÉNERO <i>HELICOBACTER</i> DEL ANÁLISIS. HAN SIDO REPRESENTADOS LOS 19 GÉNEROS MAYORITARIOS. LAS LÍNEAS ROJAS INDICAN CORRELACIONES NEGATIVAS Y LAS LÍNEAS VERDES CORRELACIONES POSITIVAS.	117
FIGURA 23: PCoAs DE LOS PERFILES BACTERIANOS A NIVEL ASV BASADO EN LA MATRIZ DE DISTANCIA BRAY-CURTIS (A) Y LA MATRIZ DE DISTANCIA JACCARD BINARIA PARA EL ESTUDIO POR PRESENCIA/AUSENCIA (B) DE MUESTRAS DE BIOPSIAS GÁSTRICAS DE PACIENTES NEGATIVOS A INFECCIÓN POR <i>H. PYLORI</i> AGRUPADAS POR PATOLOGÍA GÁSTRICA (DISPEPSIA, CIRCULO NEGRO; GASTRITIS, TRIÁNGULO ROJO; CANCER , CRUZ	

VERDE. EL VALOR EN CADA EJE REPRESENTA EL PORCENTAJE DE VARIACIÓN EXPLICADO POR ESE EJE.	120
FIGURA 24: COMPARACIÓN DE LA ABUNDANCIA RELATIVA DE LOS GÉNEROS MÁS ABUNDANTES DE LAS MUESTRAS DE BIOPSIAS EN FUNCIÓN DE LA PATOLOGÍA DE LAS DE PACIENTES NEGATIVOS A INFECCIÓN POR <i>H. PYLORI</i>	121
FIGURA 25: DIAGRAMA DE CAJAS MOSTRANDO LA BIODIVERSIDAD, MEDIDA POR EL ÍNDICE DE SHANNON (A) Y SIMPSON (B) EN MUESTRAS DE BIOPSIA GÁSTRICA DE PACIENTES DIAGNOSTICADOS COMO POSITIVOS A INFECCIÓN POR <i>H. PYLORI</i> AGRUPADOS EN FUNCIÓN DE SU PATOLOGÍA GÁSTRICA.	123
FIGURA 26: COMPARACIÓN DE LA ABUNDANCIA RELATIVA DE LOS 20 GÉNEROS MÁS ABUNDANTES DE LAS MUESTRAS DE BIOPSIAS EN FUNCIÓN DE LA PATOLOGÍA GÁSTRICA DE PACIENTES INFECTADOS POR <i>H. PYLORI</i>	125
FIGURA 27: DIAGRAMA DE CAJAS MOSTRANDO LA BIODIVERSIDAD, MEDIDA POR EL ÍNDICE DE SHANNON, DE LOS PACIENTES NEGATIVOS O POSITIVOS A LA INFECCIÓN POR <i>H. PYLORI</i> EN FUNCIÓN DE LA PATOLOGÍA GÁSTRICA ANTES Y DESPUÉS DE SUSTRAR LAS SECUENCIAS DE LAS BACTERIAS DEL GÉNERO <i>HELICOBACTER</i>	129
FIGURA 28: COMPARACIÓN DE LA ABUNDANCIA RELATIVA DE LOS 4 FILOS MAYORITARIOS DE LAS MUESTRAS DE BIOPSIAS DE LOS PACIENTES <i>H. PYLORI</i> NEGATIVOS (A) Y <i>H. PYLORI</i> POSITIVOS (B) EN FUNCIÓN DE LA PATOLOGÍA GÁSTRICA CON SECUENCIAS DE BACTERIAS DEL GÉNERO <i>HELICOBACTER</i> (HP) Y SIN LAS SECUENCIAS (SIN_HP)	130
FIGURA 29: COMPARACIÓN DE LA ABUNDANCIA RELATIVA DE LOS 19 GÉNEROS MAYORITARIOS DE LAS MUESTRAS DE BIOPSIAS DE LOS PACIENTES <i>H. PYLORI</i> NEGATIVOS Y <i>H. PYLORI</i> POSITIVOS EN FUNCIÓN DE LA PATOLOGÍA GÁSTRICA CON SECUENCIAS DE BACTERIAS DEL GÉNERO <i>HELICOBACTER</i> (CON) Y SIN LAS SECUENCIAS (SIN).....	131
FIGURA 30: DIAGRAMA DE VENN DE LOS GÉNEROS IDENTIFICADOS EN MUESTRAS DE BIOPSIA GÁSTRICA DE PACIENTES DIAGNOSTICADOS POSITIVOS PARA INFECCIÓN POR <i>H. PYLORI</i> (HPPOS) Y PACIENTES DIAGNOSTICADOS COMO NEGATIVOS (NEG) AGRUPADAS POR PATOLOGÍA GÁSTRICA (GAST=GASTRITIS) CON UNA ABUNDANCIA RELATIVA DEL 0,01% (A) Y/O 1% (B). EL NÚMERO EXTERIOR INDICA EL NÚMERO TOTAL DE GÉNEROS DETECTADOS.....	132
FIGURA 31: DIAGRAMA DE CAJAS MOSTRANDO LA BIODIVERSIDAD, MEDIDA POR EL ÍNDICE DE SHANNON (A) Y SIMPSON (B), EN MUESTRAS DE BIOPSIA GÁSTRICA DE LOS PACIENTES POSITIVOS A LA INFECCIÓN POR <i>H. PYLORI</i> EN FUNCIÓN DEL PAÍS DE RESIDENCIA DE LOS PACIENTES. (AUS= AUSTRIA, CHN= CHINA, ESP= ESPAÑA, MYS= MALASIA, MEX=MÉXICO, GBR=REINO UNIDO).....	135
FIGURA 32: PCOAs DE LOS PERFILES BACTERIANOS A NIVEL ASV BASADO EN LA MATRIZ DE DISTANCIA BRAY-CURTIS (A) Y LA MATRIZ DE DISTANCIA JACCARD BINARIA PARA EL ESTUDIO POR PRESENCIA/AUSENCIA (B) DE MUESTRAS DE BIOPSIAS GÁSTRICAS AGRUPADAS EN FUNCIÓN DEL PAÍS DE RESIDENCIA DE LOS PACIENTES POSITIVOS A INFECCIÓN POR <i>H. PYLORI</i> (AUS= AUSTRIA, CIRCULO NEGRO; CHN= CHINA, TRIÁNGULO ROJO; ESP= ESPAÑA, CRUZ VERDE; MYS= MALASIA, EQUIS AZUL OSCURO; MEX = MÉXICO, ROMBO AZUL; GBR= REINO UNIDO, TRIÁNGULO ROSA). EL VALOR EN CADA EJE REPRESENTA EL PORCENTAJE DE VARIACIÓN EXPLICADO POR ESE EJE.....	136
FIGURA 33: COMPARACIÓN DE LA ABUNDANCIA RELATIVA DE LOS 4 FILOS MAYORITARIOS DE LAS MUESTRAS DE BIOPSIAS DE LOS PACIENTES <i>H. PYLORI</i> POSITIVOS EN FUNCIÓN DEL PAÍS DE RESIDENCIA DEL PACIENTE (AUS= AUSTRIA, CHN= CHINA, ESP= ESPAÑA, MYS= MALASIA, MEX=MÉXICO, GBR=REINO UNIDO).	137
FIGURA 34: FILOS QUE CUMPLEN LA CONDICIÓN DE ESTAR PRESENTES EN MÁS DE UN 80% DE LAS MUESTRAS CON UNA ABUNDANCIA RELATIVA DEL 0,01% (A) Y/O 1% (B) POR CADA EN LAS MUESTRAS DE BIOPSIAS DE PACIENTES POSITIVOS AGRUPADOS POR EL	

Índice

PAÍS DE RESIDENCIA DEL PACIENTE (AUS= AUSTRIA, CHN= CHINA, ESP= ESPAÑA, MYS= MALASIA, MEX=MÉXICO, GBR=REINO UNIDO).	140
FIGURA 35: GÉNEROS QUE CUMPLEN LA CONDICIÓN DE ESTAR PRESENTES EN MÁS DE UN 80% DE LAS MUESTRAS CON UNA ABUNDANCIA RELATIVA DEL 0,01% (A) Y/O 1% (B) EN LAS MUESTRAS DE BIOPSIAS DE PACIENTES POSITIVOS AGRUPADOS POR EL PAÍS DE RESIDENCIA DEL PACIENTE (AUS= AUSTRIA, CHN= CHINA, ESP= ESPAÑA, MYS= MALASIA, MEX=MÉXICO, GBR=REINO UNIDO).	142

Resumen

1. Resumen

Título

Estudio bioinformático de muestras clínicas en la infección de *Helicobacter pylori*: Importancia del microbioma

Introducción

Helicobacter pylori es una bacteria Gram negativa con forma espiral, que lleva colonizando al ser humano desde hace más de 50.000 años, actualmente infecta entre el 25 y 80% de la población mundial. Está asociada a múltiples patologías gástricas, como gastritis, úlcera y cáncer gástrico. Estas patologías no se distribuyen de manera uniforme entre poblaciones con una tasa de infección similar y se ha propuesto, como un factor clave en la evolución de las patologías gástricas, a la microbiota gástrica no-*H. pylori*. Las nuevas técnicas proteómicas y de secuenciación masiva, junto a nuevos programas bioinformáticos que nos ayudan a procesar la información que este tipo de estudios produce, nos abren un nuevo horizonte de posibilidades.

El objetivo principal de este estudio es la utilización de las técnicas proteómicas y de secuenciación masiva, junto a un programa de elaboración propia, en el análisis de la microbiota gástrica para identificar variaciones en pacientes infectados o no por *H. pylori* en función del estado de infección, patología gástrica asociada y país de residencia.

Métodos

Para los estudios proteómicos se incorporaron 15 cepas de *H. pylori*, obtenidas de pacientes de la Comunidad de Madrid, a la base de datos del MALDI Biotyper OC 3.1 tras ser debidamente identificadas. Se analizaron 133 aislamientos diferentes con la base de datos Biotyper OC 3.1 y Biotyper junto a cepas locales para analizar su implementación en un laboratorio de microbiología clínica.

Se ha preparado un manual con los comandos y las herramientas más útiles para realizar una investigación metataxonómica que incluye: ensamblaje, control de calidad y agrupación de secuencias, asignación taxonómica y análisis de la diversidad alfa, diversidad beta y taxonomía.

Resumen

El cálculo del tamaño muestral y la comparación entre bases de datos de secuencias de referencia se realizó con las secuencias obtenidas en el proyecto de Llorca *et al.* (2017) por medio del cálculo de los índices de diversidad de Shannon y Simpson y los análisis taxonómicos.

Se seleccionaron 5 proyectos de biopsias gástricas de la base de datos pública ENA que habían sido secuenciados con la plataforma Illumina y se seleccionaron las muestras que cumplieron los requisitos de poseer, al menos, 1.000 secuencias y un porcentaje de abundancia relativa de bacterias no clasificadas menor del 25%.

Se analizó la diversidad alfa, la diversidad beta y los perfiles taxonómicos de todas las muestras en función del estado de infección por *H. pylori*, la patología gástrica asociada y el país de residencia del paciente.

Resultados

La realización de los estudios proteómicos reveló una gran plasticidad proteica en las cepas de *H. pylori* en función de la región geográfica. Este hecho implica peores valores al realizar análisis proteómicos con el MALDI-TOF, mejorando significativamente al incorporar cepas locales a su base de datos.

El cálculo del tamaño muestral mostró que hay una gran variabilidad en los perfiles de microbiota gástrica dentro de los grupos de pacientes con infección por *H. pylori*, siendo necesario un número de muestras relativamente elevado para obtener un poder estadístico suficiente y aumentar, en lo posible, la resolución de las pruebas estadísticas reduciendo la probabilidad de error tipo II.

Mediante la comparación de las bases de datos SILVA 132 y Greengenes 13.8 se observó un efecto directo en los resultados obtenidos afectando a los análisis taxonómicos, siendo especialmente notables en los perfiles de diversidad beta.

Se seleccionaron muestras que cumplieran los requisitos para ser incorporados en este estudio para el análisis de la microbiota y, de un total de 631 muestras, 331 cumplieron los criterios de inclusión, finalmente, solo pudieron ser aceptadas 223 debido a una posible fuente de contaminación en uno de los proyectos (64 muestras de biopsias procedentes de pacientes diagnosticados como

Helicobacter negativo y 159 provenientes de pacientes diagnosticados como *Helicobacter* positivo).

Mediante los análisis metataxonómicos se observó un efecto en los índices de diversidad en función del estado de infección a *H. pylori* (presentaron menor diversidad bacteriana las muestras de pacientes positivos), a la patología (mayor diversidad bacteriana en los pacientes con dispepsia respecto a gastritis volviendo a ser mayor en los pacientes de cáncer) y, aunque se aprecia un efecto del país de origen, parece estar más vinculado a la patología gástrica. Los estudios de diversidad beta muestran que los perfiles gástricos se agrupan en función de la infección por *H. pylori*, la patología asociada y el país de residencia. Estas diferencias disminuyen al retirar de las muestras las secuencias pertenecientes a bacterias del género *Helicobacter*.

Conclusiones

Las técnicas proteómicas como el MALDI-TOF son buenas herramientas para la identificación de colonias de *H. pylori* teniendo en consideración la alta plasticidad de esta especie, pudiendo ser necesario preparar una base de datos de cepas locales.

Con la herramienta bioinformática desarrollada en esta tesis hemos podido reducir tiempos y gastos asociados a toda investigación metataxonómica, facilitando así el análisis de gran cantidad de información. Se observa una gran variabilidad en los perfiles de microbiota gástrica dentro de los grupos de pacientes, por lo que es necesario planificar estos estudios en consideración a la hora de realizar reclutamientos de pacientes en el futuro. También es fundamental conocer las diferentes bases de datos de referencia ya que el uso de una u otra, afecta de manera directa en los resultados obtenidos.

H. pylori tiene un efecto directo en la diversidad de la microbiota gástrica teniendo un papel modulador principal en aquellos pacientes que hayan sido infectados por esta bacteria. Sin embargo, y aunque no podamos determinar si las patologías gástricas tienen un efecto en la microbiota o la microbiota tiene efecto en las patologías, sí hemos visto que ambas están claramente relacionadas y la modulación de una puede tener efecto en la otra.

Resumen

Mediante estos análisis hemos visto que los perfiles microbianos de las muestras gástricas parecen tener un comportamiento ligeramente diferente en función de las patologías y la región geográfica. A pesar de ello se aprecia una microbiota común caracterizada, especialmente, por bacterias del género *Helicobacter*, *Prevotella*, *Escherichia* y *Streptococcus*.

Summary

Title

Bioinformatics study of clinical samples in *Helicobacter pylori* infection: Importance of the microbiome

Introduction

Helicobacter pylori is a Gram-negative spiral bacterium, which has been colonizing humans for more than 50,000 years, currently infecting between 25 and 80% of the population. It is associated with multiple gastric pathologies, such as gastritis, ulcer and gastric cancer. These pathologies are not evenly distributed among populations with a similar infection rate and the non-*H. pylori* gastric microbiota has been proposed as a key factor in the evolution of gastric pathologies. New proteomic and mass sequencing techniques, together with new bioinformatics programs, which help us to process the information of this type of studies, open a new horizon of possibilities for us.

The main objective of this study is the use of proteomic techniques and massive sequencing, together with a program of our own elaboration, in the analysis of gastric microbiota in patients infected or not by *H. pylori* depending on the state of infection, gastric pathology and country of residence.

Methods

For the proteomic studies, 15 strains of *H. pylori* obtained from patients in the Comunidad de Madrid were selected and incorporated into the MALDI Biotyper OC 3.1 database after being identified. 133 different isolates were analyzed with the database Biotyper OC 3.1 and Biotyper together with local strains to analyze their implementation in a clinical microbiology laboratory.

A manual has been prepared with the most used commands and the most useful tools to carry out a metataxonomic research that includes steps for assembly, quality control, sequence picking, taxonomic assignment and analysis of alphadiversity, betadiversity and taxonomy.

Resumen

The sample size calculation and the comparison between reference sequences databases was done using the sequences obtained in the Llorca's et al. (2017) project, calculating the Shannon and Simpson diversity indices and taxonomic analyses.

Five gastric biopsy projects were selected from the ENA public database that had been sequenced with the Illumina platform. The samples that met the requirements of having, at least, 1,000 sequences and a relative abundance of unclassified bacteria percentage less than 25% were selected.

The alphadiversity, the betadiversity and the taxonomic profiles of all the samples were analyzed according to the state of infection by *H. pylori*, the associated gastric pathology and the country of residence of the patient.

Results

The proteomic studies revealed a great protein plasticity in *H. pylori* strains depending on the same geographical region. This fact implies worse MALDI-TOF values when carrying out proteomic analysis, the values were improving significantly when local strains were added to its database.

The sample size calculation showed that there is a great variability in the gastric microbiota profiles within the groups of patients with *H. pylori* infection being necessary a relatively high number of samples to obtain enough statistical power in order to reduce the probability of type II error.

By comparing the databases SILVA 132 and Greengenes 13.8 a direct effect was observed in the taxonomic analyses, being especially remarkable in the microbiota profiles of betadiversity.

The sample selection was carry out and, from a total of 631 samples of which 331 met the inclusion criteria, finally only 223 could be accepted due to a possible source of contamination in one of the projects (64 biopsy samples from patients diagnosed as *Helicobacter* negative and 159 from patients diagnosed as *Helicobacter* positive).

By means of the metataxonomic analyses, an effect was observed in the diversity indexes due to the *H. pylori* infection status (the samples from positive patients showed a lower bacterial diversity), to the pathology (there was a higher diversity

in patients with dyspepsia than in gastritis patients and diversity was again higher in cancer patients) and, although an effect of the country is appreciated, it seems to be more linked to the gastric pathology. Betadiversity studies showed that gastric profiles were grouped according to *H. pylori* infection, pathology and country of residence. These differences decreased when sequences belonging to bacteria of the genus *Helicobacter* were removed from the samples.

Conclusions

Proteomic techniques such as MALDI-TOF are good tools for the identification of *H. pylori* colonies, taking into consideration the high plasticity of this species, but it may be necessary to prepare a database of local strains.

With the bioinformatics tool developed in this thesis we have been able to reduce time and expenses associated with all metataxonomic research facilitating the analysis of large amounts of information. A great variability in gastric microbiota profiles within patient groups is observed, so it is necessary to plan these studies in consideration when recruiting patients in the future. It is also essential to know the different reference databases since the use of one or another directly affects the results obtained.

H. pylori has a direct effect on the gastric microbiota diversity having a major modulating role in those patients who have been infected by this bacterium. However, despite we cannot determine if gastric pathologies influence the microbiota or the microbiota influence on the pathology, we have seen that both are clearly related and the modulation of one of them can produce an effect on the other.

With these analyses we have seen that the microbial profiles of the gastric samples seem to have a slightly different behavior depending on the pathologies and the geographical region. However, a common microbiota could be seen, characterized especially by bacteria of the genus *Helicobacter*, *Prevotella*, *Escherichia* and *Streptococcus*.

Introducción

2. Introducción

En esta tesis nos encontramos con tres historias que, por diversos motivos, han cambiado la vida de millones de personas.

La primera historia transcurre en el estómago, un ambiente considerado hostil, donde el descubrimiento de *Helicobacter pylori* (*H. pylori*) y su implicación en la patología gastroduodenal supuso un hito en los campos de la Microbiología y la Gastroenterología, tanto es así que sus descubridores Robin Warren y Barry Marshall fueron reconocidos con el Premio Nobel de Medicina en el año 2005.

Sin embargo, las comunidades médicas y científicas se encontraron con una serie de problemas que no habían previsto. Se tuvieron que enfrentar a un nuevo paradigma: poner en marcha un plan que consistió en cambiar radicalmente la relación entre departamentos médicos, modificar el sistema de tratamiento de las úlceras gástricas y dar un giro al tratamiento de las enfermedades gastroduodenales, en general.

Ese ecosistema único que somos cada uno de nosotros, las constantes “luchas y alianzas” entre los microorganismos por ocupar su “espacio” y cómo somos capaces de interpretar y modificar ese mundo microscópico es la segunda historia: el microbioma.

Con la tercera historia nos armamos con las herramientas necesarias para poder realizar la “exploración”: la bioinformática.

2.1 *Helicobacter pylori*

2.1.1 Antecedentes históricos y situación taxonómica actual

H. pylori lleva colonizando al ser humano desde hace más de 50.000 años (Atherton & Blaser, 2009) y, actualmente, infecta entre el 25 y 80% de la población mundial dependiendo del índice de desarrollo del país de estudio (Yue Hu *et al.*, 2016) gracias a su capacidad de resistir el ambiente ácido del estómago (Chey *et al.*, 2017). No es una bacteria con un origen reciente, su historia está

Introducción

Íntimamente ligada a la del ser humano, su único huésped conocido, anterior a la expansión de los humanos desde África (Linz *et al.*, 2007). A pesar de estar tan ampliamente extendida entre la población, la primera observación de la bacteria no se realizó hasta 1875 y, debido a que no pudo ser cultivada estos resultados fueron olvidados (Blaser, 2005). No fue hasta el año 1982 que unos científicos australianos pudieron aislar la bacteria y comenzar así el estudio de su rol en el estómago (Marshall & Warren, 1984).

De entre toda la población afectada, solo algunos desarrollan alguna sintomatología y solo en esos casos se recomienda el tratamiento (Malfertheiner *et al.*, 2017).

Normalmente, la infección por *H. pylori* origina gastritis subclínicas y en mucha menor medida se desarrollan otras complicaciones (úlcera péptica, adenocarcinoma y linfoma gástrico) en las personas infectadas. Si nos focalizamos en la población pediátrica, esas complicaciones se presentan en un porcentaje todavía menor, variando de país en país entre 2 al 5% en los países industrializados, a un 13,5 al 20% en países en vías de desarrollo y con una alta prevalencia de infección por *H. pylori* (Hernandez *et al.*, 2014; Roma, Kafritsa, Panayiotou, Liakou, & Constantopoulos, 2001).

H. pylori no es la única especie del género *Helicobacter*, en la actualidad hay reconocidas más de 20 especies destacando, por su posible papel patógeno en humanos, *H. heilmannii*, *H. fennelliae* y *H. felis* (este último presente en animales), se han aislado de manera infrecuente en humanos con gastritis habiendo indicios de que también podrían estar implicados en cáncer y linfoma gástricos (Solnick, O'rourke, Vandamme, & Lee, 2006).

2.1.2 Morfología y características fisiológicas del género *Helicobacter*

Con un tamaño que oscila entre 0,5 y 1 μm de ancho y 2,5 y 4 μm de largo *H. pylori* es un bacilo Gram negativo con forma de espiral o ligeramente curvado, pudiendo cambiar y adoptar la forma cocoide dependiendo de factores ambientales (Sisto, Brenciaglia, Scaltrito, & Dubini, 2000) o de una degeneración

celular (J. G. Kusters *et al.*, 1997). Se ayuda de sus 4 a 8 flagelos polares para desplazarse helicoidalmente.

H. pylori ha ido reclasificándose con el paso de los años conforme más conocíamos sobre la bacteria. Originalmente, Marshall y Warren la clasificaron dentro del género *Campylobacter* bajo el nombre de *Campylobacter pyloridis* (Goodwin, Armstrong, & Marshall, 1986). Posteriormente, debido a problemas gramaticales fue corregido a *Campylobacter pylori* en 1987 (Konturek, 2003). En 1989 se estableció el género *Helicobacter* (vara en espiral), en el que fueron clasificadas otras especies junto a *H. pylori* perteneciendo al filo Proteobacteria. En 1992 se describió la subclase *Epsilonproteobacteria* (Tenover *et al.*, 1992) y se le asignó el estatus de clase en 2005 (Garrity, Bell, & Lilburn, 2006). En el año 2017 Waite postuló la necesidad de separar la clase *Epsilonproteobacteria* del filo *Proteobacteria* debido al gran número de bacterias que no podían afiliarse a este filo, proponiendo uno nuevo denominado *Epsilonbacteraeota* (Waite *et al.*, 2017). Esta denominación se encuentra en algunas versiones de las bases de datos realizadas entre los años 2017 y 2018. Un *addendum* posterior terminó reclasificándolo finalmente con el nombre de *Campylobacterota* (Waite *et al.*, 2018) siendo éste el aceptado actualmente (Waite, Chuvochina, & Hugenholtz, 2019).

Por lo tanto, la taxonomía actual incluye a esta bacteria en: Filo *Campylobacterota*; Clase *Campylobacteria*; Orden *Campylobacterales*; Familia *Helicobacteraceae*; género *Helicobacter*.

Las especies del género *Helicobacter* presentan una gran homogeneidad en cuanto a sus características bioquímicas, son quimiorganotrofas y tienen un metabolismo respiratorio. No hay fermentación ni oxidación de azúcares más allá de la oxidación de la glucosa.

No presentan la vía del glioxilato y tienen, al menos parcialmente, las vías metabólicas Entner-Doudoroff, de pentosas fosfato, y el ciclo de ácidos tricarboxílicos. No hidrolizan la gelatina, almidón, caseína o tirosina, son rojo de metilo y Voges-Proskauer negativos.

Introducción

Las tres pruebas bioquímicas más características y útiles para la identificación de *H. pylori* son: la catalasa, la oxidasa y, sobre todo, la ureasa. Todas ellas positivas y siendo la última especialmente rápida.

Aunque a nivel bioquímico presenta una alta homogeneidad, no ocurre lo mismo a nivel genético, donde se ha determinado la presencia de profagos en su genoma así como una gran variabilidad antigénica.

El ambiente ácido del estómago lo convierte en un nicho pobre en variedad bacteriana pero rico en subpoblaciones de *H. pylori*, tanto es así que no es raro encontrar en un mismo individuo diferentes cepas de *H. pylori* (Taylor *et al.*, 1995). Esta bacteria posee una gran capacidad de mutación y de recombinación, existiendo repeticiones no aleatorias de nucleótidos en su genoma que facilitan los fenómenos de recombinación (Aras, Kang, Tschumi, Harasaki, & Blaser, 2003) además de presentar un sistema de transporte tipo IV (el sistema ComB) capaz de importar ADN externo e integrarlo mediante recombinación (Karnholz *et al.*, 2006).

2.1.3 Genoma de *Helicobacter pylori*

El primer genoma completo de una cepa de *H. pylori* fue secuenciado por primera vez en 1997 (Tomb *et al.*, 1997), desde entonces, las mejoras conseguidas en el campo de la secuenciación nos han permitido disponer de un número, cada día mayor, de aislamientos. También se han mejorado los algoritmos informáticos utilizados para analizar las secuencias de nucleótidos permitiendo el solapamiento de muchas de ellas en largos cóntigos (Zhang, Chiodini, Badr, & Zhang, 2011). Al menos los genomas de 170 cepas de *H. pylori* han sido ya publicados (Lara-Ramírez *et al.*, 2011) (Muñoz-Ramírez *et al.*, 2017). El estudio y comparación de estos genomas nos ha permitido conocer mejor muchos aspectos de la genética, microbiología y patogenicidad de este microorganismo. El tamaño de los genomas secuenciados está entre 1,5 y 1,7 Mpb y tiene un contenido de G+C entre el 35% y el 40%. El genoma común o *core* entre las cepas estaría compuesto por 1.186 genes e incluye todos los genes relacionados con la adaptación y supervivencia al ambiente gástrico. Entre el 22% y el 27% son los genes auxiliares y patogénicos, mucho más variables entre las diferentes

cepas (Gressmann *et al.*, 2005). Entre estos, encontramos las islas de patogenicidad como Cag-PAI o genes para proteínas de membrana. Se pueden encontrar plásmidos en diferentes cepas de *H. pylori*; pero, su función aún no ha podido ser comprobada (Heuermann & Haas, 1995).

2.1.4 Factores de virulencia

Los principales factores de virulencia que le aportan su carácter modulador son: sus enzimas (principalmente la enzima ureasa y proteasas) y sus antígenos (VacA y CagA).

La ureasa es la enzima más característica de *H. pylori* y la que produce en más cantidad, rondando el 5% de la producción total de proteínas celulares (Rivas & Hernandez, 2012). Es una metaloenzima que produce dióxido de carbono y amoníaco a partir de la urea que se encuentra en el jugo gástrico. La liberación del amoníaco aumenta el pH gástrico hasta alcanzar un pH de entre 6 y 7 creando así un ambiente más básico entorno a *H. pylori*, momento en el cual la enzima se autorregula dejando de producir más amoníaco (McNulty, Ulmschneider, Luecke, & Ulmschneider, 2013).

Como curiosidad, la urea fue la primera molécula orgánica sintetizada publicada en 1828 mientras que la ureasa fue la primera enzima cristalizada en 1926 siendo ambas un hito en la biología molecular (H. L. Mobley, Mendz, & Hazell, 2001).

El segundo grupo de enzimas son las proteasas, las cuales tienen un efecto indirecto sobre la modulación de la virulencia al encargarse de deshacer gradualmente la estructura polimérica del mucus provocando una pérdida de la viscosidad del mucus y haciéndolo más permeable a las bacterias. Esta pérdida de viscosidad afecta directamente a la estructura del mucus, provocando el contacto del jugo gástrico con las células de la pared gástrica.

Entre las toxinas producidas por *H. pylori*, la toxina VacA es producida por aproximadamente el 50% de las cepas, por lo que no es imprescindible para su colonización. La proteína VacA es una toxina codificada por el gen *vacA*, esta

Introducción

induce vacuolización en las células epiteliales, la muerte celular y la destrucción de la integridad epitelial (Salama, Hartung, & Müller, 2013). Esta vacuolización viene causada por un aumento en la carga de moléculas alcalinas, mediada por el ensamblado de la proteína VacA en la bicapa lipídica de las células del hospedador, lo que provoca un desequilibrio de la presión osmótica, pudiendo provocar la citólisis de las células gástricas. La pérdida de la integridad epitelial podría ser utilizada por microorganismos para una posterior colonización.

Siguiendo con las proteínas antigénicas, la proteína CagA (*citotoxin-associated gen A*), es uno de los factores de virulencia de *H. pylori* más extensamente estudiados. El gen *cagA* está incluido dentro de la isla de patogenicidad denominada *cag* (*cag*-PAI) y se ha sugerido que la proteína asociada al gen A (*cagA*) con sus productos, está vinculada a una mayor patogenicidad de las cepas de *H. pylori*. La citotoxina CagA tiene un carácter pleiotrópico siendo capaz de perturbar múltiples vías de señalización de las células del hospedador actuando sobre la ruptura de las uniones intercelulares, en la desregulación del ciclo celular y de apoptosis lo que terminará favoreciendo la transformación de las células epiteliales gástricas en células malignas.

El porqué *H. pylori* expresa proteínas destinadas finalmente a la muerte del hospedador puede resultar chocante a primera vista. No podemos olvidar que *H. pylori* se suele establecer en las etapas tempranas de la vida del hospedador mientras que los casos de cáncer gástrico suelen ser pacientes adultos mayores de 50 años. En las etapas tempranas de la infección, la proteína CagA va a provocar un estrés oncogénico lo que dará lugar a una apoptosis mediada por las cadenas de reacción de las ARF/p53/p21 o p16-pRB (Beausejour & Campisi, 2006; Sherr & Weber, 2000) reduciendo el número de células epiteliales, lo que reduce la secreción ácida aportándole una ventaja competitiva pudiendo afectar a la composición de la microbiota gástrica. Mientras las células tengan un correcto manejo de la senescencia derivará a una gastritis o ulceración; pero, con el paso del tiempo puede causar inestabilidad genética, pudiendo aumentar la capacidad de las células en adquirir resistencia al estrés oncogénico. Así pues, si una célula falla en desencadenar la senescencia puede desencadenar una respuesta pro-oncogénica haciendo inevitable que, con el paso del tiempo, deriven o aparezcan células cancerígenas (Hatakeyama, 2014).

2.1.5 Patogenicidad

La colonización por *H. pylori* viene mediada por los factores de virulencia, los cuales facilitan su resistencia al ambiente gástrico, la adhesión a la pared gástrica y modulación de las células gástricas en su beneficio. Estos factores afectarán a la dinámica propia del estómago y favorecerán la aparición de una serie de patologías como disminución del pH gástrico, dispepsia, inflamación de la mucosa gástrica aguda y/o crónica, úlceras gástricas, cáncer o linfoma gástricos tipo MALT.

a) Dispepsia

La dispepsia es esa sensación de ardor o malestar en la parte superior del abdomen o la parte inferior del pecho de origen generalmente desconocido por el paciente. También es posible que la causa se deba a problemas psicológicos, tales como el estrés, la ansiedad, la depresión o problemas de sueño. Esta sintomatología es la principal por la que los pacientes acuden al médico en búsqueda de un diagnóstico.

b) Gastritis

El inicio de la colonización por *H. pylori* conduce a una respuesta inmune mediada por neutrófilos y monocitos que se infiltran en la mucosa y el cuerpo gástrico. Según lo reciente que sea la colonización se pueden distinguir dos formas de gastritis (Johannes G. Kusters, Van Vliet, & Kuipers, 2006):

b.1) Gastritis aguda

La gastritis aguda producida por *H. pylori* generalmente cursa asintomática para el paciente. Todos los individuos infectados presentan manifestaciones histológicas con una inflamación de la mucosa gástrica proximal y distal y solamente una minoría presentará síntomas inespecíficos como la dispepsia, pudiendo cursar con náuseas y/o vómitos.

b.2) Gastritis crónica

Cuando la colonización se vuelve persistente, se alcanza un equilibrio entre la producción de ácido, que afecta negativamente al crecimiento bacteriano, y los factores de virulencia de la bacteria, asociados a la inflamación de la mucosa

Introducción

gástrica y daño sobre las propias células parietales. Este equilibrio suele localizarse en la región del antro gástrico, donde hay una menor carga de células secretoras y la infección suele ir acompañada de pocas bacterias en la superficie.

En el caso en el que, por la medicación, bacterias o causas externas, la producción de ácido se encuentre comprometida en el paciente, la bacteria se propagará de manera mayoritariamente uniforme derivando a una pangastritis aumentando la hipoclorhidria del ambiente gástrico (Kuipers *et al.*, 1995).

c) Úlcera péptica

Las úlceras pépticas son pequeñas lesiones con un diámetro de, al menos, 0,5 cm que penetra a través del estrato de fibras musculares que forma parte de la mucosa gástrica. Estas úlceras están generalmente asociadas con *H. pylori* siendo esta bacteria la responsable de aproximadamente el 90% de las úlceras, tanto duodenales como gástricas (Metzger *et al.*, 2001).

d) Carcinoma gástrico

La inflamación crónica causada por *H. pylori* puede provocar la pérdida de la estructura normal de la mucosa gástrica con la pérdida de glándulas secretoras y la aparición de fibrosis. Este hecho conlleva la aparición de metaplasia intestinal, la cual consiste en un cambio fenotípico de las células gástricas a una estructura de fenotipo intestinal. Estas áreas afectadas se irán extendiendo con el tiempo aumentando la probabilidad de cáncer gástrico entre 5 y 90 veces dependiendo de la extensión y severidad de la atrofia (Slipponen, Kekki, Haapakoski, Ihamäki, & Siurala, 1985).

En base a los numerosos estudios que relacionan la colonización de *H. pylori* y su relación con el cáncer gástrico la Organización Mundial de la Salud (OMS) categorizó a la bacteria *H. pylori* como carcinógeno de clase 1 (Moller, Heseltine, & Vainio, 1995).

e) Linfoma gástrico del tejido linfoide asociado a mucosas (MALT)

El linfoma tipo MALT, llamado también linfoma de tejido linfoide asociado a mucosa es un tipo de cáncer incluido dentro del grupo de los linfomas no Hodgkin. El caso del linfoma gástrico está casi siempre asociado a la infección por *H. pylori*. El tratamiento frente a *H. pylori* ha demostrado una completa

remisión del linfoma entre el 60 y 80% de los pacientes frente al 10% que presentaron mejorías y el resto de los pacientes que no mostraron respuesta frente al linfoma mediante el tratamiento de *H. pylori* (Johannes G. Kusters *et al.*, 2006).

2.1.6 Métodos de diagnóstico

Es fundamental conocer a que pacientes hay que realizar el diagnóstico y ofrecer tratamiento a todos aquellos que tuvieron una prueba positiva de infección activa (Chey *et al.*, 2017).

Como hemos visto, *H. pylori* está muy relacionada con la úlcera péptica (duodenal o gástrica) y las gastritis atróficas, pudiendo ser el primer paso para la evolución a cáncer gástrico. También puede desarrollarse un tipo de linfoma, poco frecuente, que es el linfoma gástrico tipo MALT. En todos estos pacientes se recomienda realizar una prueba de diagnóstico de *H. pylori* (Chey *et al.*, 2017; Malfertheiner *et al.*, 2017).

Los métodos diagnósticos pueden ser catalogados en dos grandes grupos: Métodos invasivos (dependientes de la realización de una endoscopia y toma de biopsia gástrica) o métodos no invasivos (no requieren de la toma de la biopsia). La elección del método de diagnóstico dependerá esencialmente de la experiencia del centro donde se diagnostique, el paciente y la epidemiología de la infección en el área, ya que todos los métodos aportan una serie de ventajas e inconvenientes frente a los otros (Miqueleiz *et al.*, 2020).

Aunque el método diagnóstico ideal sería aquel mínimamente invasivo para el paciente, barato, seguro, sensible, preciso y fácilmente practicable en los centros de diagnóstico ninguno de los actuales cumple con todas estas características. En el caso de los métodos invasivos el paciente ha de someterse a una intervención quirúrgica, siendo el único método eficaz para el diagnóstico por cultivo, fundamental para casi todos los estudios de sensibilidad y aportando otros datos muy relevantes para el clínico, como el estado histológico o conocer las lesiones de la mucosa para poder clasificar y diagnosticar la gastritis.

a) Métodos invasivos

En caso de realizarse un procedimiento invasivo se recomienda la toma de una biopsia de antro gástrico y otra de cuerpo gástrico (Lan, Chen, Li, Chang, & Lin, 2012).

Los métodos invasivos son necesarios para el aislamiento de la cepa de *H. pylori* siendo éste el método más específico para la detección y/o su diagnóstico. El mayor problema reside en su sensibilidad, la cual varía en gran medida dependiendo de diversos factores: la toma de la muestra, el transporte, el almacenamiento, los medios de cultivo utilizados, las habilidades del personal técnico y las condiciones de incubación (Miqueleiz *et al.*, 2020). Las mejoras en estos aspectos y, en especial, en el medio de cultivo, son fundamentales para mejorar la sensibilidad de la técnica (López-Brea *et al.*, 2004).

En los últimos 30 años se ha ido mejorando el medio de cultivo para *H. pylori*. La base de todos ellos suele ser sangre (Fig. 1-A) o suero (Fig. 1-B), un ambiente de cultivo microaerofílico, una alta humedad y tiempos de incubación que pueden llegar hasta los 10-15 días (Ndip, MacKay, Farthing, & Weaver, 2003). El uso de antibióticos no es extraño para evitar posibles contaminaciones en la toma de la muestra, siendo la vancomicina, trimetoprima, sulfametoxazol, bacitracina o anfotericina, en solitario o en combinación, los antibióticos más utilizados.

A



B



Figura 1: *H. pylori* cultivado en medio de Columbia agar sangre (A) y en medio selectivo *pylori* agar (biomerieux) (B).

Otras técnicas que también requieren de una endoscopia para poder realizarse son la prueba de la ureasa o diversas pruebas moleculares. La primera se basa en la producción de la enzima ureasa por *H. pylori*. La biopsia se pone en contacto con un medio con urea y un indicador de pH; en caso de haber actividad ureasa se hidroliza la urea en dióxido de carbono y amoníaco. Esta prueba presenta unos índices de sensibilidad y especificidad de en torno a 80-90% (Naumov & Fenjvesi, 2011).

En el caso de las técnicas moleculares se encuentran tanto técnicas invasivas como no invasivas dependiendo del tipo de muestra a analizar, la biopsia gástrica es la más común, aunque también se pueden realizar sobre muestras de heces, saliva, agua, etc. (Simala-Grant & Taylor, 2004). La técnica molecular más habitual es la reacción en cadena de la polimerasa (PCR) en todas sus versiones (convencional o a tiempo real) tanto para la detección del microorganismo como para la detección de genes que confieran resistencia a antibióticos.

b) Métodos no invasivos

Las técnicas no invasivas se recomiendan en la mayoría de los casos de diagnósticos y, especialmente, en los pacientes adultos con síntomas de dispepsia no estudiada con no más de 60 años y sin características que indiquen una alarma al médico (Chey *et al.*, 2017).

b.1) Prueba del aliento

Es el método de detección no invasivo más ampliamente utilizado (Graham & Miftahussurur, 2018; A. Miqueleiz-Zapatero *et al.*, 2020). Se trata de un método de detección indirecto ya que no detecta *H. pylori* sino su actividad ureasa por lo que puede haber falsos positivos por la presencia de *Helicobacter heilmannii* en el estómago (Y. K. Wang *et al.*, 2015) o bacterias con actividad ureasa en la boca (Lopes, Vale, & Oleastro, 2014). La técnica precisa de la ingesta por parte del paciente de urea marcada con un isótopo de carbono 13. La actividad ureasa de *H. pylori* romperá la urea marcada absorbiéndose y difundiéndose en la sangre hasta llegar a los pulmones liberándose en el aliento. Este método sirve para la detección de una infección actual, por lo que su uso está recomendado para hacer el seguimiento de los tratamientos utilizados en la erradicación de *H. pylori* (Perri *et al.*, 2005).

b.2) Detección del antígeno de *H. pylori* en heces

Este método detecta antígenos de *H. pylori* en heces por lo que la obtención y conservación de la muestra es fácil. Actualmente hay diversos kits comerciales diferenciados por el uso de anticuerpos, monoclonales o policlonales, y por el método por el que se llevan a cabo, ELISA o inmunocromatográfica. Es un buen método de diagnóstico inicial y de seguimiento después de 4 a 6 semanas desde la erradicación de *H. pylori* pudiendo ser utilizado para confirmar la aparición de nuevas infecciones.

Los principales inconvenientes son la variabilidad de la sensibilidad y la especificidad en función de la población en la que se utilice debido a las diferencias antigénicas de diferentes cepas de *H. pylori* (Ritchie *et al.*, 2009) por lo que deberá ser testado en cada población (Shimoyama, 2013).

b.3) Serología

Las técnicas serológicas tienen su base en la detección de anticuerpos específicos frente a *H. pylori* en suero. Están especialmente diseñadas para estudios epidemiológicos debido a la dificultad de diferenciar una infección activa de individuos con una exposición previa. A diferencia de las técnicas no invasivas previas, no se ve afectada por el uso de antibióticos o antiácidos.

Se sustenta en la respuesta inmunológica del paciente frente a la presencia de *H. pylori* a nivel sistémico por un aumento transitorio de IgM seguido de un aumento de IgG e IgA más duradero. La efectividad de la técnica depende pues de la respuesta del paciente la cual suele ser menor en pacientes pediátricos menores de 12 años (Malfertheiner *et al.*, 2007).

2.1.7 Tratamiento

El tratamiento para *H. pylori* solo está recomendado en aquellos pacientes con sintomatología, no recomendándose las pruebas indiscriminadas ni el tratamiento de los pacientes positivos sin síntomas (Malfertheiner *et al.*, 2017).

Actualmente, solo 6 antibióticos en terapia combinada están disponibles para el tratamiento de erradicación de *H. pylori*: claritromicina, amoxicilina, metronidazol, levofloxacina, tetraciclina y rifabutina. Las diferentes pautas de tratamientos dependerán de la combinación de los antibióticos utilizados y el uso o no de inhibidores de bomba de protones (IBP) o sales de bismuto.

Las pautas más utilizadas actualmente son la combinación de IBP (Inhibidor de Bomba de Protones), claritromicina y amoxicilina, conocida como la terapia triple estándar, o la cuádruple terapia que incorpora sales de bismuto, un IBP y dos de los citados antibióticos.

La toma de antibióticos puede ser secuencial (5 días de amoxicilina con IBP y 5 días de IBP con claritromicina y metronidazol), concomitante (10 días de IBP con amoxicilina, claritromicina y metronidazol) o tratamientos híbridos (7 días de amoxicilina con IBP y 7 de IBP con amoxicilina, claritromicina y metronidazol) (Kavitt & Cifu, 2017).

A la hora de plantear un tratamiento es importante contar con el perfil de resistencia antimicrobiana de cada paciente ya que podremos suministrar el régimen antibiótico más adecuado logrando un tratamiento más eficaz con menos efectos secundarios (Alba, Blanco, & Alarcón, 2017) y menor probabilidad de aparición de cepas resistentes.

No podemos olvidar que el uso de varios antibióticos incrementa el riesgo de efectos secundarios en los pacientes (Chey *et al.*, 2017) y la presencia de mecanismos de resistencia en *H. pylori* afecta a la eficacia del tratamiento antimicrobiano para perjuicio de éstos. Este último es un problema creciente que se ha visto incrementado por el aumento del uso indiscriminado de antibióticos en las últimas décadas.

2.2 Estudio del microbioma y sus limitaciones

A pesar de que los organismos unicelulares fueron las primeras formas de vida en encontrarse de manera omnipresente en “casi” todos los posibles ambientes que pueden darse en la tierra, no fueron descubiertos hasta 1677. Un comerciante de telas llamado Antony van Leeuwenhoek fue el primero en realizar observaciones de “animálculos” y desde este punto despegó el estudio de los diferentes microorganismos. En esa época también se empezó a ver una cierta relación entre los microorganismos con el ser humano cuando se empezaron a distinguir microorganismos en la placa dental. Actualmente se estima que el número total de bacterias, solo en el océano, es de $1,3 \times 10^{29}$ y aunque tenemos tendencia a verlas desde un punto de vista patogénico existen solo 1.400 especies patogénicas para el ser humano (incluyendo virus, bacterias, hongos, protozoos y helmintos). Este número representa menos del 1% del total de especies del planeta (“Microbiology by numbers,” 2011) .

Si nos centramos en el cuerpo humano encontramos que la mayoría de bacterias se localizan en el colon con una estimación de $0,9 \times 10^{11}$ bacterias/g heces húmedas que a un volumen medio, en una persona adulta de 70 kg, de 480 mL hacen $3,9 \times 10^{13}$ (Sender, Fuchs, & Milo, 2016b). En el resto del cuerpo humano se estima un número de bacterias de, como mucho, 10^{12} dándonos un ratio 1:1 de células bacterianas con células humanas (Sender, Fuchs, & Milo, 2016a).

No es de extrañar que, con una relación tan estrecha entre los microorganismos y nosotros, no surjan diferentes relaciones tanto de comensalismo, como de parasitismo; pero, sin lugar a duda, también hay una clara tendencia a relaciones simbióticas, las cuales han ido coevolucionando conforme nos hemos ido desarrollando y adaptando como especie. Esta coevolución nos ha proporcionado una serie de beneficios fisiológicos como pueden ser a la digestión (Jandhyala *et al.*, 2015), a la mejora del sistema inmune (Belkaid, 2015) y a la protección frente a patógenos (Ubeda, Djukovic, & Isaac, 2017) entre otros.

Todos estos beneficios no se pueden atribuir a una única especie en particular, sino que es el resultado de una compleja interacción de todos los microorganismos tanto entre sí como con el hospedador.

Descubrir esas relaciones, identificar todas y cada una de las partes de las diferentes vías de interacción y, finalmente, una posible modulación de las relaciones individuales, serán los diferentes puntos de partida de todos los estudios de microbioma.

Las técnicas de secuenciación masiva fueron desarrolladas en la década de los 90 para suplir las limitaciones de las técnicas clásicas dependientes de cultivo, permitiéndonos obtener un mejor conocimiento de las comunidades microbianas. Estas técnicas nos han aportado una ingente cantidad de información y ha surgido la necesidad de herramientas potentes y prácticas que nos ayuden a estudiar las complejas interacción y relaciones del mundo microbiológico. Con ello en mente, surgieron las denominadas herramientas “bioinformáticas” que en los últimos 20 años, han realizado importantes contribuciones a la ciencia, tanto en forma de reservorios de datos como en capacidad de análisis. Las investigaciones clínicas y biológicas actuales son un claro ejemplo de ello, siendo solo el prelude de lo que está aún por llegar.

2.2.1 Revisión de definiciones

Desde que en 1920 fuera acuñada la palabra *genoma* por el botánico alemán Hans Winkler para referirse a la totalidad de genes presentes en una célula, el sufijo “-oma” ha ido derivando su significado. El sufijo “-oma” proviene del término griego “-*ōma*” que significa “resultado de un proceso” o “tumor” (tenemos muchos ejemplos de ello con palabras como el adenoma o el linfoma siendo un tumor de las glándulas o un tumor de los tejidos linfáticos, respectivamente) que derivó a masa o conjunto de algo.

Este neologismo ha sido utilizado desde su acuñación por la biología molecular derivando al sufijo “-ómico/a” para referirse al estudio de ese conjunto, como es el ejemplo de la genómica, estudio del conjunto de genes o proteómica para referirse al estudio de las proteínas.

Esto, junto al uso de la informática, de los ordenadores y de los programas que nos ayudan en el día a día ha ido calando en nuestra sociedad y se encuentra tan unida a la rama clínica y la ciencia que no concebimos una sin la otra, como

Introducción

si hubieran ido de la mano desde el principio de los tiempos, cuando, en realidad, no fue así.

Podemos remontarnos a muchos eventos históricos: cuando se acuñó por primera vez la palabra “genoma” en 1920 o cuando en 1953 Watson y Crick propusieron la estructura de doble hélice del ADN (Ubeda *et al.*, 2017) o tal vez cuando se secuenció la primera proteína (insulina bobina) por F. Sanger en 1955 (Sanger, 1959). Todo ello sin siquiera la existencia del primer circuito integrado, ideado por Jack Kilby en los laboratorios de Texas Instruments en 1958.

Las herramientas han ido evolucionando y la ciencia con ellas, hasta el punto de que no solo somos capaces de conocer el genoma de una única célula, sino que somos capaces de conocer todo el material genético de una muestra de estudio con un número muy elevado de células. A este tipo de estudios se les denomina metagenómicos, siendo el estudio de todo el material genético de todas las células de la muestra, incluyendo plásmidos, y destacando el potencial genético de la población. Podemos estudiar el ARN total transcrito por medio de los estudios metatranscriptómicos, todos los metabolitos producidos por medio de los estudios metametabólicos o a nivel proteico con los estudios metaproteómicos.

Podemos realizar búsquedas selectivas de genes como pueden ser los genes de resistencia y realizar un estudio de resistoma o utilizar un gen marcador que sea lo suficientemente diferente entre cada especie de microorganismo como puede ser gen 16S ARNr y realizar estudio metataxonómicos. El siguiente salto es conocer no solo los microorganismos sino todos sus metabolitos y el ambiente de la muestra, estos serían los estudios de microbioma.

Es necesario definir qué es una especie bacteriana ya que, aunque la pregunta surgió en el siglo XIX, no fue hasta 1987 cuando se llegó a un primer consenso: una especie bacteriana es una colección de cepas que se caracterizan por compartir una similitud del 70% o más en experimentos de hibridación ADN-ADN con una variación de temperatura no superior a 5 grados (Wayne *et al.*, 1987). La definición fue ampliada y se recomienda la identificación polifásica, utilizando criterios fenotípicos a la par que la secuenciación (Stackebrandt *et al.*, 2002).

Actualmente se añadió la similitud media de sus genes compartidos en un 94-96% y una identidad del gen 16S de en torno a 98,7% (Richter & Rosselló-Móra, 2009).

2.2.2 Microbioma humano

Como “seres humanos” somos más un ecosistema que un “ser”. Se estima una relación de 23.000 genes humanos por más de 3 millones de genes bacterianos. Ya desde antes del nacimiento vamos formando nuestro “pequeño ecosistema” donde hay translocación de las bacterias desde la placenta al embrión “no nato” (Gomez-Arango *et al.*, 2017). Pero sin lugar a duda es en el nacimiento donde nos colonizamos con una gran variedad de microorganismos (Favier, De Vos, & Akkermans, 2003) y podemos empezar a utilizar el término de microbiota. Así pues, la microbiota humana sería todo el conjunto de microorganismos que residen en nuestro cuerpo (tantos hongos, virus, parásitos, arqueas y, por supuesto, bacterias). Esta relación de los microorganismos con nosotros y nosotros con ellos dependerá de las diferentes regiones anatómicas y condiciones fisiológicas, dándonos diferentes perfiles de microbiota que se han adaptado a cada una de ellas (B. Wang, Yao, Lv, Ling, & Li, 2017).

2.2.3 Funciones de la microbiota

De todos los posibles nichos, es en el intestino donde la microbiota, por su gran importancia funcional/operacional para el hospedador, es considerada como “el órgano olvidado” (O’Hara & Shanahan, 2006) y ha atraído el interés de la comunidad científica. Por poner unos ejemplos donde la microbiota desempeña un papel relevante tenemos: la digestión de restos de fibra con un fuerte impacto en la salud del hospedador (Makki, Deehan, Walter, & Bäckhed, 2018), la liberación de metabolitos beneficiosos para el hospedador (Martinez, Leone, & Chang, 2017), la protección contra microorganismos patógenos (tanto por exclusión competitiva como por la producción de metabolitos) y regulación de la inmunidad (Ubeda *et al.*, 2017).

Introducción

Este hecho está cambiando el paradigma de la microbiología clínica donde un microorganismo es el causante de una enfermedad infecciosa. Con el tiempo se ha vuelto claro que algunas bacterias comensales pueden proteger a su hospedador por ellas solas, lo que se conoce como “resistencia por colonización” gracias a la producción de metabolitos, mejorando el sistema inmunitario del hospedador (Buffie & Pamer, 2013), mientras que otras veces lo hace en función de toda la comunidad microbiana “resistencia por comunidad”, siendo este último el origen de los trasplantes fecales (del Campo, Alarcón, D’Auria, Delgado, & Ferrer, 2018). Con todo ello, se empieza a tratar la enfermedad microbiana como una perturbación del estado de equilibrio de la microbiota, que sería el estado de salud. A esa perturbación la denominamos disbiosis y para conocer la extensión de esa disbiosis no bastaría solo el cultivo, aislamiento y caracterización de los microorganismos patógenos, sino que sería necesario caracterizar y modular toda la microbiota del paciente para devolverla a su estado de equilibrio.

2.2.4 Técnicas para el estudio de la microbiota

Intentar imponer el cultivo de todos los microorganismos como método de estudio, a parte del problema del tiempo y los recursos necesarios, es técnicamente imposible. Lamentablemente, la inmensa mayoría de los microorganismos son considerados “bacterias viables pero no cultivables” estimándose la recuperación por cultivo en solo un 20% de los microorganismos totales a fecha de la publicación (Eckburg *et al.*, 2005) aunque con el tiempo se van pudiendo recuperar más.

Esta limitación técnica está llevando a múltiples autores a plantearse nuevas fórmulas, suplementos o condiciones de cultivo para poder abarcar el mayor número de microorganismos posible. A esta técnica se la denomina culturómica, mediante la cual el abanico de especies cultivables crece, pero a expensas de un número de condiciones tanto de temperatura, tipo de atmósfera o tiempo de incubación y medios de cultivo muy elevado haciéndose, en la práctica, inabarcable para un laboratorio microbiológico clínico (Lagier *et al.*, 2018).

En los últimos años están surgiendo técnicas de secuenciación no dependientes de cultivo, las cuales, gracias al abaratamiento de los costes junto al desarrollo de aplicaciones y análisis bioinformáticos están siendo día a día más accesibles y fáciles de implementar para abordar el estudio de estas comunidades microbianas complejas con una gran aceptación por la comunidad científica. Estas técnicas tienen la capacidad de distinguir organismos individuales en una muestra compleja a partir del material genético de los mismos permitiendo estudiar comunidades microbianas.

Con las nuevas técnicas de secuenciación masiva somos capaces de detectar, sin necesidad de cultivo, todo el material genético de la muestra, amplificarlo y secuenciarlo independientemente del estado de viabilidad de las bacterias presentes, convirtiendo así las muestras en pequeños ecosistemas. Estas técnicas vienen con sus propios desafíos técnicos, como la correcta identificación de esos microorganismos los cuales no han sido aun correctamente caracterizados, la necesidad de utilizar el gen 16S rRNA o utilizar fragmentos de secuencias relativamente cortos que no abarquen el gen entero (Yarza *et al.*, 2014).

2.2.5 Clasificación taxonómica

El camino para determinar las comunidades microbianas empezó cuando podíamos detectar aquellas especies que, mediante técnicas tradicionales dependientes de cultivo, conseguíamos identificar. Con este protocolo se llevan a cabo multitud de clasificaciones y se va ganando el conocimiento necesario para llevar a cabo su ordenación mediante su genoma, pero también mediante su morfología, bioquímica, metabolismo, patogenicidad e incluso epidemiología (Cowan, 1971). De todo ello se encarga la denominada rama de la sistemática, utilizada en la categorización de las especies microbianas. Estas técnicas requieren de tiempo y son especialmente difíciles de llevar a cabo con bacterias con requerimientos metabólicos complejos. En consecuencia, tenemos aproximadamente 11.000 identificaciones entre bacterias y archeas. Se estima que serían necesarios más de 1.000 años para clasificar todas las especies restantes a la velocidad actual (Rosselló-Móra, 2012).

Introducción

Ya sea por un método o por otro, el objetivo final es poder identificar todos esos microorganismos, o al menos clasificarlos dentro de algún nivel taxonómico. Se ha demostrado la importancia, tanto de la abundancia como de la diversidad microbiana, en la salud de los organismos eucariotas. Este hecho ha puesto de manifiesto en la comunidad científica la necesidad de entender mejor tanto la composición como el rol que cada uno de esos microorganismos tiene en el hospedador haciendo falta poder distinguir un microorganismo de otro.

Desde 1965 se han utilizado las proteínas y los ácidos nucleicos como cronómetros moleculares ya que muchas de estas macromoléculas experimentan mutaciones por azar y aumentan con el paso del tiempo de una manera lineal. Entre todas las macromoléculas, el ácido ribonucleico ribosomal (ARNr), y en particular el de la región 16S, es el más utilizado para los estudios de filogenia y taxonomía bacterianas.

Los ribosomas son orgánulos con una alta complejidad y especialidad, constituidos por 2 subunidades de diferente tamaño. La subunidad grande presenta un índice de sedimentación de 50S (unidades de Svedberg) y la pequeña de 30S. Es esta última la más utilizada para los estudios de taxonomía y evolución y está compuesta a su vez por 21 proteínas diferentes y el ARNr 16S. El ARNr 16S es un polirribonucleótido de aproximadamente 1.500 nucleótidos (nt) dividido a su vez en 9 regiones variables entre las que se encuentran regiones muy conservadas. Las mutaciones en este gen suelen tener un efecto letal en las bacterias y en las que el efecto no es letal no aportan una ventaja competitiva frente a las que no la presentan, por lo que la evolución de este gen es relativamente lenta haciéndolo idóneo para estudios de evolución y taxonomía.

El gen ARNr 16S es ampliamente utilizado para estudios de la microbiota por su gran número de ventajas, aunque plantea también una serie de problemas técnicos: El número de copias del operón ribosómico no es el mismo en todas las bacterias, pudiendo variar de entre 1 a 15 copias. Estas copias a su vez no tienen por qué ser perfectamente homólogas pudiendo haber una pequeña heterogeneidad la cual puede llegar al 1,4% en algunas cepas como las de *Veillonella* (Marchandin *et al.*, 2003). Estos dos factores ponen de manifiesto los problemas principales de la técnica: no es realmente cuantitativa, ya que el

número de copias en cada genoma es diferente y la necesidad de delimitar muy bien el porcentaje de variación aceptable para cada especie bacteriana.

De acuerdo con la definición de especie bacteriana se ha demostrado que las cepas de la misma especie generalmente presentan una similitud mayor del 97% en el gen del ARNr 16S. Así pues, cepas con una disimilitud mayor al 97% en el gen del ARNr 16S es improbable que pertenezcan a la misma especie (Stackebrandt & Goebel, 1994). A pesar de esto se han encontrado cepas de *Mycobacterium chlonae* y *Mycobacterium abscessus* con una similitud inferior al 50% en experimentos ADN-ADN y con una similitud del 99% en el gen ARNr 16S por lo que no se aportará siempre una identificación definitiva a nivel de especie (Patel, 2001).

En general, un porcentaje de disimilitud mayor al 94,5% es una buena evidencia para distinguir géneros, 86,5% para la diferenciación a nivel de familia, 82% para diferenciar órdenes, 78,5% clases y un 75% es necesaria para poder distinguir entre filos (Yarza *et al.*, 2014).

Con todo ello, el estudio del gen 16S ARNr ha abierto un nuevo paradigma en la microbiología con la posibilidad de poder estudiar poblaciones bacterianas.

Como hemos dicho, hay microorganismos que aún no han sido identificados o no disponemos de una nomenclatura para poder referirnos a ellos, en tal caso se realiza una clasificación artificial y en cierta medida subjetiva.

Por norma general los estudios metataxonómicos precisan de miles a millones de secuencias del gen 16S ARNr por estudio. Esas secuencias pueden ser clasificadas por variantes con exactamente la misma secuencia de nucleótidos repetida a lo largo de diferentes secuencias, denominándose secuencias variantes de amplicón o ASVs por sus siglas en inglés "*Amplicon Sequence Variants*". Esas secuencias pueden compararse entre sí y unir las que tengan un gran parecido o similitud. Normalmente, la similitud entre 2 secuencias se calcula como el porcentaje de sitios que concuerdan en una alineación de secuencias por pares, siendo el umbral utilizado el del 97%, derivado de los estudios sobre el gen ARNr 16S y generando una nueva secuencia que sea representativa de las secuencias. Estas secuencias representativas de cada grupo de secuencias son las denominadas OTUs, del inglés "*Operational Taxonomic Unit*" pudiendo

Introducción

reducir el análisis a unos pocos miles de secuencias representativas a las cuales se las realizará la asignación taxonómica frente a las bases de datos.

La clasificación por OTUs y la selección de las secuencias representativas de cada grupo de secuencias es un paso fundamental y ampliamente controvertido. La elección del 97% de similitud entre las secuencias está fijado como vimos previamente por los estudios sobre el gen ARNr 16S, pero los métodos de secuenciación actuales solo suelen abarcar de 1 a 3 regiones hipervariables de dicho gen. Esto podría no ser suficiente para determinar con exactitud especies diferentes, lo que conduce a una subestimación de la diversidad real.

2.2.6 Estudios de diversidad

Una vez clasificados los microorganismos en ASVs, OTUs o en taxones, necesitamos las herramientas, tanto teóricas como prácticas, para el estudio de la diversidad. Para ello, se han recogido muchos conceptos clásicos de evolución y ecología de eucariotas con las dificultades que ello conlleva. Algunos de esos conceptos son: reposición intergeneracional (muy rápida en bacterias comparada con eucariotas), presiones bióticas (se incluyen las interacciones con la defensas del hospedador o el propio control activo del hospedador como la toma de antibióticos selectivos, probióticos, etc.), producción de metabolitos libres (muy común en la microbiota intestinal y raro entre eucariotas), efectos directos sobre el ecosistema (producción de antibiocinas, cambios en el pH, etc.) o mediciones de la diversidad y riqueza de especies entre otros. Aun así, se han visto modelos muy interesantes que comparten muchas analogías: ocupación de un hábitat desocupado (desarrollo de la microbiota en recién nacidos), formación del ecosistema después de una alteración (toma de antibióticos) o incursión de una especie invasiva (el contexto de un patógeno) (Costello, Stagaman, Dethlefsen, Bohannan, & Relman, 2012).

Además del significado que tiene en sí misma la diversidad, pudiendo entenderse como el número efectivo de taxones en la muestra (riqueza de taxones) o como el grado en el que las diferentes especies son similares en cuanto a su abundancia (equitatividad), es un parámetro útil en el estudio de las

comunidades microbianas. En 1972 se acuñaron 2 de los términos más utilizados en el estudio de la microbiota: diversidad alfa y diversidad beta con los cuales vamos a poder medir, comparar y agrupar las distintas comunidades microbianas (Whittaker & Whittaker, 1972).

La diversidad alfa se define como la relación de la riqueza y la equitatividad de los taxones que se encuentran en la muestra. Se puede medir matemáticamente y para ello solo necesitamos el número total de taxones y el número de microorganismos en cada taxón. Con ello podremos calcular abundancias relativas y calcular los índices de diversidad. Los índices de Shannon-Wiener y Simpson son los índices de diversidad ecológica más utilizados para cuantificar la diversidad de cada muestra. El índice de Shannon-Wiener (1) fue postulado en 1949 y se utiliza para medir la entropía del resultado del muestreo. Sus componentes principales son p_i (siendo el porcentaje de la comunidad explicada por el taxón i o abundancia relativa) y S , el número de especies.

$$H = - \sum_{i=1}^S (p_i \log_2 p_i) \quad (1)$$

El índice de diversidad de Simpson (2), también postulado en 1949, mide la probabilidad de extraer al azar dos secuencias de una muestra y pertenezcan al mismo taxón o dicho de otro modo la dominancia de los microorganismos.

$$S = 1 - \sum p_i^2 \quad (2)$$

Al hablar de la riqueza de los taxones determinar el número y la identidad de los taxones que se comparten entre diferentes individuos, es actualmente una de las principales cuestiones de investigación.

Introducción

La diversidad beta se define como el estudio comparativo de los perfiles de los taxones entre las diferentes muestras. Estas comparaciones pueden ser a múltiples niveles donde los 2 principales son: la comparación por presencia/ausencia de los taxones de las muestras y la comparación por la abundancia relativa de esos taxones.

Los coeficientes más utilizados son el coeficiente binario de Jaccard para estudios de comparación presencia/ausencia y el coeficiente de Bray-Curtis para estudios de comparación por abundancias relativas. Otros coeficientes más específicos para los estudios de microbiota son los coeficientes Unifrac “unweighted” para estudios de presencia/ausencia y el coeficiente Unifrac “weighted” para estudios de comparación por abundancia relativas. Estos últimos presentan la peculiaridad de realizar los estudios incorporando la información sobre la relación filogenética de los taxones observados en el estudio (Lozupone, Hamady, Kelley, & Knight, 2007).

El resultado de la aplicación de estos índices es la construcción de 2 matrices de distancia triangulares de las muestras.

Entre los taxones comunes, tienen especial interés aquellos que están presentes en todas o, al menos, en una gran mayoría de las muestras. A estos taxones comunes se les conoce como taxones *core*. El número de taxones *core* está muy condicionado a la profundidad de secuenciación y a la prevalencia requerida de dichos taxones dentro de las muestras de estudio (Salonen *et al.*, 2012).

Esto hace fundamental definir en cada estudio las condiciones que tienen que cumplir dichos taxones para poder ser incluidos dentro de la categoría de *core*. En esta tesis se entenderá como microbiota *core* aquella microbiota con una abundancia relativa mínima del 0,01% en, al menos, el 80% de las muestras dentro de los diferentes grupos de pacientes. Para ilustrar esta definición se presenta el siguiente ejemplo teórico con un único taxón:

“De un total de 100 pacientes, 50 pacientes fueron diagnosticados como positivos y 50 como negativos a infección por *H. pylori*”.

“Entre los pacientes positivos se detectaron secuencias de bacterias pertenecientes al género *Helicobacter*, con una abundancia relativa mayor del 0,01%, en 48 de ellos. Por otro lado, en los pacientes negativos se detectaron secuencias de bacterias pertenecientes al género *Helicobacter*, con una abundancia relativa mayor del 0,01%, en 20 de ellos”.

Al analizar la microbiota *core* el género *Helicobacter* no formaría parte de la microbiota *core* general (se ha detectado solo en 78 de los 100 pacientes) pero si formará parte de la microbiota *core* de los pacientes positivos (se ha detectado en 48 pacientes del total de 50 (el 96%)).

2.2.7. Microbioma gástrico

El ecosistema gástrico presenta una serie de particularidades dentro de todo el tracto gastrointestinal que lo convierten en un ambiente hostil para la supervivencia, y más aún para la colonización, de las posibles bacterias que puedan llegar a él. Este hecho, viene fundamentado principalmente en los jugos gástricos, una mezcla de componentes de enzimas proteolíticas y ácido clorhídrico que crea un ambiente muy ácido (pH de 1-2) con capacidad protectora y digestiva (Martinsen, Bergh, & Waldum, 2005).

Los primeros estudios sobre microorganismos en el estómago provienen de antes del aislamiento de *H. pylori* en 1982. Por ejemplo, en 1977 Savage *et al.* consiguieron aislar bacterias de muestras gástricas con una abundancia media de 10^3 UFC/g entre las que se encontraban bacterias de diferentes géneros como *Bifidobacterium*, *Streptococcus*, *Clostridium*, *Lactobacillus* y *Veillonella* entre otros (Savage, 1977).

Sin embargo, estos hallazgos fueron catalogados como bacterias transitorias al ser bacterias que están presentes a lo largo de todo el tracto intestinal y no como bacterias que habían colonizado el estómago.

Introducción

Con el aislamiento de *H. pylori* por Barry Marshal y Robin Warren en 1982 (Marshall & Warren, 1984) el paradigma del estómago como un ambiente estéril o como órgano de transición bacteriana se rompió.

Actualmente, tanto estudios dependientes de cultivo con medios más selectivos y optimizados (Delgado, Cabrera-Rubio, Mira, Suárez, & Mayo, 2013) como estudios independientes de cultivo como las herramientas actuales de secuenciación masiva y metataxonomía permiten distinguir no solo la presencia de *H. pylori* sino de todo un ecosistema gástrico.

Actualmente, los aislados procedentes de muestras gástricas muestran un número muy limitado de bacterias viables (aproximadamente se encuentran 10^3 unidades formadoras de colonias por mL de fluido gástrico). Con este número tan reducido se hace realmente muy difícil determinar toda la comunidad bacteriana que puede abarcar el ecosistema gástrico.

Bik *et al.* en el año 2006, utilizando el análisis del gen 16S rRNA, analizaron 1.833 secuencias entre las que encontraron 128 secuencias bacterianas diferentes (Bik *et al.*, 2006). Li *et al.* (2009) analizaron 1.223 secuencias obteniendo resultados muy similares en muestras de diferentes países. En ambos estudios se encontraron una gran abundancia de *Streptococcus* y *Prevotella*, sumando entorno al 50% de las secuencias totales (Li *et al.*, 2009).

Tanto en estos estudios como en el de Delgado *et al.* (2013), se validaron las técnicas de secuenciación masiva por pirosecuenciación junto con técnicas dependientes de cultivo y se compararon los resultados. Detectaron que las bacterias pertenecientes a los géneros *Streptococcus*, *Propionibacterium* y *Lactobacillus* eran las más abundantes tanto por cultivo como por pirosecuenciación (Delgado *et al.*, 2013).

En 2015 Tsuda *et al.* encontraron una gran abundancia de *Streptococcus* y *Prevotella* junto a bacterias del género *Rothia* y *Neisseria* entre los 5 géneros mayoritarios (Tsuda *et al.*, 2015).

En todos estos estudios se demostró que la microbiota gástrica es mucho más compleja de lo que podría parecer a simple vista y que queda mucho por saber sobre la relación entre *H. pylori* y el resto de las bacterias gástricas.

Múltiples estudios se han centrado en el efecto que provoca *H. pylori* en el resto de la microbiota gástrica pero no se ha descrito el efecto que tiene el resto de la microbiota sobre *H. pylori*. Todo ello puede complicarse debido a la dificultad que entraña comparar diferentes estudios que, por sus métodos, protocolos de trabajo o bases de datos de referencia hace que haya una gran variabilidad en los resultados. Esto, dificulta la construcción de modelos matemáticos que nos permitan inferir el resto de posibles factores que pueden afectar tanto a la infección como a diferentes patologías derivadas de la misma en los seres humanos (Ilianov Iliev *et al.*, 2019).

2.2.8 Importancia del microbioma en las patologías gástricas

H. pylori limita la producción de ácido gástrico mediante la producción de mediadores de inflamación afectando progresivamente a las glándulas gástricas y conllevando eventualmente una gastritis atrófica (Blaser & Atherton, 2004).

Esta reducción de las secreciones gástricas va acompañada de un aumento del pH en las zonas gástricas afectadas, lo que facilita la supervivencia de diferentes microorganismos en el ambiente gástrico. El cómo esta microbiota alterada puede interactuar con *H. pylori* o cómo puede afectar al desarrollo de enfermedades gástricas como el cáncer gástrico no está del todo comprendido.

Estudios previos, como el de Lofgren *et al.* (2011), demostraron la importancia de la microbiota mediante el uso de ratones libres de gérmenes. En ellos se apreció una menor velocidad en el desarrollo de gastritis atrófica y cáncer gástrico frente al grupo con una microbiota normal (Lofgren *et al.*, 2011).

Una de las hipótesis discutidas por M.J. Blaser (2004) sobre la capacidad carcinogénica de *H. pylori* trata de la producción de compuestos carcinogénicos N-nitrosos por medio de diferentes bacterias nitrificantes que utilizan los compuestos con nitrógeno del fluido gástrico (Blaser & Atherton, 2004).

Entre esas bacterias podemos encontrar a: *Escherichia*, *Lactobacillus*, *Nitrospirae*, *Clostridium*, *Veillonella*, *Haemophilus* o *Staphylococcus* (Jo *et al.*, 2016; Mowat *et al.*, 2000; L. Wang *et al.*, 2016; Williams & McColl, 2006).

Introducción

Por otro lado, bacterias pertenecientes a los géneros *Streptococcus*, *Prevotella* y *Neisseria* están asociados a un menor riesgo de desarrollar cáncer gástrico (Ferreira *et al.*, 2018; I. Yang *et al.*, 2016).

En los estudios de Ferreira *et al.* (2018) caracterizaron la microbiota gástrica de pacientes con carcinoma gástrico con una reducción de la diversidad microbiológica principalmente representada por bacterias intestinales comensales. Este perfil bacteriano presentaba un incremento de la función nitrato reductasa (Ferreira *et al.*, 2018).

Todos estos estudios nos incitan a pensar que la microbiota gástrica está involucrada en la carcinogénesis gástrica, por ello, el inicio y el desarrollo de la carcinogénesis podrían ser controladas mediante la modificación de la microbiota. Esto podría incrementar la tasa de éxito terapéutico y finalmente ayudar en una prevención precisa contribuyendo en estrategias de tratamientos personalizados (Peng, Li, Hu, & Lu, 2019).

Se conocen varios moduladores de la microbiota gástrica, Koga *et al.* (2018) sugiere que el aumento del consumo de medicamentos antiácidos como los inhibidores de bombas de protones genera un aumento de la masa microbiana en el estómago haciendo que sea más fácil su migración al intestino (Koga, Ohtsu, Kimura, & Asami, 2019).

Se ha estudiado el efecto de los probióticos en modelos animales y se ha observado que parecen influir en la colonización de *H. pylori* y atenúan la inflamación gástrica inducida por esta bacteria (Isobe *et al.*, 2012; Kabir *et al.*, 1997).

También se ha observado que *H. pylori* no solo modula la microbiota gástrica sino que es capaz de afectar de manera significativa a la microbiota intestinal. En poblaciones pediátricas se ha observado que aquellos niños asintomáticos de entre 6 y 12 años tienen hasta el doble de probabilidades de tener un incremento en número y variedad de bacterias en su microbiota de colon (Benavides-Ward *et al.*, 2018; Brawner *et al.*, 2017a).

El ambiente gástrico es una barrera frente a multitud de microorganismos y es fundamental profundizar en su conocimiento para conocer cuál es su microbiota,

cómo cambia con las diferentes patologías y poder modularla para realizar tratamientos personalizados.

2.3 Herramientas bioinformáticas

La bioinformática queda definida por la Real Academia Española como “La aplicación de la informática a la investigación biomédica.” Es una disciplina en continua evolución con un desarrollo exponencial año tras año por lo que no es de extrañar que esta definición pueda ser actualizada con el paso de tiempo.

La disciplina de la bioinformática surge como el esfuerzo y unión interdisciplinar de la biología molecular, la genética (genómica, proteómica y transcriptómica entre otros), las ciencias biomédicas (epidemiología genética y molecular), las ciencias de la computación, las matemáticas, la física y la estadística para ofrecer soluciones a los problemas de la investigación en ciencias (en especial ciencias de la salud y biológicas) que tratan una ingente cantidad de datos, cada día mayor y más compleja. Así pues, la bioinformática aplica los principios de la ciencia de la información y tecnológicas para hacer uso de esos vastos, diversos y complejos datos de las ciencias de la vida.

2.3.1 Proteómica

La proteómica se empezó a utilizar en microbiología clínica con la implementación de la espectrometría de masas “*Matrix-Assisted Laser Desorption Ionization Time of Flight*” (MALDI-TOF MS) y, desde entonces, los métodos y tiempos de diagnóstico han vivido una gran revolución.

El MALDI-TOF consiste en una técnica de ionización suave con un espectrómetro de masas con un analizador de tiempo de vuelo. Con ello, podemos separar y analizar diferentes moléculas según su masa y su carga mediante la medición de un flujo de iones acelerados por impulso de campo eléctrico. La trayectoria y los tiempos serán proporcionales a la masa y la carga de cada uno de esos iones.

Introducción

En microbiología esto nos proporciona un espectro característico para cada especie dada en base a macromoléculas como las proteínas de diversos tamaños pudiendo llegar a 100kDa, siendo una técnica rápida e innovadora para la identificación bacteriana (Sauer *et al.*, 2008)

Se suele disponer de la muestra (generalmente una colonia bacteriana o fúngica pudiendo haber sido pretratada o no) sobre la placa metálica a la cual se le añade la matriz. La matriz será clave para absorber la energía del láser y ayudar en la desorción de los iones de la muestra.

Cada análisis producirá un espectro en el que cada pico corresponde con un tipo de molécula y será necesario su comparación con una base de datos de espectros para poder realizar una identificación correcta. Será pues fundamental contar con una muy buena base de datos de referencia para mejorar las determinaciones.

2.3.2 Secuenciación masiva

A partir de que Fred Sanger y sus colaboradores describieron en 1977 un nuevo mecanismo que permitía secuenciar y determinar el orden de los nucleótidos (A, C, G y T) a partir de un oligonucleótido de ADN (Sanger, Nicklen, & Coulson, 1977), la biología molecular en este campo no ha parado de crecer y cambiar conforme han ido surgiendo nuevas tecnologías, adaptándose a herramientas más baratas y eficientes como son las herramientas de secuenciación masivas o "*Next-generation sequencing*". Con el paso de los años estas herramientas han ido aumentando tanto su precisión como su rendimiento ya que nos permiten obtener, de miles a cientos de miles de secuencias en una única carrera o proceso, ofreciendo todo un nuevo abanico de posibilidades: secuenciación de genomas enteros, secuenciación de transcriptomas completos (RNA-seq), estudios epigenéticos (como son los estudios de metilación), identificaciones metataxonómicas, etc.-

La identificación taxonómica de los microorganismos a partir del gen ARNr 16S siempre implica 4 etapas independientemente de la plataforma o técnica utilizadas: extracción del material genético, amplificación del gen ARNr 16S,

determinar la secuencia del amplicón y finalizar con el análisis correcto de la secuencia.

Las diferentes plataformas o técnicas presentan diferentes enfoques para abordar estas etapas las cuales también dependerán del tipo y número de muestras que queramos analizar.

Conocer las diferencias entre las diferentes plataformas junto con sus limitaciones es fundamental para poder confirmar o refutar diferentes publicaciones previas. Esto hace importante no olvidar plataformas, que a pesar de no seguir disponibles en el mercado, tuvieron una gran importancia en la ciencia como es el caso de la plataforma 454 Life Sciences GS FLX System (GS FLX) de Roche.

-Illumina (MiSeq, HiSeq y NextSeq): su tecnología se basa en la polimerización del ADN, incorporando nucleótidos marcados con fluorescencia. Esta técnica es una de las más potentes en cuanto a información generada y costes. Permite secuenciar fragmentos pareados de 300 pares de bases. Las diferencias entre los secuenciadores radican principalmente en el número de secuencias que pueden producir (HiSeq 300 millones y MiSeq 25 millones), lo que se refleja en el tiempo de procesado de las muestras (1 semana comparada con 1 día) y en su precio (2 veces más caro HiSeq). Con todo esto vemos que cada tecnología resulta interesante para diferentes proyectos que necesiten un número mayor o menor de secuencias.

-Ion Torrent (PGM y S5): detecta los iones H^+ liberados por la polimerasa tras la incorporación de un nucleótido. Esta tecnología permite ahorrar costes y tiempo respecto a otras tecnologías permitiendo secuenciar fragmentos de 100 pares de bases.

-SOLiD (Life Technologies): tecnología de secuenciación por ligación de una batería de octámeros marcados con fluorescencia. El tamaño de las secuencias es significativamente más bajo, permitiendo secuenciar fragmentos de 75+35 pares de bases.

Introducción

-Roche/454FLX: estuvo ampliamente extendida entre los años 2007 y 2015 hasta que fue discontinuado a mediados del año 2016. Se trata de una técnica de pirosecuenciación basada en la polimerización no fluorescente midiendo la liberación de pirofosfato mediante una serie de reacciones enzimáticas liberando luz en cada incorporación de un nucleótido; esto produce imágenes que se analizan y, una vez interpretadas por el ordenador, devuelven las secuencias de nucleótidos. Su ventaja radicaba en tamaño de las lecturas generadas, permitiendo secuenciar fragmentos de alrededor de 400 pares de bases. Su gran inconveniente eran los costes y la dificultad de secuenciación de regiones homopoliméricas.

Continuamente se están desarrollando nuevas tecnologías para suplir las deficiencias de las anteriores (principalmente longitud de las secuencias, facilidad de uso, tasa de error y tiempo de análisis). Hay 2 tecnologías conocidas como de tercera generación: la secuenciación a tiempo real de molécula única o "SMRT" "*Single Molecule, Real-Time*" por sus siglas en inglés (Bentley *et al.*, 2008) y la aproximación de construcción de secuencias grandes a partir de secuencias más cortas (T. D. Harris *et al.*, 2008). Entre estas 2 tecnologías la SMRT está mucho más extendida, siendo la plataforma Pacbio la más ampliamente distribuida.

La plataforma Pacbio (Pacific biosciences SMRT sequencing) también utiliza nucleótidos marcados pero en lugar de hacer ciclos de unión y lavado se realiza la lectura a tiempo real según son incorporados por la polimerasa a la cadena de ADN. Su principal ventaja es el tiempo de secuenciación y la longitud de las secuencias que pueden ser obtenidas (la longitud media es de 10.000 pb pero se pueden obtener secuencias de hasta 60.000 pb). Su principal problema es la alta tasa de errores que presentan sus secuencias siendo errores aleatoriamente distribuidos que pueden suponer hasta el 10% de los nucleótidos de la secuencia (Koren *et al.*, 2012).

A pesar de las numerosas diferencias entre las diferentes plataformas y protocolos de trabajo y tomando en consideración sus limitaciones, se ha mostrado como las conclusiones biológicas obtenidas a partir de los datos son consecuentes entre las plataformas y protocolos. Aun así, sigue siendo

fundamental elegir el equipo y sistema adecuado para el diseño del experimento (Allali *et al.*, 2017).

2.3.3 Los repositorios de datos

Con la implementación de las nuevas tecnologías de secuenciación nos encontramos con la posibilidad de poder secuenciar pequeñas y grandes secuencias, organismos o incluso ecosistemas enteros a tiempos y costes mucho más asequibles. Generar datos se ha convertido en una tarea relativamente sencilla y en poco tiempo somos capaces de generar un gran volumen. Sin embargo, tenemos que ser capaces de poder procesarlos, limpiarlos, analizarlos y poder obtener un rendimiento de estos. Hasta hace relativamente poco tiempo el procesamiento de los datos corría a cargo de los propios investigadores lo que podía llevar a ralentizar la investigación o, dependiendo del volumen de los datos, a imposibilitarla. Con el fin de facilitar esta tarea se busca crear métodos de automatización: filtrado automático por calidad, la agrupación de secuencias en ASVs y/o OTUs así como la anotación de las secuencias de manera automática. La anotación automática nos permite manipular gran cantidad de información en tiempos mucho más pequeños pero precisan de un gran recopilatorio de información previa para realizar asignaciones taxonómicas fiables.

Para este fin, hay muchas bases de datos encargadas de almacenar secuencias de ARNr 16S como son Greengenes (McDonald *et al.*, 2012), Ribosomal database project (Cole *et al.*, 2014), Silva (Quast *et al.*, 2013) que nos van a ayudar como bases de datos de secuencias de referencia entre las que podemos encontrar más de 4 millones de secuencias listas para ser utilizadas como identificadores para un conjunto de secuencias que queramos comparar.

En las investigaciones bioinformáticas utilizamos diferentes bases de datos en función de la información que sea precisa para el trabajo. Actualmente, hay disponibles multitud de repositorios con información de diversa índole. Así pues podemos encontrar repositorios con secuencias de ADN de genomas completos a genes de funciones específicas, de proteínas o de dominios proteicos. Estos repositorios podrán ser actualizados periódicamente o no en función del mantenimiento o información que sea necesario incorporar.

Introducción

Las bases de datos de ADN constituyen un pilar fundamental en las investigaciones genómicas. Podemos encontrar múltiples bases de datos creadas con diferentes fines lo que genera uno de los principales problemas en bioinformática: las diferentes anotaciones. Trabajar con bases de datos que nombren a la misma cosa de maneras diferentes puede ser un verdadero quebradero de cabeza para los investigadores que tengan que lidiar con ello, por lo que es fundamental la selección de una buena base de datos con una misma estructura y datos correctamente incorporados.

Entre los mayores repositorios de secuencias de ADN tenemos el “International Nucleotide Sequence Database Collaboration” (INSDC), formado por el “DNA Data Bank of Japan” (DDBJ) (Mashima *et al.*, 2017), por el GenBank (actualmente, se encuentran 124,277,818,310 pb, en un total de 132,015,054 entradas, en su principal división) (Clark, Karsch-Mizrachi, Lipman, Ostell, & Sayers, 2016) y por el “European Molecular Biology Laboratory” EMBL (el cual, en su base de datos “European Nucleotide Archive” (ENA), contiene más de 301,588,430,608 nucleótidos con 199,575,971 entradas) (Kanz *et al.*, 2005).

2.3.4 Etapas en el estudio de la microbiota y sus limitaciones

Con el aumento del interés en el estudio de la microbiota, y con cada día más grupos de investigación iniciándose en él, se hace necesario remarcar los diferentes pasos que todo análisis de la microbiota debería realizar. Todos han de partir con una idea o problema a resolver, obtener y tratar la muestra según el protocolo de trabajo, extraer los ácidos nucleicos o proteínas, secuenciar, analizar e interpretar los resultados. Así mismo, es interesante reunir las limitaciones que podemos encontrarnos en cada uno de los diferentes pasos para un estudio como es el de la presente tesis ya que tienen que conocerse previamente al empezar su diseño.

a) La idea:

Éste es el punto más importante. Tiene como objetivo la definición, planificación y justificación del estudio. Una correcta definición y planificación será lo que más errores y problemas futuros evitará. Será aquí donde deberemos buscar la mayoría de la información y definir por escrito la hipótesis de trabajo y los objetivos. Habrá también que analizar la dificultad del estudio, posible sesgos en la toma de las muestras y con todo ello ver si realmente es viable y está justificada o no la propuesta del estudio.

Los estudios de microbiota, incluyendo los de microbiota gástrica, pueden englobarse dentro de 3 grupos de estudios: estudios longitudinales, estudios de intervención y estudios de cohortes.

Los estudios longitudinales son especialmente relevantes en la evolución de la microbiota. Consisten en tener información constante de varios pacientes que van pasando por los diferentes estadios de las patologías gástricas y ver el cambio de la microbiota de un paciente sano a gastritis activa que termina en una posible gastritis crónica o ver su evolución a úlcera o metaplasia.

Los estudios de intervención tienen interés para determinar el efecto que tiene un tratamiento determinado en la microbiota de los pacientes realizándose un estudio de la microbiota antes y después del tratamiento en los mismos pacientes, por ejemplo con una nueva rutina de antibióticos o una operación quirúrgica.

Los estudios por cohortes aportan información puntual de un grupo "cohorte" de pacientes con unas características particulares en un momento particular. Cuanto más homogéneo sea el grupo menor variabilidad de los resultados encontraremos.

En todos los tipos de estudios la obtención de la muestra presentará el principal inconveniente, siendo las muestras principales la biopsia y el jugo gástrico (Sung *et al.*, 2016), obtenidas mediante gastroscopia con todo lo que ello conlleva para el paciente.

La prueba por gastroscopia debe ser realizada con un fin diagnóstico para tratar una enfermedad. Por motivos éticos y por la naturaleza del método de extracción

Introducción

de la muestra debemos descartar los estudios longitudinales y la mayoría de los estudios de intervención (Miao, Wan, & Wang, 2020), dejando los estudios por cohortes como los estudios más habituales para los estudios de microbiota gástrica.

Estos estudios por cohortes son los que presentan una mayor heterogeneidad (no se realizan, normalmente, sobre un mismo paciente), la selección de pacientes será siempre a posteriori de su diagnóstico y se realizará la gastroscopia por prescripción médica justificada (no habrá pacientes totalmente sanos).

Al diseñar este trabajo se detectaron las siguientes limitaciones para realizar un estudio bioinformático de muestras clínicas con datos procedentes de diferentes proyectos metataxonómicos:

- No disponer de muestras propias, sin conocer en cada momento toda la información que se podría estimar interesante en este tipo de estudios y limitando el estudio a unos factores muy definidos cuyo conocimiento es obligatorio para que la muestra pueda ser considerada para el estudio.
- Limitar el estudio a un estudio por cohortes, utilizando muestras de pacientes con unas características particulares en un momento particular. Este hecho limita la posibilidad de realizar posibles estudios de seguimiento del paciente o longitudinales y presentara una mayor heterogeneidad en los resultados.
- Estudio limitado a pacientes enfermos. No podemos olvidar que la muestra de estudio seleccionada en esta tesis es la biopsia gástrica y ésta ha de ser realizada con un fin diagnóstico para tratar una enfermedad. Todas las muestras serán de pacientes con alguna patología, pudiendo ser causada, o no, por *H. pylori* y no vamos a poder obtener conclusiones directas sobre población totalmente sana.

Como vemos, este tipo de limitaciones obliga al científico a realizar un planteamiento muy riguroso de su estudio. Definir los objetivos de su estudio y realizar una selección de aquellos proyectos que cumplan sus criterios metodológicos puede ser arduo, pero reducirá en gran medida la posibilidad de cometer sesgos que afecten a sus resultados y, por ello, a sus conclusiones.

b) La muestra:

Independientemente del tipo de estudio a realizar hay que tener precaución con la obtención y manipulación de la muestra. Las principales fuentes de contaminación en las muestras gástricas son: una manipulación incorrecta de la muestra en la sala de endoscopia o en el quirófano, contaminación por tejido contaminante (generalmente bucal o faríngeo) (Layfield, Witt, Metzger, & Anderson, 2011), limpieza deficiente de los aparatos de endoscopia o del laboratorio (Banerjee *et al.*, 2008), tiempo prolongado a temperatura ambiente o retrasos en la entrega o al análisis (Garza-González, Perez-Perez, Maldonado-Garza, & Bosques-Padilla, 2014).

Entre las bacterias que tradicionalmente se consideran contaminantes de biopsias gástricas encontramos *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Enterococcus faecalis*, *Candida spp*, y *Pseudomonas aeruginosa*. La presencia de estas bacterias disminuye la sensibilidad de los cultivos para el diagnóstico de *H. pylori* (Ribeiro *et al.*, 2004).

Una correcta manipulación nos aportará mucha más seguridad en los resultados y nos ayudará a distinguir si un microorganismo es realmente contaminante o es propio de la muestra.

Al valorar el tipo de muestra para este estudio se investigó cual podría ser la muestra más representativa del ambiente gástrico. Las principales muestras utilizadas para el estudio de la microbiota gástrica son el jugo o la biopsia gástrica, siendo esta última la más ampliamente utilizada este tipo de estudios. Estudios como el de Sung *et al.* (2016) compararon la idoneidad de ambos tipos de muestra para el estudio de la microbiota gástrica concluyendo que las muestras de mucosa gástrica reflejan mejor la microbiota gástrica al tener una mayor carga bacteriana (Sung *et al.*, 2016). A pesar de que las muestras de biopsia gástrica representan mejor el perfil de la microbiota gástrica es un tipo de muestra con una serie de limitaciones:

- Las muestras de biopsia gástrica son muestras obtenidas a través de procedimientos invasivos para el paciente por lo que su obtención ha de estar debidamente justificada.

Introducción

- La toma de la muestra no siempre es fácil y puede contaminarse con bacterias tanto del quirófano o por una manipulación incorrecta.
- Por norma general el análisis metataxonómico no suele ser el primer análisis que se le realiza la prueba (puede ser realizado tras pasar la muestra por una siembra por arrastre) por lo que el tiempo desde la toma hasta la extracción del material genético puede ser variable.
- La carga de bacterias de las muestras gástricas es, generalmente, muy baja y esto se hace especialmente patente en aquellas muestras de pacientes negativos a infección por *H. pylori* (Loayza *et al.*, 2015). Una menor carga de ADN bacteriano aumentará de manera muy significativa la probabilidad de secuencias contaminantes y aumentará la heterogeneidad de la muestra (Dahlberg *et al.*, 2019).

El uso de las muestras gástricas para los análisis metataxonómicos no es sencillo. Obliga al científico a ser más crítico con los resultados obtenidos y a utilizar solo aquellas muestras que aporten robustez y seguridad al estudio. Como hemos observado hay que realizar estudios de idoneidad tanto al nivel de las muestras como al nivel de proyectos completos, para así intentar disminuir en lo posible los factores externos al estudio que puedan introducir variabilidad a los resultados.

c) El método de extracción del ADN:

Actualmente no existe un método de extracción de ADN considerado de referencia o *gold-standard*. Se recomienda a los investigadores que sean lo más rigurosos posible en la búsqueda de un protocolo que se adecúe al tipo de muestra a analizar e intenten utilizar el mismo método para estudios posteriores para evitar posibles desviaciones inherentes a los métodos de extracción. Existen diferentes sistemas comerciales para diferentes tipos de muestras.

Las muestras gástricas son muestras con relativa baja carga de ADN bacteriano, el tamaño de la muestra generalmente es pequeño, contiene mucho ADN humano y presencia de proteínas que pueden interferir en los procesos de purificación del ADN. Habrá que valorar cada uno de estos factores para elegir el mejor protocolo de extracción.

d) El método de secuenciación:

A la hora de obtener nuestras secuencias del ADN purificado, tendremos que tomar dos decisiones importantes: el secuenciador y el fragmento del 16S ARNr que será amplificado y secuenciado.

Generalmente, para amplicones de un tamaño de unas 250-300 pares de bases se suelen utilizar Illumina Miseq o el Ion Torrent PGM. Hay multitud de estudios únicamente centrados en la comparación de los perfiles microbianos en función del secuenciador utilizado y la región hipervariable secuenciada (Marine *et al.*, 2020) (Clooney *et al.*, 2016)(Salipante *et al.*, 2014).

Igual que en casos anteriores se recomienda realizar una búsqueda bibliográfica en profundidad para poder disponer de un criterio fundamentado y poder realizar las comparaciones con otros estudios sin depender de factores de distorsión como son el secuenciador y la región secuenciada.

En general, el uso de la plataforma Miseq de Illumina mediante la secuenciación de fragmentos 2x250-2x300 que contenga las regiones V3-V4 permiten diferenciar entre múltiples microorganismos diferentes aportando una gran diversidad a la muestra (Clooney *et al.*, 2016).

e) Análisis de los datos:

Actualmente podemos encontrar múltiples programas informáticos o páginas web que ofrecen unos servicios de análisis, con mejor o peor acierto, a las necesidades que puede presentar un investigador. Estos programas se van actualizando día a día. Sin embargo, a pesar del continuo crecimiento de las herramientas, los usuarios no están familiarizados con las limitaciones y posibles sesgos asociados a cada tratamiento de datos.

Podemos encontrar 3 tipos de plataformas en función de lo abiertas o modificables que sean para los investigadores: plataformas cerradas o propias, plataformas semiabiertas y plataformas abiertas.

- Plataformas cerradas o propias: en esta categoría podemos encontrar aquellas que están diseñadas para realizarse de manera secuencial sin posibilidad, o dificultando mucho, la modificación del programa por el investigador. Dentro de esta categoría tenemos las herramientas propias

Introducción

de las empresas de secuenciación como la plataforma Illumina Reporter (Illumina Inc, 2015) o plataformas online como Calypso (Zakrzewski *et al.*, 2017). Generalmente son las más fáciles de usar por los investigadores y no requiere de conocimientos previos de programación.

- Plataformas semiabiertas: Dentro de esta categoría tenemos aquellas que nos proporcionan herramientas diversas y dependerá del investigador el uso de unas u otras en función de cómo quiera analizar sus datos. Suelen ser las más completas debido a la gran variedad de herramientas que ofrecen pudiendo realizar los análisis por etapas lo que genera una muy buena curva de aprendizaje para los investigadores. Dentro de esta categoría tendríamos QIIME2 (Bolyen *et al.*, 2019) o MOTHUR (Schloss, 2019). Se han realizado estudios para determinar cuál de ellas presenta más ventajas sin encontrarse grandes diferencias salvo en los taxones minoritarios (López-García *et al.*, 2018; Tawfik, Azab, Ahmed, & Fayyad, 2018). Estas herramientas son generalmente cerradas u ocultas dentro de scripts y la modificación de alguna exigirá un nivel muy elevado de programación y terminará derivando en una posible destrucción de la herramienta y puede que de la plataforma.
- Plataformas libres: dentro de esta categoría nos encontramos con aquellas que han ido generándose gracias al apoyo de grupos de investigación como puede ser la plataforma de R (R Core Team, 2018). Dentro de esta plataforma se han ido creando librerías de comandos con diferentes funciones que aportan mucho a los análisis de los datos. En este tipo de plataformas cada investigador tiene que buscar y modular las funciones para realizar un análisis de sus datos. La ventaja principal es la necesidad de conocer bien que es lo que se necesita en cada punto del análisis y como llevarlo a cabo. Las plataformas libres como R suelen tener las diferentes herramientas que presentan las plataformas semiabiertas pudiendo realizar todos los análisis con un solo programa. Los principales inconvenientes son la necesidad de entender su lenguaje y la búsqueda de las herramientas/funciones que quieras realizar. Una vez solventados los inconvenientes, la plataforma ofrece un acceso sencillo a multitud de técnicas tanto estadísticas, de transformación de datos como gráficas.

La principal limitación en el análisis de los datos metataxonómicos va a ser pues, el investigador. El gran número de programas o plataformas puede abrumar al investigador en un principio y hace difícil determinar cuál puede ser el más afín a los objetivos de su estudio. En este punto se hace patente la necesidad de tener un equipo multidisciplinar en el que cada miembro aporte su conocimiento al proyecto. Contar con un bioinformático es muy recomendable para poder brindar una mayor libertad y gama de herramientas para poder realizar un análisis lo más correcto y fiel a las muestras.

Objetivos

3. Objetivos

1. Aplicar herramientas bioinformáticas en el estudio proteómico para la identificación de *H. pylori* y analizar cómo mejora su identificación con la modificación de su base de datos de referencia.
2. Desarrollar un documento con instrucciones, escritas en códigos de programación e incorporando diferentes lenguajes, para la ejecución de diversas funciones para el análisis metataxonómico.
3. Inferir cual es el número de muestras gástricas que es necesario estudiar para obtener un poder estadístico del 95%.
4. Analizar cómo afecta la utilización de diferentes bases de datos de referencia (Greengenes 13,8 y Silva 132) a la reproducibilidad de los estudios metataxonómicos.
5. Analizar la microbiota gástrica de pacientes infectados o no por *H. pylori*, obtenida de diferentes estudios metataxonómicos, para extraer el efecto de esta bacteria en la microbiota gástrica, así como el efecto de la diferente patología que presenta el paciente y del país de realización del estudio.

Material y métodos

4. Material y métodos

4.1 Pacientes y muestras

Todas las muestras incluidas en este estudio proceden de pacientes a los que se les prescribió una gastroscopia con el fin de buscar alteraciones en la pared gástrica, tales como úlceras, tumores o zonas con histología compatible con gastritis. A la mayoría de los pacientes les fue prescrita dicha prueba debido a dolor abdominal, sospecha de gastritis, síntomas compatibles con reflujo gastroesofágico o diagnóstico de cáncer gástrico.

Se obtuvieron muestras de biopsia gástrica que fueron enviadas al laboratorio de microbiología para su cultivo y conservación para secuenciación.

Se realizó el cultivo de las cepas en una placa de medio selectivo, *Agar Pylori* (*Pylori* Agar; BioMèrieux), que se incubó en atmósfera microaerofílica (5% O₂, 10% CO₂, 85% N₂) a 37°C durante un máximo de 15 días. En este periodo de tiempo fueron revisadas periódicamente para detectar la presencia de colonias compatibles con *H. pylori*. Posteriormente, dichas colonias fueron cultivadas en agar sangre (Columbia Agar con sangre de oveja al 5%; Biomerieux) y conservadas en medio *Tryptic Soy Broth* (TSB) con 15% de glicerol para estudios posteriores.

4.2 Identificación de *H. pylori* por MALDI-TOF

4.2.1 Selección de las cepas

Se seleccionaron 15 cepas de *H. pylori* obtenidas de muestras locales para ser incorporadas en la base de datos del MALDI Biotyper OC 3.1.

Como criterio de aceptación, fue necesario realizar una identificación a varios niveles para poder confirmar dicha determinación. Con esta finalidad, se emplearon tanto técnicas moleculares como técnicas dependientes de cultivo.

a) Identificación por técnicas de visualización microscópica

Se llevó a cabo una impronta mediante el arrastrado de 4-5 colonias con una torunda por toda la superficie de un portaobjetos limpio previamente desengrasado y con 1 gota de solución salina continuando con la fijación de las bacterias por calor.

Para la tinción, se realizó la técnica de tinción de Gram modificada, utilizando fuchina durante 5 min.

Para la visualización con el microscopio se utilizó el objetivo de 100 aumentos con aceite de inmersión.

Solo se aceptaron microorganismos Gram negativos curvados como compatibles con *H. pylori*.

b) Identificación por técnicas dependientes de cultivo

Sobre las colonias que mostraron un aspecto compatible con *H. pylori* se realizaron las pruebas bioquímicas de ureasa, catalasa y oxidasa, aceptándose exclusivamente aquellas colonias que obtuvieron resultados positivos en todas las pruebas.

c) Identificación por técnicas moleculares

Varias colonias de un cultivo puro de cada cepa se resuspendieron en 300µl de solución salina. La extracción del material genético se realizó de forma automatizada por medio de la plataforma NucliSens® easyMAG (bioMérieux, Marcy l'Etoile, Francia), siguiendo el manual del fabricante y obteniéndose un volumen final de 25 µL.

Se realizó una PCR con el kit comercial GenoType® HelicoDR. Este experimento consistió en una PCR seguida de hibridación inversa que detecta la presencia de *H. pylori* y de las mutaciones genéticas más comunes que generan la resistencia de *H. pylori* a los antibióticos claritromicina y levofloxacina.

Una vez realizados todos los ensayos, se utilizaron subcultivos de no más de 48 horas para su incorporación en la base de datos del MALDI.

4.2.2 Extracción proteica e introducción en la base de datos

Se cultivó cada cepa de *H. pylori* en agar columbia de 3 a 5 días en CO₂ y se realizó una extracción proteica llevada a cabo según se describe a continuación: Se resuspendió una colonia en 330 µL de agua grado HPLC y 300µL de etanol puro (alcohol anhidrico) homogenizándose con el vortex durante 1 minuto.

Se centrifugó durante 2 minutos y eliminó el sobrenadante.

Se añadieron 20µL de acetonitrilo puro resuspendiéndose las bacterias precipitadas en el vortex 1 minuto.

Se incorporaron 20 µL de ácido fórmico al 70% y se mezcló con vortex.

Se centrifugó a máxima velocidad durante 2 minutos.

En la tarjeta de 96 posiciones se pusieron 2 µL del sobrenadante, por cada cepa de estudio, en 8 posiciones, una posición con BTS (Bacterial Test Standard consistente en un extracto de alfa DH5 *Escherichia coli*) y la última con BTS más el extracto. A todas las posiciones se les añadió 1µL de Matriz (ácido α-ciano-4-hidroxicinámico en solvente orgánico).

Una vez secada la matriz en todas las cepas y el BTS, se puso la tarjeta en el espectrómetro y se obtuvieron los espectros de cada una de las cepas siguiendo el protocolo de la casa comercial.

Se compararon los espectros obtenidos para cada una de las posiciones de cada cepa eliminándose aquellos desplazados o que se salieran de la tónica general. Todas las cepas analizadas y aquellas previamente incluidas por la base de datos de la casa comercial fueron comparadas mediante un análisis de componentes principales visualizándose los resultados mediante un dendograma de agrupaciones utilizando el software Biotyper OC 3.1

Material y métodos

4.2.3 Ensayo de mejora de la identificación por MALDI

Se estudiaron un total de 60 cepas de *H. pylori*. Para este experimento fueron transferidas a una placa metálica a las que se le añadió 1 µl de ácido fórmico y se esperó hasta su secado. Posteriormente, 1 µl de la solución matriz (ácido α-ciano-4-hidroxicinámico con solvente orgánico) fue añadido y secado.

Se evaluaron por medio de la base de datos MALDI Biotyper OC 3.1 original y, posteriormente, con la base de datos con las 15 cepas incorporadas.

4.2.4 Ensayo de factibilidad en clínica

Se estudiaron un total de 133 aislamientos de *H. pylori*. Cada muestra fue extendida sobre la placa de metal y se añadió 1µl de ácido fórmico y, posteriormente, 1µl de solución matriz (ácido α-ciano-4 hidroxicinámico + solvente orgánico).

La base de datos utilizada fue la MALDI Biotyper OC 3.1, a la que se incorporaron 15 cepas de *H. pylori* obtenidas de muestras locales siguiendo el procedimiento recomendado para el MALDI-TOF-MS (Bruker).

4.3 Análisis bioinformático para estudio taxonómico

Con el fin de poder establecer un protocolo de trabajo claro y personalizable, se han querido incluir los comandos y herramientas más utilizadas para poder realizar una investigación sin la necesidad de tener (o casi sin tener) amplios conocimientos de programación, pero entendiendo cada uno de los procesos para poder razonar que procesos realizar. Estos módulos proporcionan un molde para poder ir desarrollando en cada grupo los comandos que más se ajusten a sus necesidades.

4.3.1 Ensamblaje, control de calidad y agrupación de las secuencias

El proceso preliminar al análisis de las secuencias comienza con el estudio de su origen y de los protocolos utilizados para su secuenciación. Así pues, es necesario conocer de antemano si las regiones identificadoras o *barcodes* y las regiones de unión o *linkers* han sido sustraídos de las secuencias y conocer la región utilizada del gen 16S para poder realizar un mejor análisis taxonómico e intuir la longitud media esperada de las secuencias a analizar.

En este estudio se utilizó el programa FastQC (Andrews *et al.*, 2012) para comprobar la calidad de las secuencias y detectar la presencia de fragmentos de *barcodes* y/o *linkers*. Se analizó el tamaño de las secuencias y se visualizaron los gráficos de calidad por posición de las bases en busca de regiones con baja calidad de la secuenciación (generalmente encontradas en los extremos de las secuencias).

Posteriormente se utilizó el programa Oracle VM VirtualBox para poder ejecutar una máquina virtual con una imagen de la distribución de QIIME 2 Core – 2019.1 preinstalada en el sistema operativo Linux Ubuntu 18.04.1 LTS (Bionic Beaver) (Bolyen *et al.*, 2019).

A continuación se cargaron los archivos con las secuencias por medio de la terminal del sistema.

Se procedió al truncado de las regiones de baja calidad, ensamblado de las secuencias, corrección y eliminación del ruido mediante la herramienta DADA2 (Callahan *et al.*, 2016), agrupando las secuencias por ASVs.

4.3.2 Asignación taxonómica

La asignación taxonómica comienza con la elección de una base de datos de secuencias de referencia del gen 16S ARNr. En este caso se seleccionó la base de datos SILVA 132, mediante la plataforma QIIME2 con la herramienta *Classifier* (Bokulich *et al.*, 2016) mediante el algoritmo *sklearn* (Pedregosa *et al.*, 2012). Este algoritmo se sustenta en la creación de grupos o clases (en este caso, cada grupo será un valor taxonómico diferente que podrá estar incluido dentro de un grupo más grande (nivel taxonómico superior)). Para la creación de esos grupos es imprescindible la utilización de secuencias procedentes de una base de datos de referencia por lo que la elección de dicha base de datos tendrá mucha relevancia en los resultados posteriores. El algoritmo clasificará las secuencias representativas dentro de cada uno de los grupos mediante el uso del teorema de Bayes, empezando por los grupos taxonómicos más pequeños (especie o género), en caso de no obtener un nivel de significancia suficiente pasará a los niveles taxonómicos superiores (familia, orden, clase, etc.) hasta ser incluido o denominarlo no-clasificado obteniéndose, para cada secuencia, la máxima asignación taxonómica (MAT). Con las secuencias representativas, o con su identificación taxonómica, se obtendrán las tablas de abundancias absolutas de cada secuencia por muestra.

Los comandos utilizados para la plataforma QIIME se disponen en el anexo I.

4.3.3 Análisis metataxonómico en muestras gástricas

En primer lugar, en el caso de tener las muestras separadas por estudios o archivos, se debe realizar la fusión de las tablas de abundancias absolutas. Para ello se utilizó la función “merge ()” de R, la cual va incorporando las asignaciones taxonómicas por filas y las muestras por columnas.

Para el correcto funcionamiento del script es necesario separar la tabla de abundancias absolutas en 3 tablas diferentes: a nivel de ASV/OTU/MATs, a nivel de género y a nivel de filo. Para ello se utilizó el programa EXCEL.

a) Estimación de la diversidad microbiana

Para estimar la diversidad microbiana se utilizaron las tablas de abundancia absoluta mediante la normalización de cada una de las muestras. Para ello se transformó el número de secuencias detectadas de cada nivel taxonómico de cada muestra al porcentaje relativo del número total de secuencias de cada muestra. Con ello se obtiene el porcentaje de abundancia relativa de cada taxón en cada una de las muestras.

La diversidad alfa se calculó usando diferentes medidas: la medida de la riqueza de especies o ASVs observadas, la medida del índice de Shannon y la medida del índice de Simpson. La significación estadística entre grupos de la alfa diversidad se calculó mediante una prueba no paramétrica en función del número de factores de la variable a estudiar: la prueba de Wilcoxon-Mann-Whitney, si la variable tiene 2 factores, o la prueba de Kruskal-Wallis si la variable presenta 3 o más factores, con el cálculo posterior de la prueba de Wilcoxon-Mann-Whitney por cada par de factores. Al realizarse comparaciones múltiples, se ajustaron todos los p-valores mediante la prueba de Bonferroni.

El análisis de la diversidad beta se llevó a cabo calculando tanto las matrices de distancia de Bray-Curtis como las matrices de distancia mediante el coeficiente de Jaccard binario con la tabla de abundancias normalizada. Con ambas matrices se realizó una prueba MANOVA con 999 permutaciones "PERMANOVA". Se utilizó una prueba de permutaciones para determinar si la dispersión de las muestras en alguno de los grupos con respecto a su centroide presentaba diferencias significativas con uno o más grupos. Para la comparación multinivel a pares se realizó la prueba PERMANOVA con una corrección de los valores P por el método de Holm-Bonferroni con la librería de R "pairwiseAdonis" (Versión: 0.0.1).

También se realizó un análisis de agrupamiento jerárquico utilizando como conjunto de diferencias una serie de filos o géneros mayoritarios. El método de agrupación es la agrupación por unión completa donde se agrupan las muestras o grupos de muestras por la distancia más corta. La distancia es, entre la serie de comparaciones de todos los pares de muestras de 2 agrupaciones, la distancia entre esos dos elementos (uno en cada grupo) que están más alejados entre sí.

Material y métodos

Los diagramas de Venn se elaboraron con el objeto de determinar una microbiota *core*. Se pusieron una serie de condicionantes: una abundancia relativa de los filos o de los géneros mínima y un porcentaje de las muestras agrupadas según el factor de estudio que cumplieran dicha abundancia relativa mínima. En este estudio se seleccionaron los filos o géneros con una abundancia relativa mayor a 0,01%, 0,1% y 1% en, al menos, un 80% de las muestras en uno o más grupos. Para determinar posibles interacciones entre los géneros se realizó una red de correlaciones entre los géneros mayoritarios. Debido a la condición del *script* el número máximo de correlaciones será 20, siempre y cuando el tamaño muestral lo permita. Con tamaños muestrales donde el grupo de muestras no llegue a 21 se analizarán $n-1$ géneros, siendo n el número de muestras incluidas en el factor con menor número de muestras.

Todos los análisis y gráficos fueron representados en R mediante el *script* descrito en el anexo II y anexo III.

Para la realización del *script* se utilizaron diferentes librerías o paquetes de funciones de R que pueden catalogarse en: librerías de análisis, librerías gráficas y librerías de personalización de gráficas y tablas.

- Librerías de análisis:

Las librerías de análisis son, como su propio nombre indica, aquellas librerías que aportan funciones de análisis al *script*.

-Vegan (versión:2.5.6) (Oksanen *et al.*, 2019): el paquete de funciones Vegan, diseñado originalmente para el análisis de comunidades ecológicas, aporta las herramientas para el análisis de la diversidad, tanto para la diversidad alfa (con los algoritmos para el cálculo de los índices de Shannon, Simpson y Fisher) como para el análisis de la diversidad beta (con los algoritmos para calcular los índices de Gower, Bray–Curtis, Jaccard binomial, Kulczynski, Morisita, Horn–Morisita y Chao).

-Tidyverse (version:1.2.1) (Wickham, 2017) el paquete Tidyverse es, en si mismo, un conjunto de paquetes para el análisis de datos. Incluye tanto paquetes de análisis como paquetes de visualización de resultados

incluyendo paquetes como `forecast`, `tidyr` o `ggplot2` que veremos más adelante.

-`FactoMineR` (versión:1.42) (Lê, Josse, & Husson, 2008): es un paquete de R dedicado al análisis de datos multivariante especialmente utilizado para el análisis mediante PCAs (análisis de componentes principales) de variables no correlacionadas. Incluye funciones para la creación de gráficos y estadísticos.

-`PairwiseAdonis` (versión:0.0.1): esta herramienta facilita los análisis pareados de conjuntos de muestras para diferenciar aquellos conjuntos que presentan diferencias estadísticamente significativas entre sí.

-`Magrittr` (version:1.5) (Bache & Wickham, 2014): este paquete incluye una función operacional que permite agilizar la escritura del *script* al conducir el resultado de una función a otra sin necesidad de crear una variable intermedia.

-`Reshape2` (versión:1.4.3) (Wickham, 2007): esta herramienta facilita el análisis de los datos al facilitar la transformación de las tablas de datos o *dataframes* en listas continuas. Es especialmente útil para preparar tablas/listas intermedias para facilitar su representación gráfica.

-`Gtools` (versión:3.8.1) (Warnes, Bolker, & Lumley, 2018): paquete que incluye funciones para la realización de combinaciones múltiples o control de valores nulos, vacíos o perdidos (NA).

- **Librerías gráficas**

Las librerías gráficas contribuyen a lograr una mejor comprensión del tema desarrollado de los resultados obtenidos. Están diseñadas para resumir información. Son especialmente útiles cuando la cantidad de datos es tan grande que sólo un gráfico es capaz de ayudar a visualizar todo el conjunto sin perder los puntos más significativos.

-`Ggplot2` (versión:3.2.0) (Wickham, 2016): esta librería es una de las mejores herramientas que tiene R para realizar representaciones gráficas de los datos de manera relativamente sencilla e intuitiva. Su conocimiento permite crear múltiples tipos de gráficos y personalizarlos con gran libertad.

Material y métodos

-Factoextra (versión:1.0.5.999) (Kassambara & Mundt, 2017): herramienta ligada a FactoMinerR para realizar las representaciones gráficas de manera fluida tras analizar los datos con FactoMinerR. Incluye múltiples herramientas de representación de análisis multivariable.

-Qgraph (versión: 1.6.3) (Epskamp *et al.*, 2012): paquete de funciones para la representación gráfica de redes de correlaciones de una manera más sencilla. Su simplicidad puede, a su vez, limitar el tipo de análisis y gráficos obtenidos.

-Gplots (versión:3.0.1.1) (Warnes *et al.*, 2019): como Ggplot2, es una herramienta que aporta funciones para la realización de gráficos. En este caso presenta gráficos un poco más especializados (gráficos de cajas o gráficos de calor).

-Ggbiplot (versión: 0.55) (Vu, 2011): herramienta ligada a Ggplot2, Ggbiplot incorpora algunas funciones adicionales a para la representación gráfica de los PCoAs.

-Venn (versión:1.7) (Dusa, 2018): esta herramienta facilita la creación de los diagramas de Venn, incluyendo la generación de tablas intermedias y su representación gráfica directamente.

- Librerías de personalización de gráficas y tablas

Las librerías o paquetes de personalización contribuyen, generalmente, a mejorar el diseño del informe final, facilitando la elección de colores para las gráficas o aportando diferentes diseños de tablas que ayudan a hacer más comprensibles tanto las gráficas como las tablas de resultados.

-RColorBrewer (version:1.1.2) (Neuwirth, 2014): este paquete incluye una selección de paletas de colores listas para poder ser utilizadas en las diferentes herramientas de creación de gráficos como Ggplot2.

-Viridis (version:0.5.1) (Garnier, 2018): esta herramienta permite seleccionar gamas de colores continuas en forma de degradados. Los colores que aporta son especialmente interesantes para la impresión en escala de grises.

-Forcats (version:0.4.0) (Wickham, 2019): este paquete de funciones son útiles tanto para los análisis como para la representación gráfica. Facilita la reordenación de los niveles de los factores dentro de una variable. Esto

afecta a como el programa entiende dicha variable y facilita, por ejemplo, que unos factores aparezcan antes o después en los gráficos.

-Ggpubr (version:0.2.1) (Kassambara, 2019): esta herramienta facilita la colocación de p-valores en un gráfico de ggplot como pueden ser los diagramas de cajas.

-Knitr (version:1.23) (Xie, 2019): herramienta para realizar cambios visuales en el formato de las tablas obtenidas al realizar el *script*.

b) Selección y estudio de los fillos y géneros mayoritarios

Para la selección de los fillos o géneros mayoritarios se seleccionaron aquellos con una abundancia relativa promedio mayor al resto. El número de fillos a géneros seleccionados puede variar, siendo 5 fillos y 19 géneros los seleccionados en este estudio. El resto de los fillos y géneros minoritarios se agruparon en un único fillo o género denominado “minoritarios”.

4.4 Selección de proyecto para el cálculo del tamaño muestral

Para el cálculo del tamaño muestral necesario para obtener un poder estadístico del 95% se utilizaron las secuencias del estudio realizado por Llorca *et al.* (2017) en el cual caracterizaron la microbiota gástrica por primera vez de pacientes pediátricos. Para ello seleccionaron 51 pacientes pediátricos con patología gástrica, reclutados desde octubre del año 2013 a enero del año 2014 en la sección pediátrica de Gastroenterología del Hospital Niño Jesús y del Hospital Doce de Octubre (Madrid, España), a los que les prescribieron la realización de una endoscopia gástrica con la toma de biopsia de antro gástrico.

El diagnóstico de la infección por *H. pylori* fue realizado mediante 4 métodos: cultivo, prueba rápida de la ureasa, histología compatible con la infección por *H. pylori* y qPCR. Se consideraron positivas aquellas muestras que fueron positivas por cultivo o positiva en, al menos, 2 de las otras pruebas diagnósticas.

Con este criterio 33 pacientes se consideraron negativos y 18 positivos para infección por *H. pylori*. Con el objetivo de reducir la variabilidad entre las muestras solo se seleccionaron aquellas que presentaban un número mínimo de 1.000 secuencias pareadas y menos de un 25% de indeterminaciones.

De las 51 muestras, 39 cumplieron las condiciones para el estudio (Tabla 1).

Tabla 1: Principales características de los pacientes y su diagnóstico de las muestras de Llorca *et al.* (2017).

España	Mujeres	Hombres	Total
Edad años (DESVEST)	11,5 (3,73)	10,26 (3,73)	10,77 (3,70)
H.p.-	10 (25,64%)	13 (33,33%)	23 (58,97%)
Dispepsia	5 (12,82%)	7 (17,95%)	12 (30,76%)
Gastritis	5 (12,82%)	6 (15,38%)	11 (28,20%)
Cáncer	0 (0,00%)	0 (0,00%)	0 (0,00%)
H.p.+	6 (15,38%)	10 (25,64%)	16 (41,03%)
Dispepsia	1 (2,56%)	2 (5,13%)	3 (7,69%)
Gastritis	5 (12,82%)	8 (20,51%)	13 (33,33%)
Úlcera	0 (0,00%)	0 (0,00%)	0 (0,00%)
Metaplasia	0 (0,00%)	0 (0,00%)	0 (0,00%)
Cáncer	0 (0,00%)	0 (0,00%)	0 (0,00%)
Total general	16 (41,03%)	23 (58,97%)	39 (100%)

4.5 Selección del resto de proyectos y muestras

Se realizó una búsqueda de diferentes proyectos independientes entre los años 2004 y 2019 en la plataforma ENA utilizando las palabras clave “*Gastric*”, “*Gastric microbiome*”, “*Gastric microbiota*” y “*Gastric microbiota illumina*”. Los factores que hicieron elegibles a los proyectos eran: muestras gástricas, humanas, secuenciadas por medio de una plataforma Illumina y cuyos datos de secuenciación y demográficos fueran accesibles. Además, era imprescindible que dispusieran de los siguientes datos: país, estado de infección por *H. pylori* y región del gen 16S secuenciada, siendo deseable conocer el estado de salud o diagnóstico del paciente. Así mismo, solo se seleccionaron aquellas muestras que presentaban un número mínimo de 1.000 secuencias pareadas y menos de un 25% de indeterminaciones.

Se seleccionaron para el estudio 5 proyectos diferentes: PRJEB21497 (Castaño-Rodríguez *et al.*, 2017), PRJEB22107 (Klymiuk *et al.*, 2017), PRJEB21104 (Parsons *et al.*, 2017), PRJNA310127 (Yu *et al.*, 2017) y PRJNA313391 (2016). Las muestras de estos proyectos, junto con las del proyecto de LLorca *et al.* (2017) sumaban un total de 660 muestras gástricas de pacientes desde 2004 a 2016.

Material y métodos

Proyecto PRJEB21497 (Castaño-Rodríguez *et al.*, 2017):

Se caracterizó la microbiota gástrica de 36 pacientes adultos procedentes de Malasia y República de Singapur con diferentes desórdenes gastrointestinales diagnosticados a partir de una biopsia de antro gástrico: 12 pacientes recientemente diagnosticados de cáncer gástrico no cardias, 20 pacientes con dispepsia funcional y 4 con úlcera gástrica. El diagnóstico para *H. pylori* se realizó por medio de detección de anticuerpos IgG para *H. pylori* en suero.

De las 36 muestras, 33 cumplen las condiciones para el estudio (tabla 2).

Tabla 2: Principales características de los pacientes cuyas muestras han sido incluidas y su diagnóstico (PRJEB21497).

Malasia	Mujeres	Hombres	Total
Edad años (DESVEST)	55,67 (12,03)	56,39 (12,77)	56,06 (12,25)
H.p.-	3 (9,09%)	6 (18,18%)	9 (27,27%)
Dispepsia	3 (9,09%)	5 (15,15%)	8 (24,24%)
Gastritis	0 (0,00%)	0 (0,00%)	0 (0,00%)
Cáncer	0 (0,00%)	1 (3,03%)	1 (3,03%)
H.p.+	12 (36,36%)	12 (36,36%)	24 (72,73%)
Dispepsia	5 (15,15%)	5 (15,15%)	10 (30,303%)
Gastritis	0 (0,00%)	0 (0,00%)	0 (0,00%)
Úlcera	0 (0,00%)	4 (12,12%)	4 (12,12%)
Metaplasia	0 (0,00%)	0 (0,00%)	0 (0,00%)
Cáncer	7 (21,21%)	3 (9,09%)	10 (30,30%)
Total general	15 (45,45%)	18 (54,55%)	33 (100%)

Proyecto PRJEB22107 (Klymiuk *et al.*, 2017)

Se caracterizó la microbiota gástrica de 30 pacientes adultos procedentes de Austria con diferentes desórdenes gastrointestinales diagnosticados a partir de una biopsia de antro gástrico. El diagnóstico para el estado de *H. pylori* se realizó mediante PCR específica.

De las 30 muestras, 27 cumplen las condiciones para el estudio (tabla 3).

Tabla 3: Principales características de los pacientes cuyas muestras han sido incluidas y su diagnóstico (PRJEB22107).

Austria	Mujeres	Hombres	Total
Edad años (DESVEST)	50,53 (18,78)	48,58 (11,61)	49,67 (15,74)
H.p.-	3 (11,11%)	6 (22,22%)	9 (33,33%)
Dispepsia	2 (7,41%)	4 (14,81%)	6 (22,22%)
Gastritis	1 (3,70%)	2 (7,41%)	3 (11,11%)
Cáncer	0 (0,00%)	0 (0,00%)	0 (0,00%)
H.p.+	12 (44,44%)	6 (22,22%)	18 (66,67%)
Dispepsia	3 (11,11%)	1 (3,70%)	4 (14,81%)
Gastritis	9 (33,33%)	4 (14,81%)	13 (48,15%)
Úlcera	0 (0,00%)	1 (3,70%)	1 (3,70%)
Metaplasia	0 (0,00%)	0 (0,00%)	0 (0,00%)
Cáncer	0 (0,00%)	0 (0,00%)	0 (0,00%)
Total general	15 (55,56%)	12 (44,44%)	27 (100%)

Material y métodos

Proyecto PRJEB21104 (Parsons *et al.*, 2017):

Se caracterizó la microbiota gástrica de 95 pacientes adultos residentes en Inglaterra con diferentes desórdenes gastrointestinales que requerían de una biopsia de antro gástrico. En dicho estudio se analizó el efecto del uso de un inhibidor de bomba de protones y el efecto de la atrofia autoinmune en la microbiota intestinal. Aquellas muestras que no cumplieron con los requisitos de esta tesis no fueron incluidas. El diagnóstico a infección por *H. pylori* se realizó mediante el resultado positivo de alguna de las siguientes pruebas: prueba rápida de ureasa, prueba serológica o histología compatible con infección por *H. pylori*.

De las 95 muestras presentadas en su artículo, 65 cumplen las condiciones para el estudio (tabla 4).

Tabla 4: Principales características de los pacientes cuyas muestras han sido incluidas y su diagnóstico (PRJEB21104).

UK	Mujeres y hombres
Sexo	57,5 % Mujeres
Rango de edad	60 (48-70)
H.p.-	23 (38,98%)
Dispepsia	18 (27,69%)
Gastritis	0 (0,00%)
Cáncer	5 (7,69%)
H.p.+	42 (64,62%)
Dispepsia	0 (0,00%)
Gastritis	26 (40,00%)
Úlcera	0 (0,00%)
Metaplasia	3 (4,62%)
Cáncer	7 (10,77%)
Total general	65 (100%)

Proyecto PRJNA310127 (Yu *et al.*, 2017):

Se caracterizó la microbiota gástrica de 80 pacientes adultos residentes en China y 80 pacientes residentes en México todos con cáncer gástrico. En este caso no se valoró el estado de infección por *H. pylori* considerándose solo la presencia de la bacteria en las muestras de biopsia

De las 160 muestras presentadas en su artículo, 59 (34 de México y 25 de China) cumplen las condiciones para el estudio (tabla 5).

Tabla 5: Principales características de los pacientes cuyas muestras han sido incluidas y su diagnóstico (PRJNA310127).

China / México (Portugal)	Mujeres	Hombres	Total
H.p.-	0 (0,00%)	0 (0,00%)	0 (0,00%)
Dispepsia	0 (0,00%)	0 (0,00%)	0 (0,00%)
Gastritis	0 (0,00%)	0 (0,00%)	0 (0,00%)
Cáncer	0 (0,00%)	0 (0,00%)	0 (0,00%)
H.p.+	23 (38,98%)	36 (61,02 %)	59 (100%)
Dispepsia	0 (0,00%)	0 (0,00%)	0 (0,00%)
Gastritis	0 (0,00%)	0 (0,00%)	0 (0,00%)
Úlcera	0 (0,00%)	0 (0,00%)	0 (0,00%)
Metaplasia	0 (0,00%)	0 (0,00%)	0 (0,00%)
Cáncer	23 (38,98%)	36 (61,02 %)	59 (100%)
Total general	23 (38,98%)	36 (61,02 %)	59 (100%)

Material y métodos

Proyecto PRJNA313391:

Se caracterizó la microbiota gástrica de 119 pacientes adultos residentes de China por el hospital municipal de Qingdao con infección por *H. pylori*. Estas muestras se hicieron públicas sin estar asociadas a la publicación de un artículo y solo se dispone de los datos facilitados a la hora de depositar las secuencias entre las cuales no se encuentra el método de extracción ni el método de diagnóstico para la infección por *H. pylori*. Se dispone del sexo, secuenciador utilizado, estado de infección y patología asociada entre los datos asociados a las secuencias.

De las 119 muestras, 108 cumplen las condiciones para el estudio (tabla 6).

Tabla 6: Principales características de los pacientes cuyas muestras han sido incluidas y su diagnóstico (PRJNA313391).

China	Mujeres	Hombres	Total
Edad años (DESVEST)	55,53 (12,33)	55,75 (12,76)	55,69 (12,36)
H.p.-	0 (0,00%)	0 (0,00%)	0 (0,00%)
Dispepsia	0 (0,00%)	0 (0,00%)	0 (0,00%)
Gastritis	0 (0,00%)	0 (0,00%)	0 (0,00%)
Cáncer	0 (0,00%)	0 (0,00%)	0 (0,00%)
H.p.+	32 (29,63%)	76 (70,37%)	108 (100%)
Dispepsia	0 (0,00%)	0 (0,00%)	0 (0,00%)
Gastritis	12 (11,11%)	17 (15,74%)	29 (26,85%)
Úlcera	0 (0,00%)	0 (0,00%)	0 (0,00%)
Metaplasia	6 (5,56%)	20 (18,52%)	26 (24,07%)
Cáncer	14 (12,96%)	39 (36,11%)	53 (49,07%)
Total general	32 (29,63%)	76 (70,37%)	108 (100%)

Resultados

5. Resultados

5.1 Aplicación de bioinformática a estudios proteómicos

Los análisis por medio de MALDI TOF de 60 cepas de *H. pylori* provenientes de cultivos de 60 muestras de biopsias gástricas se muestran en la tabla 7.

Se puede observar como inicialmente, el 45% de las cepas presentaron una puntuación menor de 1,6 al ser analizadas con la base de datos Biotyper OC 3.1. Con la incorporación de las cepas locales se obtuvo una mejoría en la identificación de las 60 cepas, mejorándose a su vez cuando se combinan las 2 bases de datos.

Tabla 7: Número y porcentaje por puntuación de cepas identificadas como *H. pylori* en función de la base de datos utilizada.

Puntuación	Base de datos Biotyper OC 3.1	Base de datos de cepas locales	Bases de datos combinadas
	Número de cepas (%)	Número de cepas (%)	Número de cepas (%)
>=2	0 (0%)	10 (16,67%)	10 (16,67%)
1,8-2	12 (20%)	28 (46,67%)	28 (46,67%)
1,6-1,79	21 (35%)	14 (23,33%)	14 (23,33%)
1,4-1,59	20 (33,33%)	5 (8,33%)	7 (11,67%)
<1,4	7 (11,67%)	3 (5%)	1 (1,67%)
Media	1,640±0,187	1,818±0,205	1,832±0,196
Total	60	60	60

Se puede observar que el 13% de las cepas presentaron una puntuación menor de 1.6 al utilizar la base de datos locales o las bases de datos combinadas.

Al analizar la base de datos del software MALDI Biotyper OC 3.1 de Bruker se encontraron 7 cepas de *H. pylori* pertenecientes a colecciones internacionales. Al incorporar las 15 cepas locales y realizar el estudio por solapamiento de picos de las cepas de *H. pylori* se aprecian 3 agrupaciones diferentes (Fig. 2) y se distingue una agrupación de las cepas locales separadas de las aportadas por la base de datos Bruker con la excepción de la cepa *H. pylori* 151 RLT.

Resultados

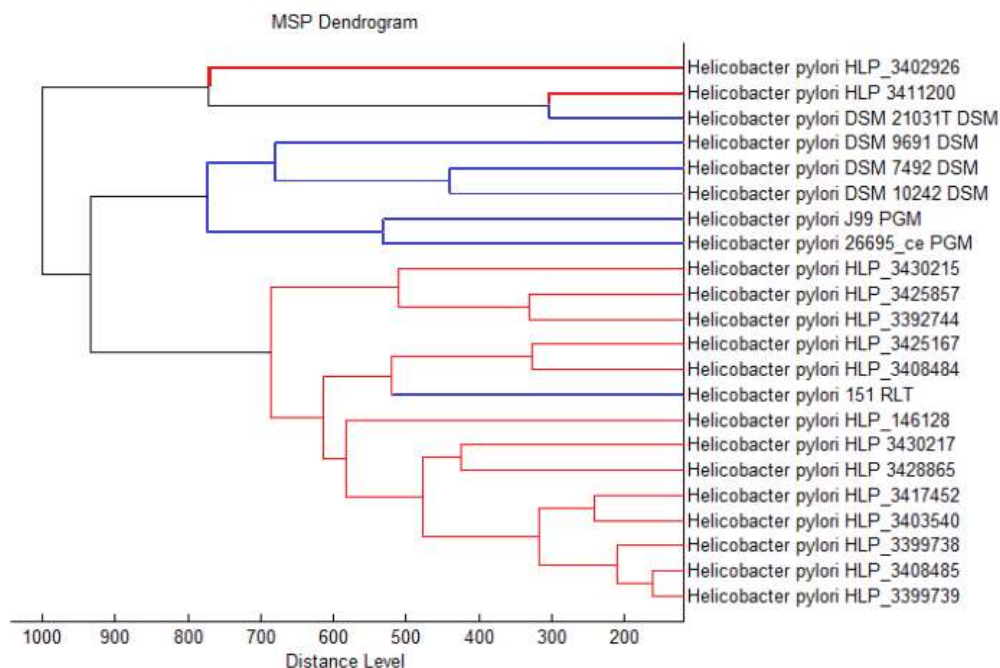


Figura 2: Dendrograma de las cepas de *H. pylori* presentes en la base de datos Bruker (azul) y las incorporadas a la base de datos local (HLP, rojo).

Para determinar la importancia de su implementación a nivel clínico se analizaron de rutina un total de 133 aislamientos de *H. pylori* tras la incorporación de las cepas locales a la base de datos del MALDI-TOF. Los resultados concordaron con los obtenidos previamente, tal y como se indica en la tabla 8.

Tabla 8: Número y porcentaje de cepas identificadas como *H. pylori* con sus valores por MALDI con las bases de datos combinadas.

Puntuación	Número de cepas (%)
> 2	20 (15,00%)
1,7- =<2	77 (57,90%)
1,4 - =<1.7	32 (24,10%)
=< 1,4	4 (3,00%)
Total	133

A pesar de la alta variabilidad que presentan las cepas de *H. pylori* entre sí, el MALDI-TOF es un método eficaz para la identificación de aislamientos clínicos, aunque puede ser necesario incluir cepas locales en la base de datos para considerar identificaciones aceptables con scores a partir de 1,7.

5.2. Documento de instrucciones para el análisis metataxonómico

En el anexo I se encuentra el documento con instrucciones para la ejecución de diversas tareas para el análisis bioinformático metataxonómico.

Las funciones incluidas son las siguientes:

- Ensamblaje, control de calidad y agrupación de las secuencias procedentes de la plataforma Illumina.
- Asignación taxonómica.
- Preparación de las tablas de abundancia por muestra.

En el anexo II se encuentra el documento con instrucciones escritas en códigos de programación, para directamente realizar las siguientes tareas:

- Preparación y normalización de las tablas de abundancias.
- Estimación de la diversidad microbiana: diversidad alfa y diversidad beta.
- Análisis de agrupamiento jerárquico.
- Elaboración de diagramas de Venn para determinar una microbiota *core*.
- Selección y estudio de los filis y géneros mayoritarios.

Resultados

5.3 Cálculo del tamaño muestral

El estudio del tamaño muestral necesario para obtener una potencia estadística de, al menos, un 0.95 se realizó con las secuencias de biopsia gástrica de Llorca *et al*, (2017) analizadas con la base de datos SILVA 132.

El estudio presenta los valores de los índices de diversidad de Shannon y Simpson con sus desviaciones estándar para un tamaño muestral de 39 biopsias de antro gástrico divididas en 2 grupos de pacientes ambos con síntomas dispépticos: 16 muestras provenientes de pacientes diagnosticados como positivos a infección por *H. pylori* y 23 pacientes con diagnóstico negativo a *H. pylori* (Tabla 9).

Tabla 9: Cálculo de los índices de diversidad en función del estado de infección de los pacientes por *H. pylori* y potencia de la prueba estadística WilkonxonMann-Whitne

Índices	Media± DESVEST		Tamaño del efecto	Tamaño muestral requerido (Poder 0.95)
	Hp Negativo (n=23)	Hp positivo (n=16)		
Shannon	1,76±0,68	1,18±0,63	0,82	95
	0,61±0,19	0,44±0,21		
Simpson	0,61±0,19	0,44±0,21	0,79	98

El tamaño del efecto que ejerce la condición infección por *H. pylori* sobre los índices de diversidad es grande (aprox. 0.8) por lo que el número total de muestras necesario para poder detectar esas diferencias con un poder estadístico de 0.95 es de entre 95 y 100 muestras.

Como se puede ver gráficamente en la figura 3, un aumento del tamaño muestral por encima de 100 aumentaría muy ligeramente la potencia del estadístico para determinar el efecto de la infección por *H. pylori* sobre la diversidad bacteriana gástrica reflejada por los índices de Simpson y Shannon.

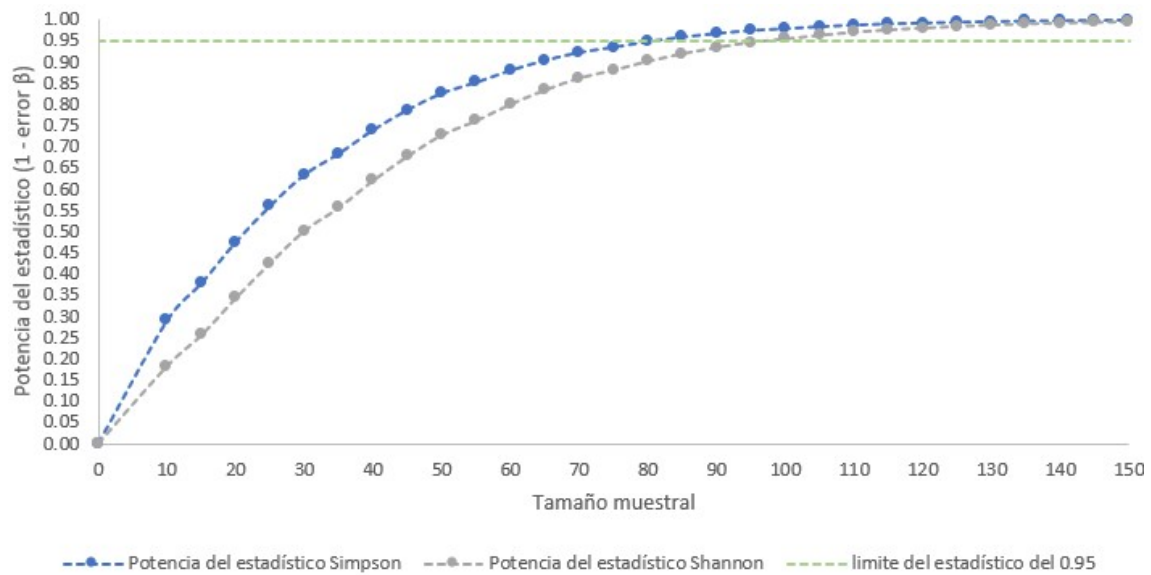


Figura 3: Potencia de la prueba estadística WilkonxonMann-Whitne para las comparaciones de los índices de Simpson y Shannon de acuerdo con el tamaño muestral.

Por lo tanto, son necesarias, al menos, 100 muestras para obtener un poder estadístico del 0.95% en la comparación del índice de Shannon entre pacientes diagnosticados como positivos o negativos para la infección por *H. pylori*, siendo lo ideal contar con, al menos, 140 muestras para un estudio donde las desviaciones estándar sean pequeñas. Como no se sabe si la microbiota gástrica es constante a lo largo de los países y en los diferentes estudios, se utilizará un número mayor de muestras.

Resultados

5.4 Determinación del efecto de la utilización de diferentes bases de datos de referencia en estudios metataxonómicos

El efecto de la elección de la base de datos utilizada sobre los resultados obtenidos se realizó con las muestras de Llorca *et al.* (2017).

Se analizaron las mismas secuencias de alta calidad con las bases de datos greengenes GG_13 (2013) y Silva 132 (2018).

Debido al gran efecto que causan las secuencias de *H. pylori* por su elevada abundancia relativa se procedió a normalizar su asignación taxonómica en ambas bases de datos.

Así pues, en la base de datos Greengenes el género *Helicobacter* está incluido de la siguiente forma:

```
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter
```

Mientras que en la base de datos Silva, aparece de la siguiente forma:

```
D_0__Bacteria.D_1__Epsilonbacteraeota.D_2__Campylobacteria.D_3__Campylobacterales.D_4__Helicobacteraceae.D_5__Helicobacter
```

Lógicamente se refieren a la misma bacteria y no realizar estos cambios haría imposible la comparación del resto de los microorganismos.

Una vez realizado el cambio taxonómico se analizaron las secuencias de las 39 muestras que pasaron los criterios de inclusión mediante el uso de las base de datos SILVA 132 y Greengenes 13.8 obteniéndose un total de 357 asignaciones taxonómicas diferentes provenientes de 587.232 secuencias totales.

5.4.1 Diversidad alfa

El análisis de las secuencias analizadas no presenta diferencias significativas por el uso de las bases de datos Silva y/o Greengenes en los valores de los índices de Shannon o Simpson (p-valor 0.84 y p-valor = 0.97 respectivamente) (Fig 4).

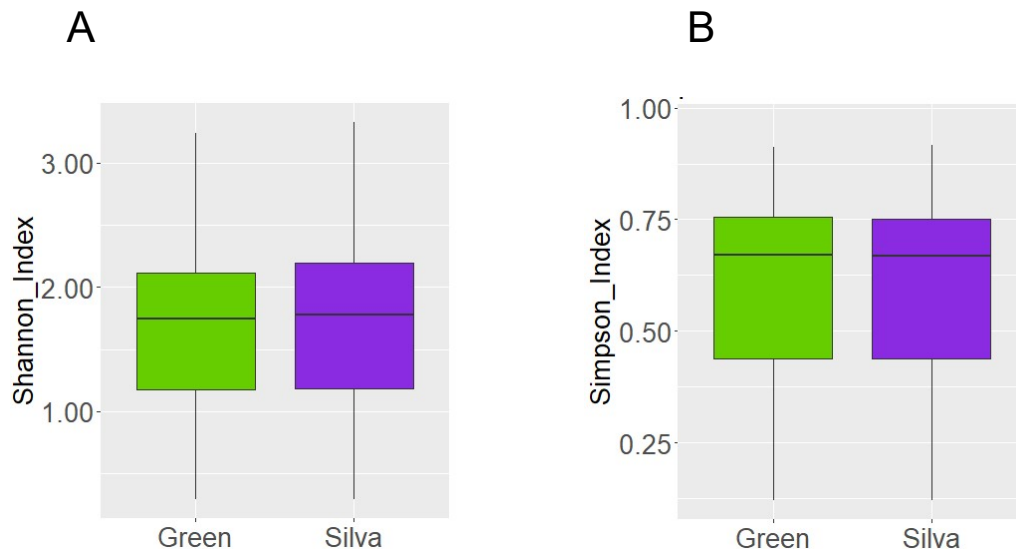


Figura 4: Valores de Shannon (A) y Simpson (B) de acuerdo con la base de datos utilizada (Green= Greengenes 13.8, en verde y Silva= SILVA 132, morado).

5.4.2 Diversidad beta

El análisis de la diversidad beta, calculada de acuerdo con la abundancia relativa de asignaciones taxonómicas (Bray-Curtis) y la presencia/ausencia de dichas asignaciones (matriz de distancias de Jaccard binaria), indicaron que las muestras analizadas se agrupan en función del perfil microbiológico obtenido según la base de datos utilizada (p valor < 0.001 y p valor < 0.001, respectivamente) obteniéndose patrones similares espacialmente separados entre sí (Fig. 5).

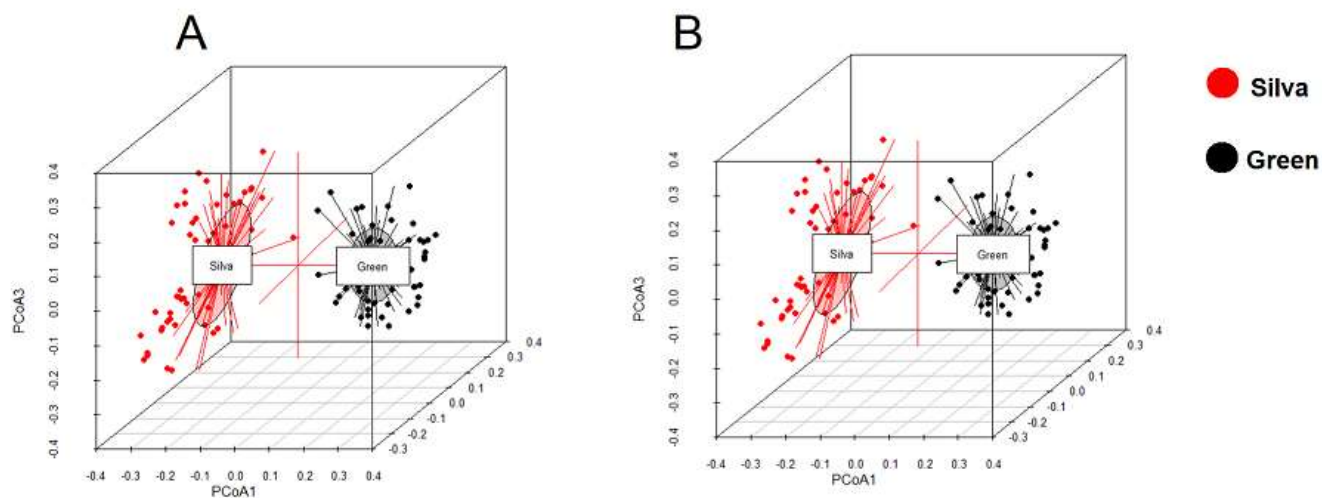


Figura 5: PCoAs de los perfiles microbiológicos de las muestras de biopsia gástrica en función de la abundancia relativa de las diferentes asignaciones taxonómicas por medio de la matriz de distancia de Bray-Curtis (A) y la presencia/ausencia de dichas asignaciones por medio de la matriz de distancias de Jaccard binaria (B) obtenidas por medio de la base de datos Silva 132 (Silva, rojo) y Greengenes 13,8 (Green, negro).

5.4.3 Asignaciones taxonómicas

Al analizar las mínimas asignaciones taxonómicas de las secuencias analizadas se aprecia una correlación entre las diferentes taxonomías asignadas por las bases de datos utilizadas. Cuando la asignación taxonómica permanece constante entre las bases de datos, no se encuentran diferencias significativas en la abundancia relativa de las dichas secuencias. Por el contrario, se aprecian diferencias en aquellas asignaciones taxonómicas que han podido sufrir un cambio en la taxonomía (como es el ejemplo del género *Rothia* que fue reasignada del orden Actinomycetales al orden Micrococcales o del género *Helicobacter* que fue reasignada del filo Proteobacteria al filo Epsilonbacteraeota), o que han sido clasificadas en niveles taxonómicos más profundos por la base de datos Silva (Tabla 10).

Al analizar los 20 géneros mayoritarios solo se obtuvieron diferencias significativas en 3 de ellos, coincidiendo 2 de ellos con géneros no detectados por la base de datos Greengenes 13.8 (Tabla 11).

Resultados

Tabla 10: Comparativa de las 20 asignaciones taxonómicas más abundantes. En gris posibles reasignaciones taxonómicas.

Mínima asignación taxonómica (MAT)	Greengenes 13.8		SILVA 132		P-value†
	n (%)	Mediana (IQR)	n (%)	Median (IQR)	
Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Pseudomonadaceae <i>Pseudomonas</i>	38 (97,44%)	28,96 (5,62-51,82)	38 (97,44%)	29,11 (5,85-52,1)	0,88
Bacteria Epsilonbacteraeota Campylobacteria Campylobacteriales <i>Helicobacteraceae</i> <i>Helicobacter</i>	28 (71,79%)	2,23 (<0,01-43,25)	28 (71,79%)	2,23 (<0,01-43,27)	0,92
Bacteria Firmicutes Bacilli Bacillales Staphylococcaceae <i>Staphylococcus</i>	29 (74,36%)	0,12 (0,01-0,91)	29 (74,36%)	0,12 (0,01-0,91)	0,91
Unassigned	36 (92,31%)	2,98 (0,32-10,51)	0 (0%)	<0,01 (<0,01-<0,01)	<0,001
Bacteria	0 (0%)	<0,01 (<0,01-<0,01)	36 (92,31%)	2,88 (0,22-8,9)	<0,001
Bacteria Firmicutes Bacilli Lactobacillales Streptococcaceae <i>Streptococcus</i>	33 (84,62%)	0,27 (0,05-0,91)	33 (84,62%)	0,27 (0,05-0,91)	0,92
Bacteria Bacteroidetes Bacteroidia Bacteroidales S247	33 (84,62%)	1,03 (0,02-8,45)	0 (0%)	<0,01 (<0,01-<0,01)	<0,001
Bacteria Bacteroidetes Bacteroidia Bacteroidales Muribaculaceae unculturedbacterium	0 (0%)	<0,01 (<0,01-<0,01)	33 (84,62%)	0,25 (0,02-7,3)	<0,001
Bacteria Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae <i>Bacteroides</i>	29 (74,36%)	0,5 (0,01-2,25)	29 (74,36%)	0,5 (0,01-2,25)	0,91
Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae	31 (79,49%)	0,27 (0,02-1,17)	0 (0%)	<0,01 (<0,01-<0,01)	<0,001
Bacteria Proteobacteria Gammaproteobacteria Pseudomonadales Moraxellaceae <i>Acinetobacter</i>	35 (89,74%)	1,2 (0,3-2,86)	35 (89,74%)	1,2 (0,3-2,86)	0,94
Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae <i>EscherichiaShigella</i>	0 (0%)	<0,01 (<0,01-<0,01)	29 (74,36%)	0,1 (<0,01-0,81)	<0,001
Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae <i>Sphingobium</i>	31 (79,49%)	0,49 (0,04-1,51)	31 (79,49%)	0,49 (0,03-1,51)	0,95
Bacteria Actinobacteria Actinobacteria Corynebacteriales Corynebacteriaceae <i>Corynebacterium1</i>	0 (0%)	<0,01 (<0,01-<0,01)	15 (38,46%)	<0,01 (<0,01-0,06)	<0,001

Resultados

Continuación Tabla 10

Mínima asignación taxonómica (MAT)	Greengenes XX		SILVA 132		P-value†
	n (%)	Mediana (IQR)	n (%)	Median (IQR)	
Bacteria Proteobacteria Betaproteobacteria Burkholderiales Oxalobacteraceae	35 (89,74%)	1,42 (0,18-2,93)	0 (0%)	<0,01 (<0,01-<0,01)	<0,001
Bacteria Proteobacteria Gammaproteobacteria Betaproteobacteriales Burkholderiaceae <i>Herbaspirillum</i>	0 (0%)	<0,01 (<0,01-<0,01)	35 (89,74%)	1,43 (0,18-2,95)	<0,001
Bacteria Firmicutes Bacilli Lactobacillales Lactobacillaceae <i>Lactobacillus</i>	32 (82,05%)	0,48 (0,07-1,32)	32 (82,05%)	0,54 (0,07-1,32)	0,89
Bacteria Actinobacteria Actinobacteria Actinomycetales Corynebacteriaceae <i>Corynebacterium</i>	23 (58,97%)	0,03 (<0,01-0,34)	0 (0%)	<0,01 (<0,01-<0,01)	<0,001
Bacteria Actinobacteria Actinobacteria Actinomycetales Micrococcaceae <i>Rothia</i>	15 (38,46%)	<0,01 (<0,01-0,06)	0 (0%)	<0,01 (<0,01-<0,01)	<0,001
Minor_genera	39 (100%)	11,16 (6-17,44)	39 (100%)	12,63 (6,48-21,06)	0,38

n (%): Número de muestras que detectaron la MAT (frecuencia relativa de detección).

† Prueba Wilconxon de rangos con la corrección de Bonferroni.

Resultados

Tabla 11: Taxonomía de los 20 géneros mayoritarios ordenados por la abundancia relativa de los filios (grises) y los géneros.

	Greengenes 13.2		Silva 132		P-valort†
	n (%)	Mediana (IQR)	n (%)	Mediana (IQR)	
Proteobacteria	51 (100%)	31,63 (10,44-62,13)	51 (100%)	31,57 (10,51-62,17)	0,93
<i>Pseudomonas</i>	50 (98,04%)	25,25 (5,85-48,17)	50 (98,04%)	25,26 (6,12-48,22)	0,89
<i>Acinetobacter</i>	46 (90,2%)	1,1 (0,3-2,82)	46 (90,2%)	1,1 (0,3-2,82)	0,93
<i>Escherichia</i>	0 (0%)	<0,01 (<0,01-<0,01)	35 (68,63%)	0,06 (<0,01-0,6)	<0,001
<i>Herbaspirillum</i>	0 (0%)	<0,01 (<0,01-<0,01)	47 (92,16%)	1,18 (0,22-2,88)	<0,001
<i>Sphingobium</i>	41 (80,39%)	0,23 (0,03-1,39)	41 (80,39%)	0,23 (0,03-1,39)	0,95
<i>Haemophilus</i>	28 (54,9%)	0,09 (<0,01-0,62)	28 (54,9%)	0,06 (<0,01-0,59)	0,9
Epsilonbacteraeota	36 (70,59%)	0,62 (<0,01-17,17)	39 (76,47%)	0,8 (0,04-17,18)	0,54
<i>Helicobacter</i>	36 (70,59%)	0,62 (<0,01-17,17)	36 (70,59%)	0,62 (<0,01-17,18)	0,93
<i>Clostridium</i>	9 (17,65%)	<0,01 (<0,01-<0,01)	20 (39,22%)	<0,01 (<0,01-0,13)	0,017
Firmicutes	51 (100%)	9,11 (4,5-16,01)	51 (100%)	9,15 (4,5-15,75)	0,99
<i>Staphylococcus</i>	38 (74,51%)	0,2 (0,01-0,91)	38 (74,51%)	0,2 (0,01-0,91)	0,92
<i>Streptococcus</i>	45 (88,24%)	0,32 (0,06-1,03)	45 (88,24%)	0,32 (0,06-1,03)	0,92
<i>Blautia</i>	22 (43,14%)	<0,01 (<0,01-0,19)	24 (47,06%)	<0,01 (<0,01-0,27)	0,66
<i>Lactobacillus</i>	40 (78,43%)	0,48 (0,04-1,08)	40 (78,43%)	0,54 (0,04-1,08)	0,91
Bacteroidetes	51 (100%)	7,61 (2,1-16,92)	51 (100%)	7,65 (2,11-16,94)	0,95
<i>Bacteroides</i>	40 (78,43%)	0,5 (0,07-1,84)	40 (78,43%)	0,5 (0,07-1,85)	0,9
<i>Prevotella</i>	36 (70,59%)	0,12 (<0,01-0,75)	34 (66,67%)	0,11 (<0,01-0,75)	0,88
Filos minoritarios	50 (98,04%)	1,86 (0,75-5,43)	50 (98,04%)	1,86 (0,79-5,46)	0,93
<i>Rothia</i>	20 (39,22%)	<0,01 (<0,01-0,06)	16 (31,37%)	<0,01 (<0,01-0,05)	0,5
<i>Corynebacterium</i>	30 (58,82%)	0,03 (<0,01-0,34)	26 (50,98%)	0,01 (<0,01-0,18)	0,28
<i>Anaeroplasma</i>	19 (37,25%)	<0,01 (<0,01-0,07)	19 (37,25%)	<0,01 (<0,01-0,07)	0,97
<i>Akkermansia</i>	29 (56,86%)	0,1 (<0,01-0,56)	30 (58,82%)	0,13 (<0,01-0,56)	0,8
Minor_genera	51 (100%)	6,76 (2,62-10,11)	51 (100%)	9,71 (3,49-12,76)	0,096
Filo no clasificado	48 (94,12%)	4,07 (0,35-15,88)	48 (94,12%)	4,05 (0,33-13)	0,86
Género no clasificado	51 (100%)	22,11 (12,04-34,55)	51 (100%)	16,19 (6,3-25,33)	0,045

*n (%): Número de muestras que detectaron el filo o género (frecuencia relativa de detección).

† Prueba Wilcoxon de rangos con la corrección de Bonferroni.

5.5 Descripción taxonómica de los proyectos estudiados

Se seleccionaron los 5 proyectos descritos en el apartado de material y métodos. Las patologías aceptadas para este estudio son: muestras sin patología aparente procedentes de pacientes con dispepsia funcional, gastritis (crónica o atrófica), úlcera, metaplasia y cáncer.

Solo se seleccionaron aquellas muestras que presentaban un número de al menos 1000 secuencias pareadas y menos de un 25% de indeterminaciones.

Finalmente, se analizó la microbiota gástrica de un total de 631 muestras de las cuales solo 331 cumplieron los criterios de inclusión.

De las 331 muestras incluidas, el 39.68 % de las muestras corresponden a mujeres y la media de edad fue de 46.44 años. Las principales características de los pacientes cuyas muestras han sido incluidas y su diagnóstico se recogen en la tabla 12:

Tabla 12: Datos de diagnóstico de la infección por *H. pylori* y patología asociada de las muestras que cumplieron los criterios de inclusión de todos los proyectos.

Total de proyectos	39.68 % Mujeres
H.p.-	64 (19,34%)
Dispepsia	43 (12,99%)
Gastritis	15 (4,53%)
Cáncer	6 (1,81%)
H.p.+	267 (80,66%)
Dispepsia	18 (5,44%)
Gastritis	87 (26,28%)
Úlcera	4 (1,21%)
Metaplasia	29 (8,76%)
Cáncer	129 (38,97%)
Total general	331 (100%)

5.5.1 Proyecto Llorca *et al.* (2017)

Se realizó la comparación de los datos publicados por los estudios de referencia con los obtenidos con nuestro protocolo de trabajo.

Al realizar la asignación taxonómica se obtuvo una abundancia relativa muy parecida al estudio de referencia. Se obtuvieron 8 de sus 10 géneros mayoritarios entre los 10 géneros mayoritarios al realizar nuestra metodología. De los dos géneros que no se detectaron entre los 10 primeros (*Rothia* y *Blautia*), *Rothia* se encontró entre los 15 mayoritarios y *Blautia* no se encontró entre los géneros mayoritarios (Fig 6).

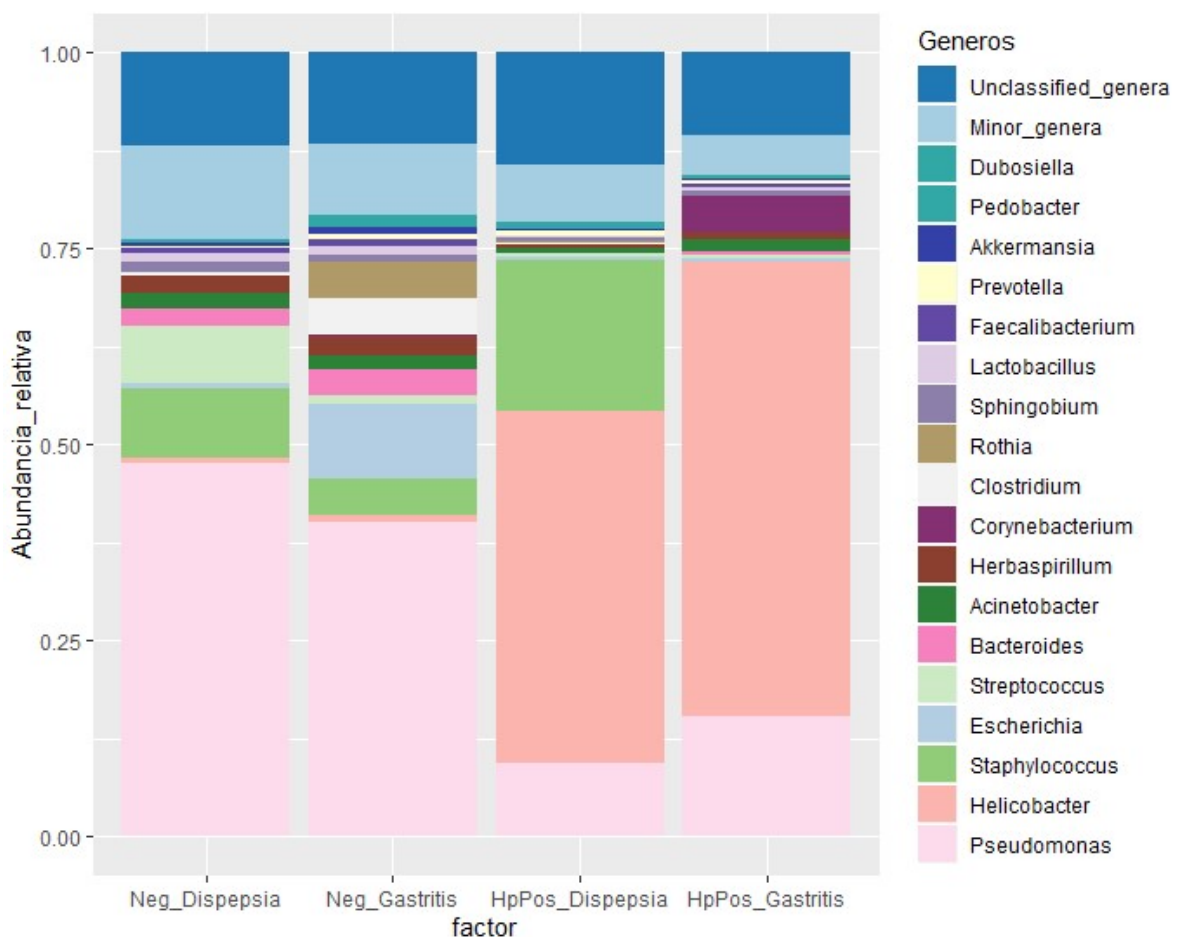


Figura 6: Abundancia relativa y composición bacteriana a nivel de género en función del estado de infección por *H. pylori* (negativo= Neg y positivo =HpPos) y las diferentes patologías gástricas asociadas de las muestras del proyecto Llorca *et al.* (2017).

Resultados

5.5.2 Proyecto PRJNA313391

Al realizar la asignación taxonómica de las secuencias obtenidas de las muestras del proyecto PRJNA313391 obtuvimos una relación de abundancias relativas muy parecida a su estudio. Se obtuvieron sus 10 géneros mayoritarios incluidos entre los 11 géneros mayoritarios de nuestro análisis. Dentro de nuestros 10 géneros mayoritarios se detectó el género *Novosphingobium* que no fue detectado por ellos (Fig 7). Entre los grupos de pacientes negativos para infección por *H. pylori*, el grupo de pacientes con cáncer solo está compuesto por un individuo por lo que la abundancia relativa de los géneros detectados puede no ser representativa de su grupo.

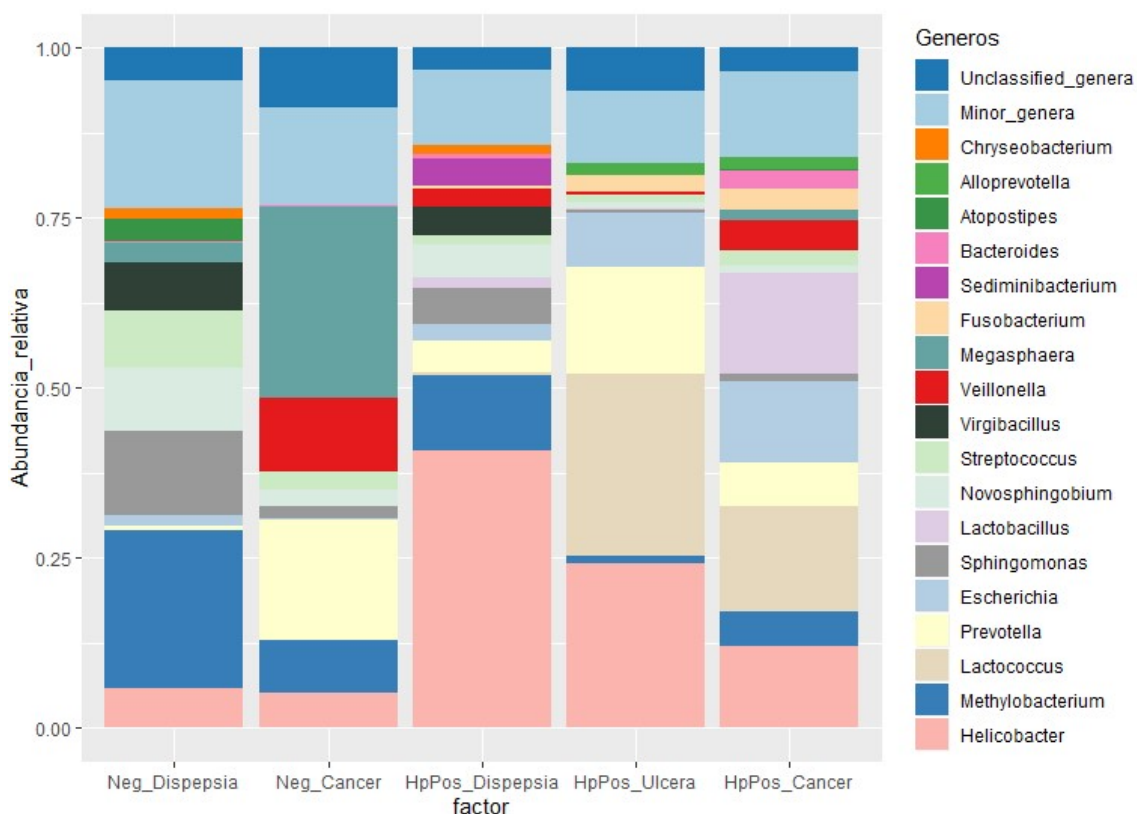


Figura 7: Abundancia relativa y composición bacteriana a nivel de género en función del estado de infección por *H. pylori* (negativo= Neg y positivo =HpPos) y las diferentes patologías gástricas asociadas de las muestras procedentes del proyecto PRJNA313391.

5.5.3 Proyecto PRJEB22107

En el estudio original del proyecto PRJEB22107 no se realizó una discusión sobre la abundancia relativa a nivel de géneros, centrándose especialmente en el nivel taxonómico de filo. Se discute sobre perfiles microbianos muy parecidos al de los estudios de Andersson *et al.*, (2008), Khosravi *et al.* (2014), Ianiro *et al.* (2015) y Llorca *et al.* (2017) haciendo hincapié en la gran abundancia de *H. pylori* entre los pacientes positivos y la importancia de las bacterias pertenecientes al género *Streptococcus* en los pacientes *H. pylori* negativos (Klymiuk *et al.*, 2017) coincidiendo los resultados que obtuvimos (Fig 8) .

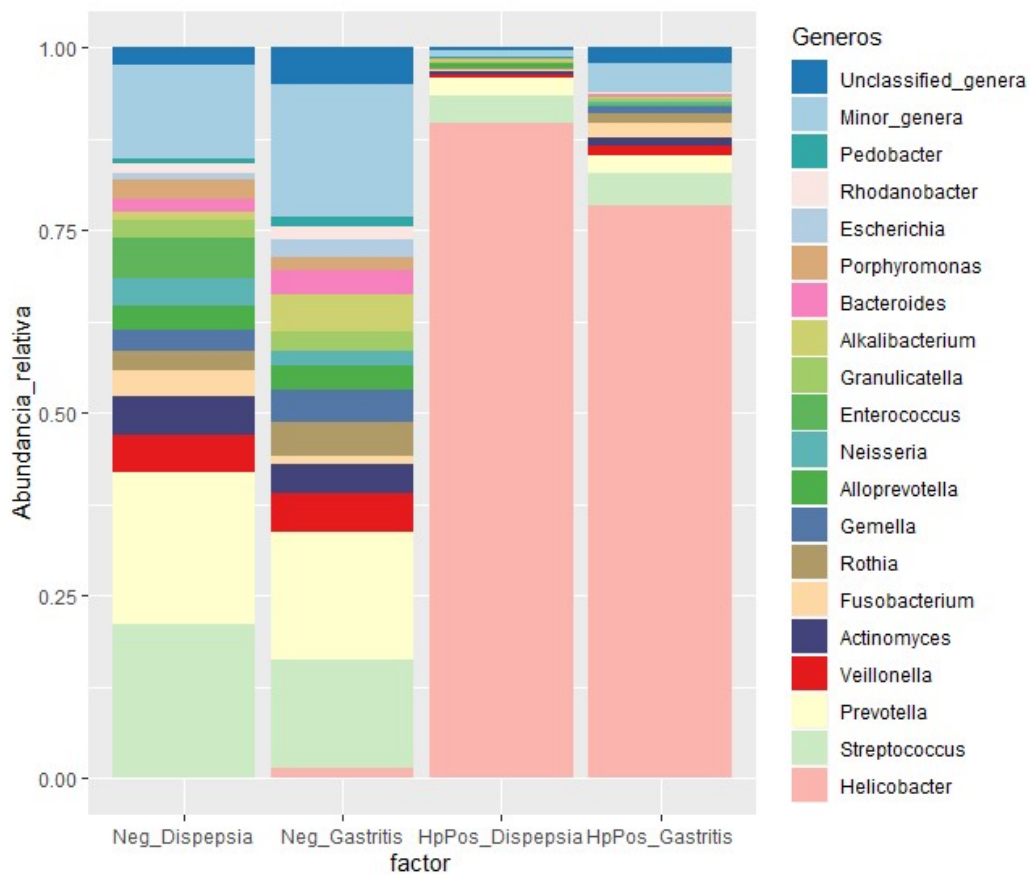


Figura 8: Abundancia relativa y composición bacteriana a nivel de género en función del estado de infección por *H. pylori* (negativo= Neg y positivo =HpPos) y las diferentes patologías gástricas asociadas de las muestras procedentes del proyecto PRJEB22107.

Resultados

5.5.4 Proyecto PRJEB21104

En el estudio de Parson *et al.* (2017) realizaron la discusión con la asignación taxonómica a nivel de familia. En este estudio detectaron una gran abundancia relativa de bacterias procedentes de las familias *Helicobacteraceae*, *Fusobacteraceae*, *Neisetiaceae*, *Prevotellaceae*, *Veillonelaceae* y *Streptococcaceae* entre otros.

Con nuestra metodología pudimos detectar muchos de los géneros que componen las citadas familias entre los 20 mayoritarios (Fig. 9):

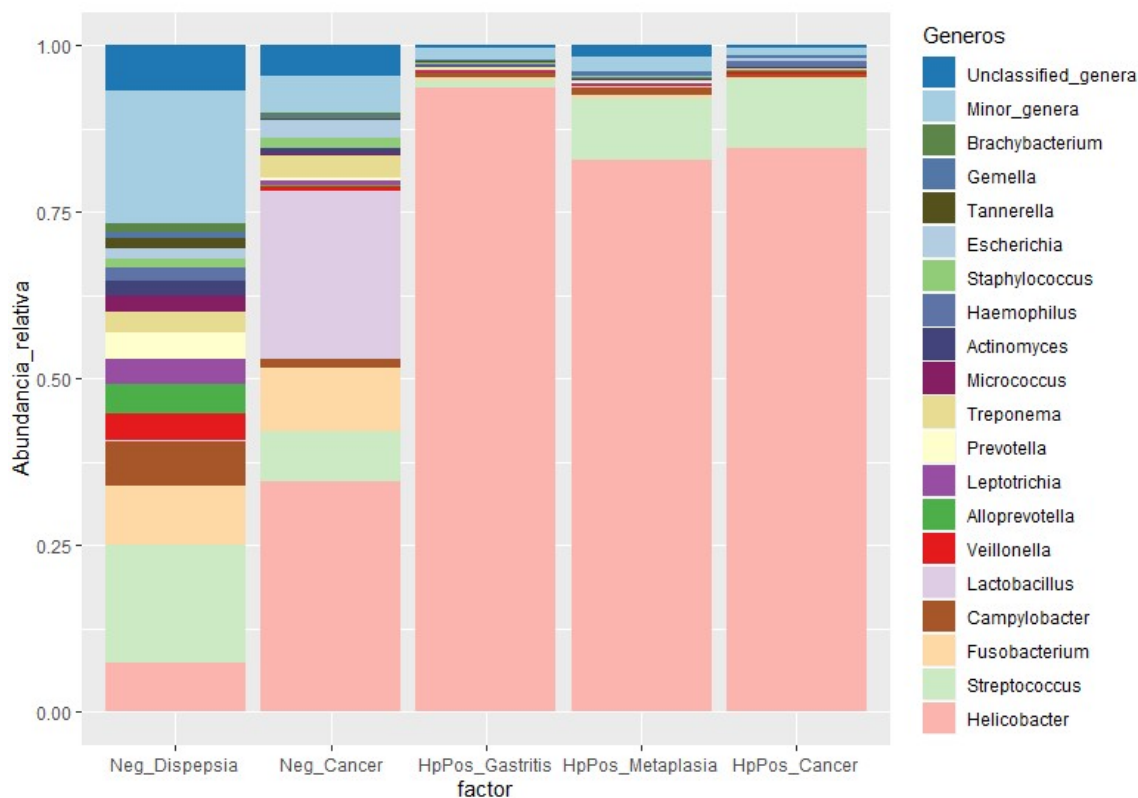


Figura 9: Abundancia relativa y composición bacteriana a nivel de género en función del estado de infección por *H. pylori* (negativo= Neg y positivo =HpPos) y las diferentes patologías gástricas asociadas de las muestras procedentes del proyecto PRJEB21104.

5.5.5 Proyecto PRJNA310127

Las secuencias almacenadas en el proyecto PRJNA310127 han sido analizadas en diferentes estudios como el de Yu *et al.* (2017) y Ferreira *et al.* (2018). En estos estudios caracterizan la microbiota de sus muestras por la gran abundancia relativa de las bacterias procedentes de los géneros *Helicobacter*, *Lactobacillus*, *Streptococcus*, *Neisseria*, *Prevotella*, *Haemophilus* y bacterias de la familia *Enterobacteriaceae*. En nuestro estudio se detectaron dichos géneros entre nuestros géneros mayoritarios (Fig. 10).

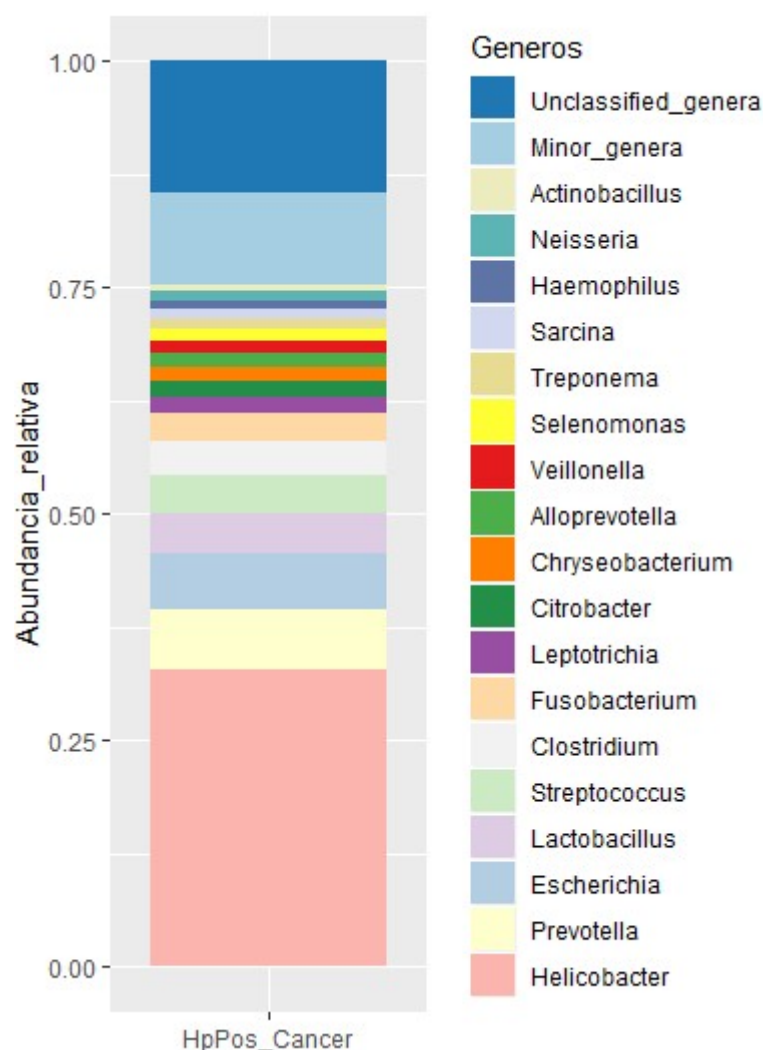


Figura 10: Abundancia relativa y composición bacteriana a nivel de género de las muestras gástricas del proyecto PRJEB21104. En este caso el grupo corresponde a pacientes con cáncer e infección por *H. pylori*.

Resultados

5.5.6 Proyecto PRJNA313391

Este grupo de secuencias no tiene un artículo en el que se reflejen las pruebas diagnósticas de *H. pylori* y no disponemos de unos resultados preliminares de los géneros mayoritarios. Estas muestras han sido ya utilizadas por Leeuw y Duval. (2020) (de Leeuw y Duval, 2020) quienes detectaron los géneros *Ochrobactrum*, *Pannonibacter* y *Sphingomonas* entre otros. Este estudio es el que más dudas genera ya que no ha sido publicado en un artículo por el equipo que lo secuenció, es el que más géneros exclusivos presenta entre los géneros mayoritarios (12 de los 20 géneros mayoritarios) y el perfil microbiológico es muy diferente a los perfiles encontrados en los otros estudios (Fig. 11).

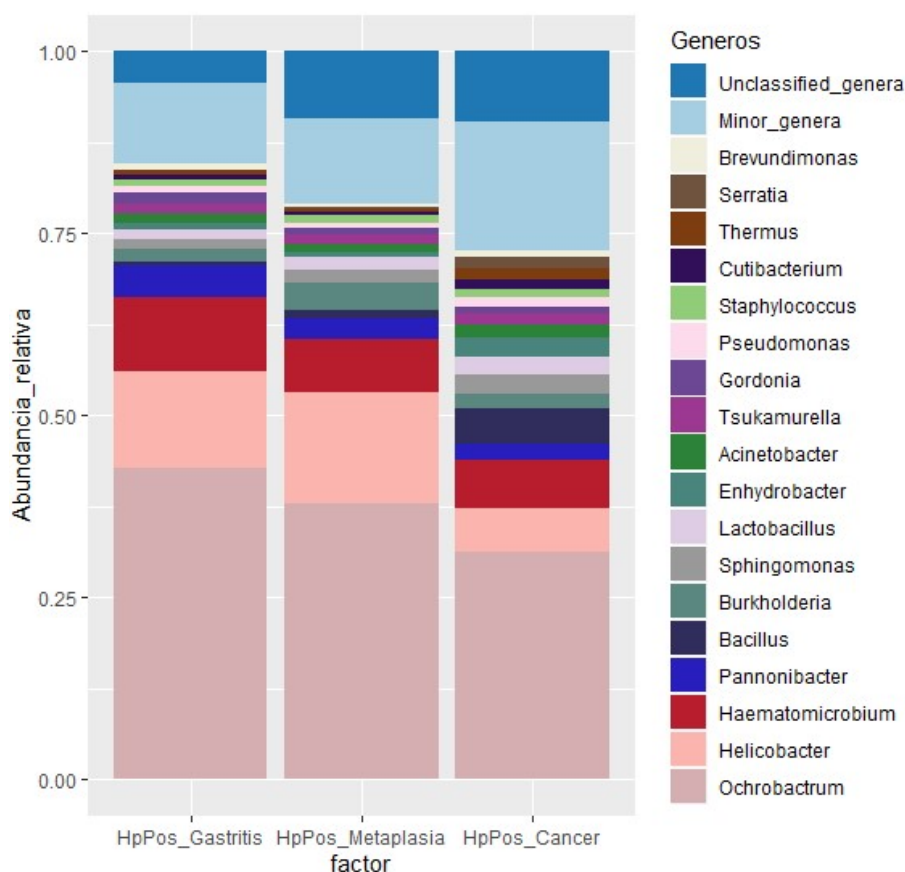


Figura 11: Abundancia relativa y composición bacteriana a nivel de género de las muestras gástricas del proyecto PRJEB21104. En este caso el grupo corresponde a pacientes con positivos para infección por *H. pylori* (HpPos) y diferentes patologías gástricas.

Para estudiar la idoneidad del proyecto se comparará con las muestras de cáncer gástrico del mismo país del estudio de Yu *et al.* (2017).

5.5.7 Estudio de idoneidad del proyecto PRJNA313391

Para determinar la viabilidad de las muestras del proyecto PRJNA313391 se realizó un estudio comparativo de muestras procedentes del mismo origen geográfico del proyecto de Yu *et al.* (2017), proyecto PRJNA310127.

Se analizaron un total de 133 muestras (108 procedentes del proyecto PRJNA313391 y 25 del estudio de Yu *et al.* (2017)) procedentes de pacientes diagnosticados como *Helicobacter* positivo con residencia en China. En total se analizaron un total de 7.090.635 secuencias filtradas de alta calidad, con un rango desde 12.725 a 171.320 secuencias por muestra (mediana (IQR)= 55.528 (44.581,25-77.890) y 25.317 (18.735-30.893) en las muestras de PRJNA313391 y Yu *et al.*) obteniéndose un total de 1554 ASVs diferentes.

La mediana de los valores de diversidad de Shannon fue de 2,65 y 1,34 mientras que la mediana de los valores de diversidad de Simpson fue de 0,7 y 0.26 respectivamente. En general, hubo diferencias significativas en función del estudio analizado en ambos índices de diversidad ($p < 0,001$ para ambos índices).(Fig. 12).

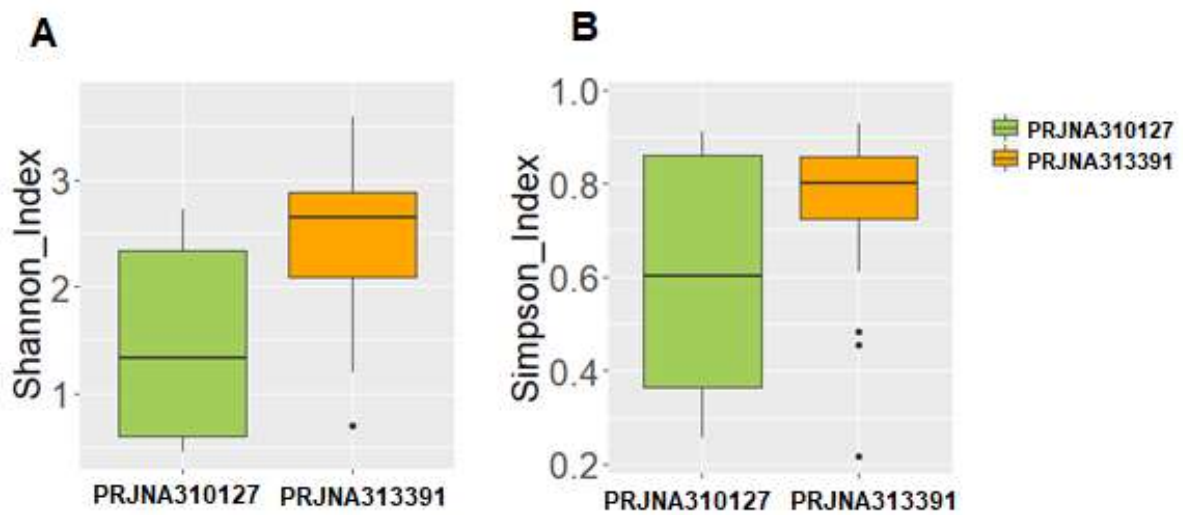


Figura 12: Índices de diversidad de Shannon (A) y Simpson (B) en muestras de los proyectos PRJNA310127 (verde) y PRJNA313391 (naranja).

Resultados

El análisis de la diversidad beta mantuvo esa misma tendencia. La diversidad beta calculada acorde a la abundancia relativa (mediante el análisis de matrices de distancia de Bray-Curtis) (Fig. 13-A) o acorde a la presencia y ausencia de las secuencias ASVs (mediante el análisis de las matrices de distancia de Jaccard binarias) (Fig. 13-B), indicaron una clara separación de los perfiles bacterianos dependiendo del tipo del estudio ($p < 0,001$ y $p < 0,001$, respectivamente; PERMANOVA).

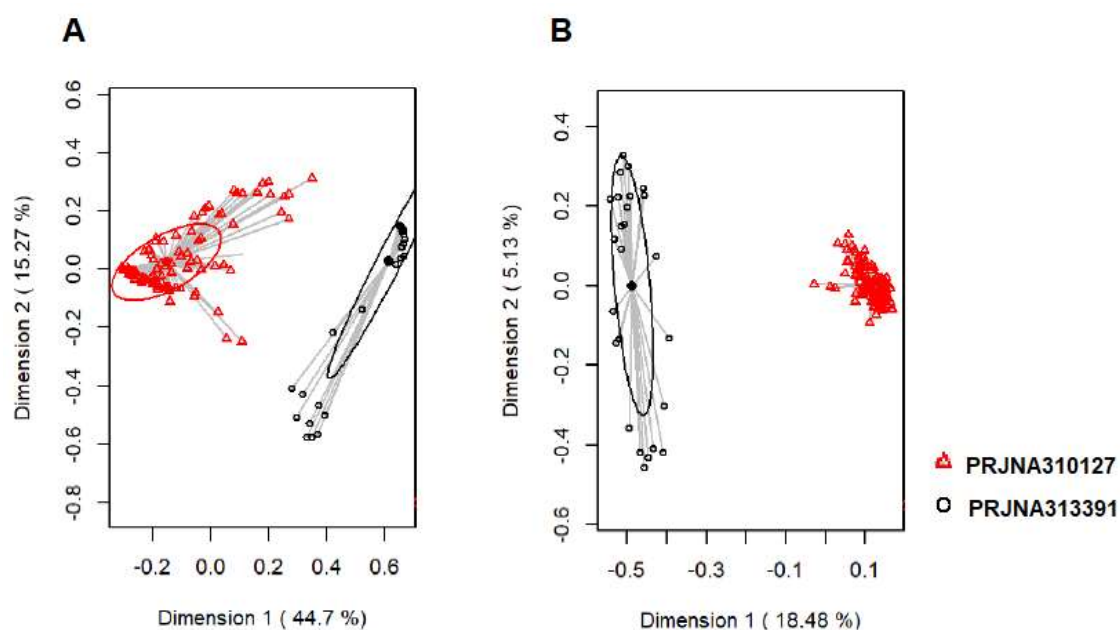


Figura 13: PCoAs de asignaciones taxonómicas (Bray-Curtis) (A) y la presencia/ausencia de dichas asignaciones (matriz de distancias de Jaccard binaria) (B) en el proyecto PRJNA310127 (triángulo, rojo) y PRJNA313391 (círculo, negro) El valor en cada eje representa el porcentaje de variación explicado por ese eje.

Las diferencias en los perfiles bacterianos se encontraron a diferentes niveles taxonómicos:

Se identificaron un total de 39 fillos diferentes siendo los fillos *Proteobacteria*, *Epsilonbacteraeota*, *Actinobacteria* y *Firmicutes* los más abundantes. Hubo un efecto significativo del tipo de muestra sobre los fillos *Proteobacteria*, *Epsilonbacteraeota*, *Actinobacteria* y el grupo de fillos minoritarios ($p < 0,001$) (Tabla 13).

Entre los géneros mayoritarios se aprecian claros patrones en función del estudio. El principal género diferenciador fue el género mayoritario detectado en el estudio PRJNA313391, *Ochrobactrum*. Se detectó en todas las muestras y presentó una abundancia relativa con una mediana (IQR) de 36,34 (25,41-44,70), mientras que no fue detectado en las muestras del estudio de Yu *et al.* El género *Helicobacter* fue detectado en ambos estudios con una abundancia relativa significativamente mayor en las muestras del estudio de Yu *et al.* que en las muestras del estudio PRJNA313391 (54,83 (0,18-76,53) y 0,78 (<0,01-11,82) respectivamente) ($p < 0.001$). Otros géneros mayoritarios también fueron detectados solo en las muestras del proyecto PRJNA313391: *Haematomicrobium*, *Pannonibacter*, *Bacillus*, *Burkholderia* o *Sphingomonas*. Por el contrario, todos los géneros mayoritarios detectados en las muestras del proyecto PRJNA310127 fueron detectados en las muestras del estudio PRJNA313391 con diferentes abundancias relativas (Tabla 13).

Resultados

Tabla 13: Abundancia relativa (%), expresada como la mediana y los rangos intercuartílicos (IQR), de los 20 géneros y 4 filos mayoritarios en función del proyecto analizado.

Filo/Género	PRJNA310127		PRJNA313391		P-valort
	n(%)#	Mediana (IQR)	n(%)#	Median (IQR)	
<i>Proteobacteria</i>	22 (88%)	0,80 (0,12-5,39)	108 (100%)	61,63 (48,62-68,63)	<0,001
<i>Ochrobactrum</i>	0 (0%)	<0,01 (<0,01-<0,01)	108 (100%)	36,34 (25,41-44,7)	<0,001
<i>Pannonibacter</i>	0 (0%)	<0,01 (<0,01-<0,01)	108 (100%)	2,63 (1,94-3,98)	<0,001
<i>Burkholderia</i>	0 (0%)	<0,01 (<0,01-<0,01)	79 (73,15%)	2,13 (<0,01-3,78)	<0,001
<i>Sphingomonas</i>	0 (0%)	<0,01 (<0,01-<0,01)	108 (100%)	1,90 (0,93-2,89)	<0,001
<i>Acinetobacter</i>	12 (48%)	<0,01 (<0,01-0,13)	108 (100%)	1,10 (0,69-1,61)	<0,001
<i>Enhydrobacter</i>	1 (4%)	<0,01 (<0,01-<0,01)	107 (99,07%)	0,36 (0,18-1,15)	<0,001
<i>Pseudomonas</i>	11 (44%)	<0,01 (<0,01-0,15)	108 (100%)	0,69 (0,38-1,19)	<0,001
<i>Epsilonbacteraeota</i>	24 (96%)	55,03 (1,22-76,53)	83 (76,85%)	0,78 (0,01-11,84)	<0,001
<i>Helicobacter</i>	21 (84%)	54,83 (0,18-76,53)	78 (72,22%)	0,78 (<0,01-11,82)	<0,001
<i>Actinobacteria</i>	9 (36%)	<0,01 (<0,01-0,46)	108 (100%)	14,64 (11,38-17,16)	<0,001
<i>Haematomicrobium</i>	0 (0%)	<0,01 (<0,01-<0,01)	108 (100%)	7,45 (5,85-9,54)	<0,001
<i>Tsukamurella</i>	0 (0%)	<0,01 (<0,01-<0,01)	108 (100%)	1,38 (1,02-1,79)	<0,001
<i>Gordonia</i>	0 (0%)	<0,01 (<0,01-<0,01)	108 (100%)	1,06 (0,82-1,47)	<0,001
<i>Cutibacterium</i>	0 (0%)	<0,01 (<0,01-<0,01)	108 (100%)	0,83 (0,36-1,23)	<0,001
<i>Firmicutes</i>	22 (88%)	4,41 (0,18-12,93)	108 (100%)	5,53 (2,92-9,51)	0,23
<i>Bacillus</i>	0 (0%)	<0,01 (<0,01-<0,01)	86 (79,63%)	0,15 (0,03-3,01)	<0,001
<i>Lactobacillus</i>	4 (16%)	<0,01 (<0,01-<0,01)	108 (100%)	1,02 (0,53-1,68)	<0,001
<i>Staphylococcus</i>	0 (0%)	<0,01 (<0,01-<0,01)	108 (100%)	0,64 (0,38-1,42)	<0,001
<i>Prevotella</i>	18 (72%)	1,91 (<0,01-6,14)	79 (73,15%)	0,03 (<0,01-0,12)	<0,001
<i>Chryseobacterium</i>	6 (24%)	<0,01 (<0,01-<0,01)	78 (72,22%)	0,04 (<0,01-0,10)	0,003
<i>Thermus</i>	0 (0%)	<0,01 (<0,01-<0,01)	101 (93,52%)	0,20 (0,06-0,59)	<0,001
Filos minoritarios	23 (92%)	8,82 (0,9-42,48)	108 (100%)	2,32 (1,31-4,32)	0,012
Géneros minoritarios	25 (100%)	15,97 (0,72-57,42)	108 (100%)	14,15 (9,85-20,86)	0,55
Filo no clasificado	25 (100%)	17,36 (12,34-21,32)	107 (99,07%)	4,04 (1,99-7,32)	<0,001
Género no clasificado	25 (100%)	19,89 (14,47-22,24)	108 (100%)	6,62 (3,72-11,48)	<0,001

n (%): Número de muestras que detectaron el filo o género (frecuencia relativa de detección).

† Prueba Wilcoxon de rangos con la corrección de Bonferroni.

Para determinar si el estudio PRJNA313391 presentaba un perfil microbiano con contaminantes se realizó un estudio de correlaciones dentro de sus 20 géneros mayoritarios reordenándolos en función de su perfil de correlaciones mediante el método hclust (Fig 14).

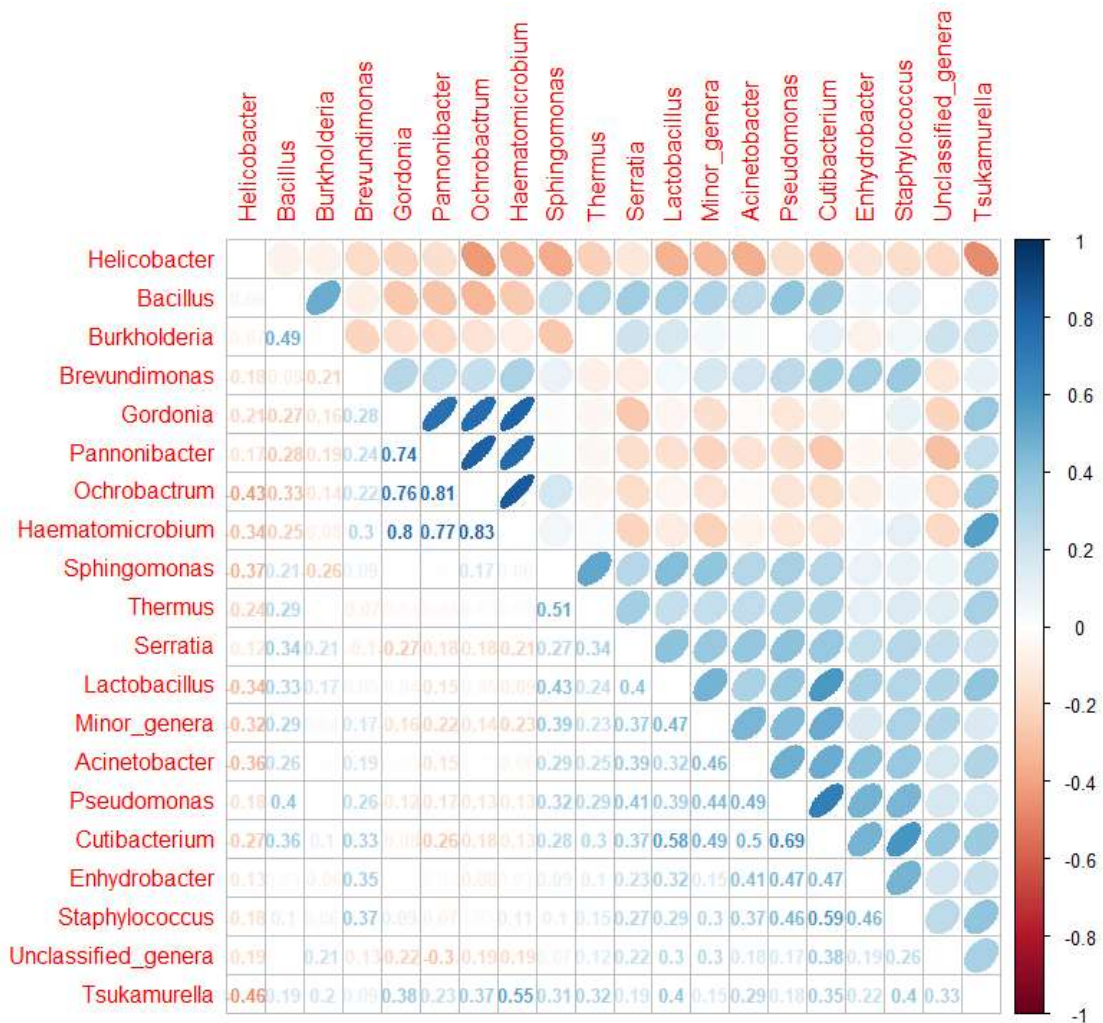


Figura 14: Gráfico de correlaciones mixto de los 20 géneros con mayor abundancia relativa media del proyecto PRJNA313391. Los géneros se han reordenado por similitud de correlaciones mediante el método hclust. Las correlaciones positivas se representan en color azul y las correlaciones negativas en color rojo.

Al realizar el estudio de correlaciones se aprecian 3 grandes grupos de géneros siendo el grupo compuesto por los géneros *Gordonia*, *Pannonibacter*, *Ochrobactrum* y *Haematomicrobium* el grupo que presenta mayores correlaciones entre sí.

Resultados

Todos estos géneros, aunque pueden ser patógenos oportunistas, se suelen encontrar en muestras de suelo y agua (Alvarado Capó *et al.*, 2007; Imirzalioglu, Hain, Hossain, Chakraborty, & Domann, 2010; Muñoz-Peña, Ocaña-Cano, Amores-Antequera, & Cantudo-Muñoz, 2016; M. Wang *et al.*, 2017). Al presentarse en correlaciones positivas significativas y encontrarse en todas las muestras del estudio, podría tratarse de contaminantes de algún reactivo.

En base a estos resultados no podemos concluir que el estudio PRJNA313391 no haya sufrido una fuente de contaminación. Debido a que estos 4 géneros presentan una abundancia relativa media conjunta de 47.73%, siendo el género *Ochrobactrum* el género mayoritario del estudio, no consideramos viable eliminar las posibles secuencias contaminantes y por tanto será excluido de los siguientes estudios.

5.6 Influencia de *Helicobacter pylori*

Para determinar el efecto de la infección por *H. pylori* en la microbiota gástrica se realizó un estudio comparativo de muestras de biopsia gástrica de los proyectos Llorca *et al.*, (2017), PRJEB21497, PRJEB22107, PRJEB21104 y PRJNA310127.

Se analizaron un total de 223 muestras (64 muestras de biopsias procedentes de pacientes diagnosticados como *Helicobacter* negativo y 159 provenientes de pacientes diagnosticados como *Helicobacter* positivo) procedentes de proyectos de diversas áreas geográficas. En total se analizaron un total de 8.062.415 secuencias filtradas de alta calidad, con un rango desde 1.098 a 343.161 secuencias por muestra (mediana (IQR)= 6.952,5 (2.776-19.033,25) y 23.864 (14.352,5-45.499) en las muestras de pacientes *Helicobacter* negativos y positivos respectivamente), obteniéndose un total de 732 ASVs diferentes.

5.6.1 Diversidad alfa.

La mediana de los valores de diversidad de Shannon fue de 2,35 en las muestras provenientes de pacientes negativos y 0,90 en las biopsias gástricas de pacientes positivos. Se mantuvo esa tendencia cuando se analizó el índice de diversidad Simpson que fue de 0,81 y 0,39, respectivamente. En general, hubo un efecto significativo del diagnóstico de infección por *H. pylori* en ambos índices de diversidad ($p < 0,001$ para ambos índices) (Fig. 15).

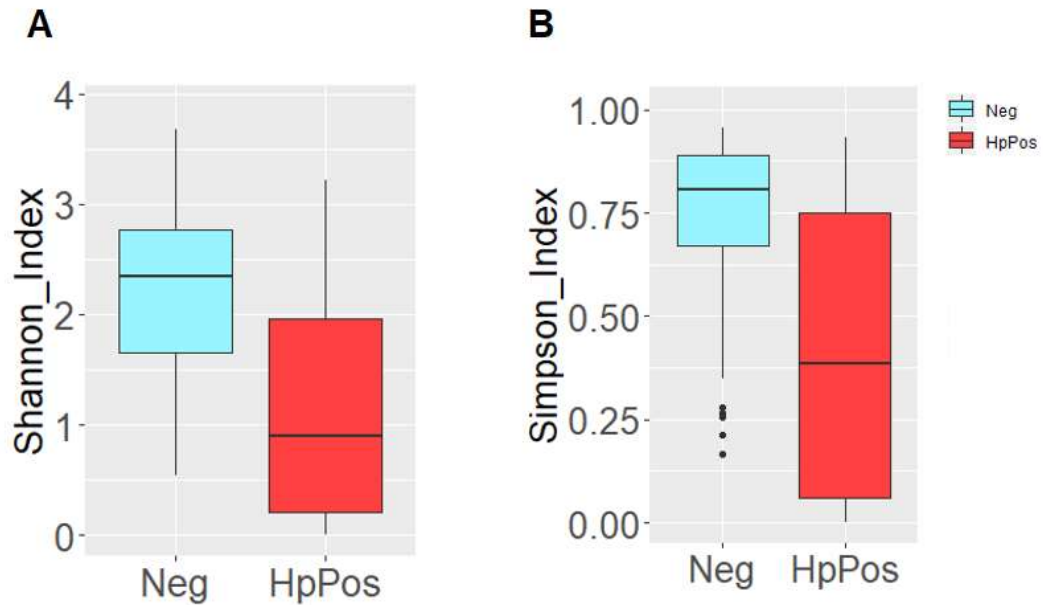


Figura 15: Mediana de los valores de diversidad de Shannon (A) y Simpson (B) en las muestras provenientes de pacientes diagnosticados sin infección de *H. pylori* (Neg, azul) y positivos a infección por *H. pylori* (HpPos, rojo).

5.6.2 Diversidad beta

La diversidad beta calculada acorde a la abundancia relativa de las secuencias ASVs (mediante el análisis de matrices de distancia de Bray-Curtis) o acorde a la presencia y ausencia de las secuencias ASVs (mediante el análisis de las matrices de distancia de Jaccard binarias), indicaron una separación significativa de los perfiles bacterianos en función de la infección por *H. pylori* ($p < 0,001$ y $p < 0,001$, respectivamente; PERMANOVA) (Fig. 16).

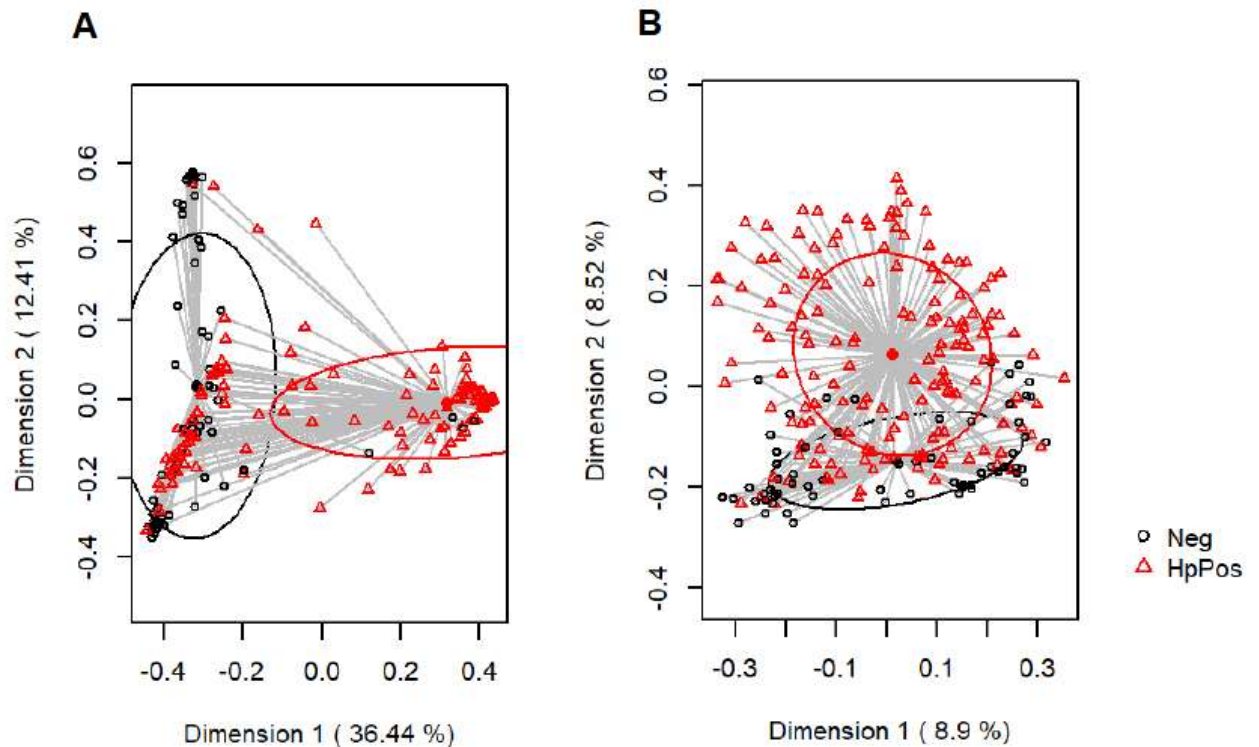


Figura 16: PCoAs de los perfiles bacterianos a nivel ASV basado en la matriz de distancia Bray-Curtis (A) y la matriz de distancia Jaccard binaria para el estudio por presencia/ausencia (B) de muestras de biopsias gástricas. El color significa el diagnóstico para la infección por *H. pylori* (Neg= negativo, círculo negro; HpPos= positivos, triángulo rojo). El valor en cada eje representa el porcentaje de variación explicado por ese eje.

5.6.3 Diferencias taxonómicas

Las diferencias en los perfiles bacterianos se encontraron a diferentes niveles taxonómicos.

Se identificaron un total de 22 filos diferentes, siendo los filos *Epsilonbacteraeota*, *Firmicutes*, *Proteobacteria* y *Bacteroidetes* los más abundantes. Hubo un efecto significativo del tipo de muestra sobre los filos *Epsilonbacteraeota*, *Firmicutes*, *Proteobacteria* y *Bacteroidetes* y el grupo de filos minoritarios ($p < 0,001$ en todas las comparaciones) (Tabla 14).

Entre los géneros mayoritarios hay una clara importancia del género *Helicobacter* en las muestras *Helicobacter* positivo. El principal género diferenciador fue el género mayoritario *Helicobacter*, detectado en 42 muestras de pacientes negativos (65,62%) y en 134 muestras de pacientes positivos (84,28%), siendo estas frecuencias de detección estadísticamente diferentes (Chi-cuadrado de

Resultados

Pearson (p-valor =0,003)). La abundancia relativa del género *Helicobacter* presentó, en general, un aumento en aquellos pacientes diagnosticados como *Helicobacter* positivos 71,49 (3,04-97,02) frente a 0,11 (<0,01-2,29) en los pacientes diagnosticados como *Helicobacter* negativos (Tabla 14).

Tabla 14: Abundancia relativa (%), expresada como la mediana y los rangos intercuartílicos (IQR), de los 20 géneros y 4 filos mayoritarios en función diagnóstico para *H. pylori*.

Filo/Género	<i>Helicobacter</i> negativo		<i>Helicobacter</i> positivos		P-valor†
	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	
<i>Epsilonbacteraeota</i>	49 (76,56%)	1,34 (0,09-8,31)	149 (93,71%)	71,61 (3,83-97,16)	<0,001
<i>Helicobacter</i>	42 (65,62%)	0,11 (<0,01-2,29)	134 (84,28%)	71,49 (3,04-97,02)	<0,001
<i>Campylobacter</i>	31 (48,44%)	<0,01 (<0,01-1,88)	64 (40,25%)	<0,01 (<0,01-0,08)	0,003
<i>Firmicutes</i>	63 (98,44%)	22,28 (10,02-38,91)	153 (96,23%)	4,41 (0,27-22,59)	<0,001
<i>Streptococcus</i>	56 (87,50%)	5,43 (0,47-17,05)	137 (86,16%)	0,33 (0,04-2,56)	<0,001
<i>Lactobacillus</i>	32 (50,00%)	<0,01 (<0,01-0,8)	60 (37,74%)	<0,01 (<0,01-0,14)	0,099
<i>Veillonella</i>	42 (65,62%)	0,33 (<0,01-3,29)	92 (57,86%)	0,03 (<0,01-0,60)	0,013
<i>Staphylococcus</i>	44 (68,75%)	0,09 (<0,01-1,11)	61 (38,36%)	<0,01 (<0,01-0,04)	<0,001
<i>Clostridium</i>	14 (21,88%)	<0,01 (<0,01-<0,01)	20 (12,58%)	<0,01 (<0,01-<0,01)	0,086
<i>Lactococcus</i>	7 (10,94%)	<0,01 (<0,01-<0,01)	21 (13,21%)	<0,01 (<0,01-<0,01)	0,57
<i>Proteobacteria</i>	64 (100%)	18,29 (9,25-60,30)	136 (85,53%)	0,73 (0,07-9,45)	<0,001
<i>Pseudomonas</i>	49 (76,56%)	0,25 (0,01-26,11)	54 (33,96%)	<0,01 (<0,01-0,02)	<0,001
<i>Escherichia</i>	41 (64,06%)	0,16 (<0,01-0,99)	71 (44,65%)	<0,01 (<0,01-0,21)	0,004
<i>Methylobacterium</i>	25 (39,06%)	<0,01 (<0,01-0,43)	40 (25,16%)	<0,01 (<0,01-<0,01)	0,028
<i>Sphingomonas</i>	20 (31,25%)	<0,01 (<0,01-0,02)	33 (20,75%)	<0,01 (<0,01-<0,01)	0,120
<i>Bacteroidetes</i>	61 (95,31%)	10,62 (2,35-18,34)	147 (92,45%)	1,19 (0,13-8,31)	<0,001
<i>Prevotella</i>	47 (73,44%)	0,74 (<0,01-4,54)	128 (80,50%)	0,34 (0,02-2,62)	0,62
<i>Alloprevotella</i>	31 (48,44%)	<0,01 (<0,01-1,7)	96 (60,38%)	0,03 (<0,01-0,31)	0,66
<i>Bacteroides</i>	31 (48,44%)	<0,01 (<0,01-1,47)	53 (33,33%)	<0,01 (<0,01-0,08)	0,002
Filos_minoritarios	62 (96,88%)	6,54 (2,30-24,69)	148 (93,08%)	0,74 (0,10-4,29)	<0,001
<i>Fusobacterium</i>	37 (57,81%)	0,68 (<0,01-4,57)	93 (58,49%)	0,05 (<0,01-0,46)	0,028
<i>Leptotrichia</i>	33 (51,56%)	0,11 (<0,01-1,3)	81 (50,94%)	<0,01 (<0,01-0,14)	0,11
<i>Treponema</i>	24 (37,50%)	<0,01 (<0,01-0,53)	50 (31,45%)	<0,01 (<0,01-0,01)	0,062
Minor_genera	64 (100,00%)	21,68 (17,16-33,74)	153 (96,23%)	2,71 (0,32-17,89)	<0,001
Filo_no_clasificado	48 (75,00%)	0,22 (0,01-1,56)	107 (67,30%)	0,35 (<0,01-10,13)	0,19
Género_no_clasificado	64 (100%)	5,94 (2,34-10,48)	144 (90,57%)	2,78 (0,12-13,96)	0,038

n (%): Número de muestras que detectaron el filo o género (frecuencia relativa de detección).

†Prueba Wilcoxon de rangos con la corrección de Bonferroni.

Dentro de los 20 géneros mayoritarios no se detectó ningún género exclusivo de ninguno de los grupos de pacientes, aunque si se vieron diferencias significativas en muchos de los géneros mayoritarios en función del diagnóstico del paciente (Tabla 14).

5.6.4 Relación de los géneros mayoritarios.

Un análisis de red de correlaciones mostró la importancia del género *Helicobacter* como modulador, en general negativo, de la microbiota gástrica en los pacientes diagnosticados con infección por *Helicobacter*. En dicha red se aprecia la posición central de *Helicobacter* y su correlación negativa con múltiples géneros, como el género *Prevotella*, y la correlación positiva de ésta última con bacterias del género *Fusobacterium* o *Veionella* (Fig. 17-A). En las muestras procedentes de pacientes negativos la modulación es mucho más sutil y el género *Helicobacter* pierde su poder modulador dejando a otros géneros, como el género *Prevotellas* o el género *Pseudomonas*, dicho papel (Fig. 17-B).

Resultados

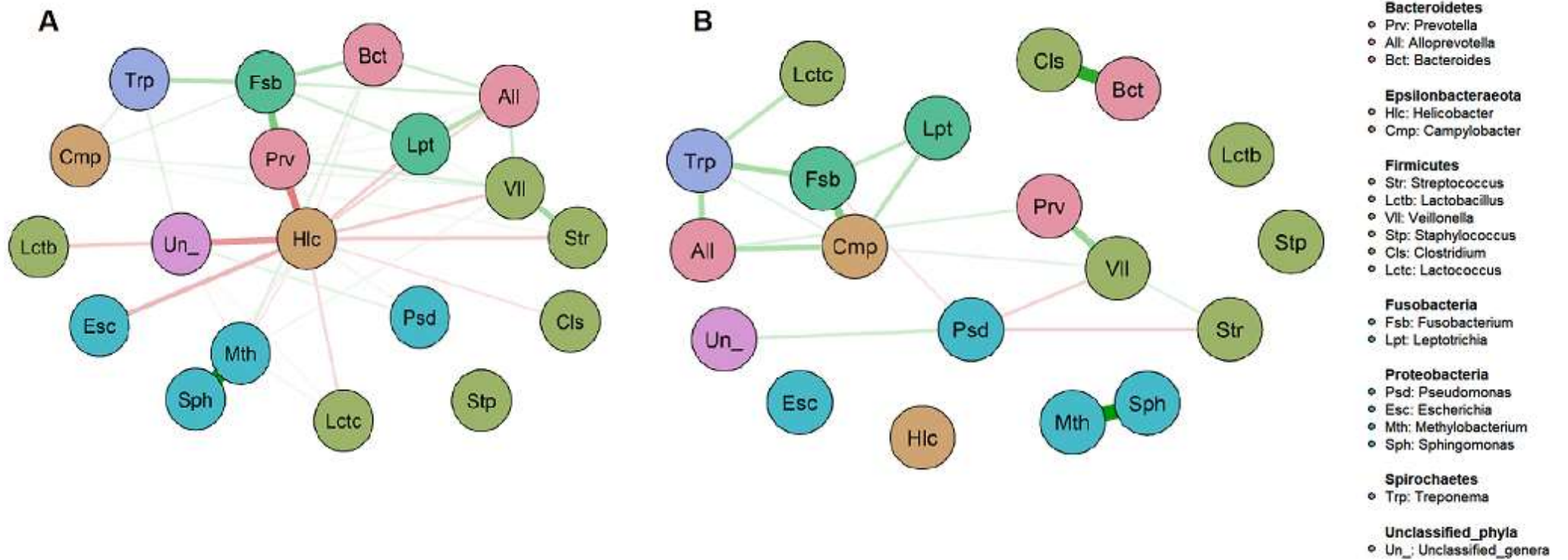


Figura 17: Estructura de relaciones estimadas basada en 223 muestras de biopsia gástrica: 159 muestras de pacientes diagnosticados positivos (A) y 64 muestras de pacientes diagnosticados negativos (B). Han sido representados los 19 géneros mayoritarios. Las líneas rojas indican correlaciones negativas y las líneas verdes correlaciones positivas.

5.6.5 Microbiota core.

Al analizar todos los géneros de todas las muestras de cada grupo de pacientes se detectaron un total de 494 géneros diferentes de los cuales solo 2 fueron categorizados como *core* junto al grupo de los géneros no clasificados (presentes en, al menos, el 80% de las muestras de alguno de los grupos con una abundancia relativa mínima del 0,01%). El género *Helicobacter* cumplió las condiciones para el grupo de los pacientes diagnosticados positivos para infección por *H. pylori* y el género *Streptococcus*, junto al grupo de los géneros no clasificados, como *core* de ambos grupos (Fig. 18).

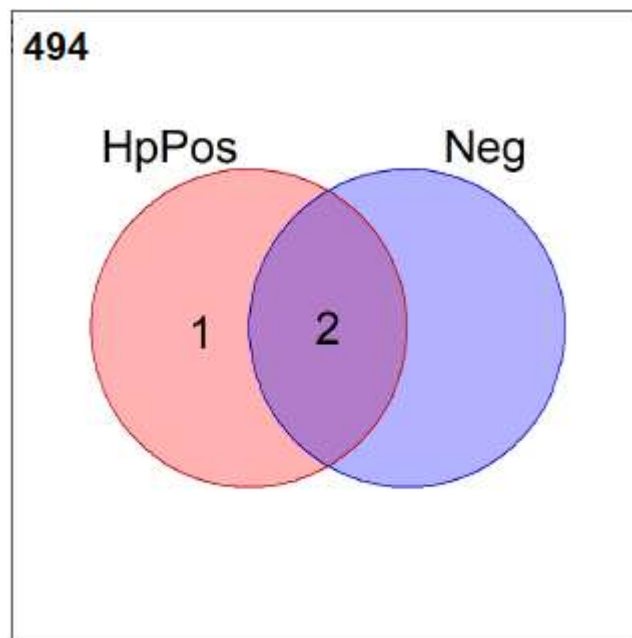


Figura 18: Diagrama de Venn de los géneros identificados en muestras de biopsia gástrica de pacientes diagnosticados positivos para infección por *H. pylori* (HpPos) y pacientes diagnosticados como negativos (Neg). El número exterior indica el número total de géneros detectados.

Resultados

5.6.6 Efecto de la sustracción de las secuencias del género *Helicobacter* de los análisis.

Para continuar con el efecto de la infección por *H. pylori* sobre la microbiota gástrica se procedió a la eliminación de las secuencias de *Helicobacter* de las muestras de biopsia gástrica. Del total de secuencias analizadas en el apartado 5.6 (8.062.415) un total de 5.942.636 secuencias fueron identificadas como provenientes de bacterias del género *Helicobacter* y fueron eliminadas. Con la sustracción de dichas secuencias de las muestras de estudio solo 185 (122 muestras provenientes de pacientes positivos y 63 de pacientes negativos) de las 223 muestras de biopsia continuaron con, al menos, 1.000 secuencias, lo que permite incluirlas para una correcta comparación.

La sustracción de las secuencias del género *Helicobacter* tuvo un efecto directo en los índices de diversidad, en la diversidad beta y las diferencias encontradas entre los filos y los géneros.

La mediana de los valores de diversidad de Shannon, especialmente en aquellas muestras provenientes de pacientes positivos, varió de 0,90 a 2,09 al eliminar el género que tenía la mayor abundancia relativa. En las muestras de los pacientes diagnosticados como negativos también varió el índice de Shannon, de 2,35 a 2,46 siendo éste un cambio menos importante.

Aún con el aumento medio del índice de diversidad de Shannon se encontraron diferencias significativas entre las muestras provenientes de pacientes positivos de aquellas provenientes de pacientes negativos ($p=0,003$). Se mantuvo esa tendencia cuando se analizó el índice de diversidad Simpson que varió de 0,39 a 0,78 en las muestras de pacientes positivos y 0,81 a 0,85 en pacientes negativos. Estas nuevas diferencias continuaron siendo significativas ($p = 0.018$). En general, hubo un efecto significativo del diagnóstico de infección por *H. pylori* en ambos índices a pesar la sustracción de las secuencias de *Helicobacter* del análisis. (Fig. 19).

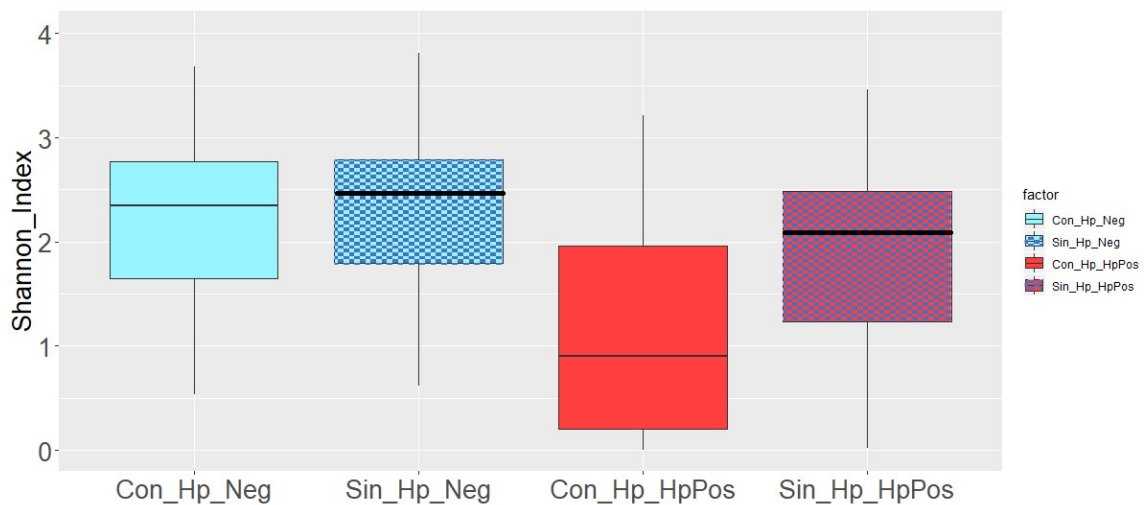


Figura 19: Diagrama de cajas mostrando la biodiversidad, medida por el índice de Shannon en muestras de biopsia gástrica de pacientes diagnosticados como negativos a infección por *H. pylori* con las secuencias de *H. pylori* (Con_Hp_Neg, azul), con las secuencias de *H. pylori* eliminadas (Sin_Hp_Neg, azul con entramado) y de pacientes diagnosticados como positivos con las secuencias (Con_Hp_HpPos, rojo) y sin las secuencias (Sin_Hp_HpPos, rojo con entramado).

Al contrario de los análisis previos, la diversidad beta (calculada mediante el análisis de matrices de distancia de Bray-Curtis) no indicó perfiles microbianos diferentes entre los pacientes positivos y los pacientes negativos ($p = 0,099$; PERMANOVA) (Fig. 20). Al analizar los perfiles microbianos en función de la presencia y ausencia mediante el análisis de las matrices de distancia de Jaccard binarias, si se encontraron diferencias significativas en los perfiles microbianos ($p = 0.019$) (Fig. 20).

Resultados

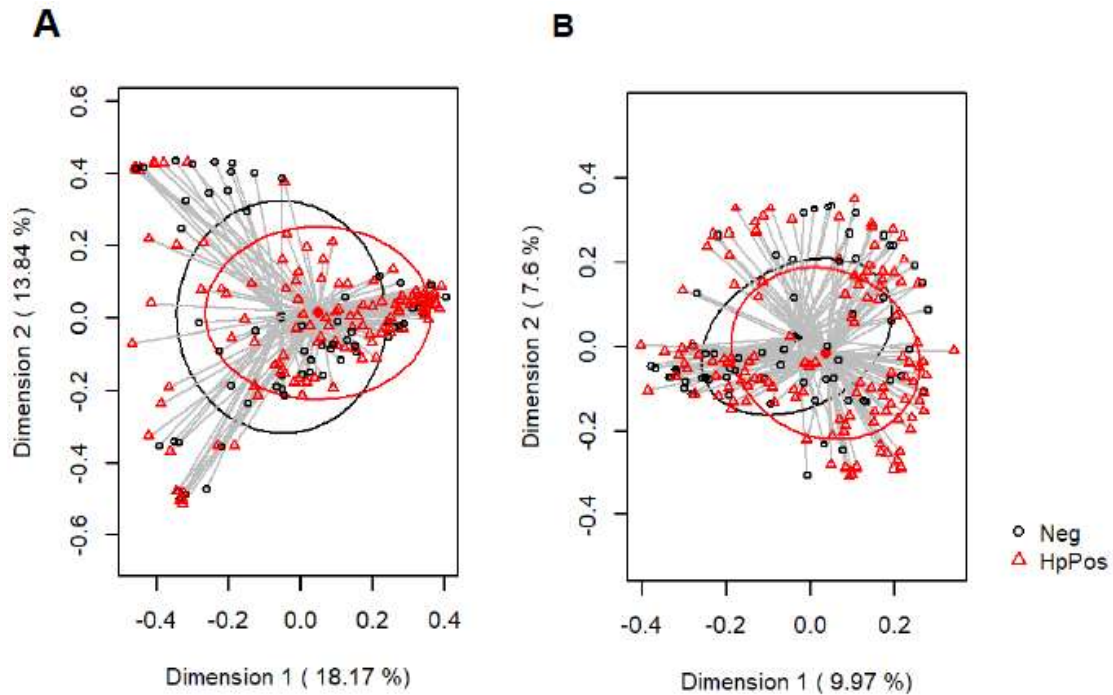


Figura 20: PCoAs de los perfiles bacterianos a nivel ASV basado en la matriz de distancia Bray-Curtis (A) y la matriz de distancia Jaccard binaria para el estudio por presencia/ausencia (B) de muestras de biopsias gástricas a las que se les ha sustraído las secuencias de *H. pylori*. El color indica el diagnóstico para la infección por *H. pylori* (Neg= negativo, círculo negro; HpPos= positivos, triángulo rojo). El valor en cada eje representa el porcentaje de variación explicado por ese eje.

Se identificaron un total de 22 filos diferentes siendo los filos *Firmicutes*, *Proteobacteria*, *Bacteroidetes* y *Actinobacteria* los más abundantes desplazando a las *Epsilonbacteraeota* a un filo detectado pero menor. Las diferencias entre los filos se hicieron menos patentes sin encontrar diferencias significativas en la abundancia relativa del filo *Firmicutes* o *Bacteroidetes* pero sí dentro del filo *Proteobacteria* y *Actinobacteria* (más abundantes en los pacientes negativos ($p < 0,001$ en ambas comparaciones)). La abundancia relativa de las secuencias cuyo filo no pudieron ser determinados fue muy superior en las muestras provenientes de pacientes positivos a *Helicobacter* ($p < 0,001$) (Tabla 15).

Tabla 15: Abundancia relativa (%), expresada como la mediana y los rangos intercuartílicos (IQR), de los 20 géneros y 4 filios mayoritarios en función del diagnóstico para *H. pylori* tras eliminar las secuencias de *H. pylori* de los análisis.

	<i>Helicobacter</i> negativo		<i>Helicobacter</i> positivos		P-valor†
	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	
<i>Firmicutes</i>	62 (98,41%)	24,66 (12,2-42,66)	118 (96,72%)	23,77 (10,36-46,04)	0,880
<i>Streptococcus</i>	55 (87,3%)	6,66 (0,45-18,45)	104 (85,25%)	2,71 (0,31-11,54)	0,120
<i>Lactobacillus</i>	32 (50,79%)	<0,01 (<0,01-0,97)	55 (45,08%)	<0,01 (<0,01-0,77)	0,810
<i>Veillonella</i>	42 (66,67%)	0,45 (<0,01-3,82)	79 (64,75%)	0,49 (<0,01-3,02)	0,850
<i>Staphylococcus</i>	44 (69,84%)	0,16 (<0,01-1,23)	49 (40,16%)	<0,01 (<0,01-0,27)	<0,001
<i>Clostridium</i>	14 (22,22%)	<0,01 (<0,01-<0,01)	19 (15,57%)	<0,01 (<0,01-<0,01)	0,310
<i>Lactococcus</i>	7 (11,11%)	<0,01 (<0,01-<0,01)	19 (15,57%)	<0,01 (<0,01-<0,01)	0,290
<i>Proteobacteria</i>	63 (100%)	18,7 (9,48-63,15)	109 (89,34%)	7,35 (1,28-28,74)	<0,001
<i>Pseudomonas</i>	49 (77,78%)	0,32 (0,02-27,41)	45 (36,89%)	<0,01 (<0,01-0,23)	<0,001
<i>Escherichia</i>	40 (63,49%)	0,17 (<0,01-0,86)	61 (50%)	0,02 (<0,01-1,25)	0,490
<i>Methylobacterium</i>	24 (38,1%)	<0,01 (<0,01-0,33)	36 (29,51%)	<0,01 (<0,01-0,22)	0,340
<i>Bacteroidetes</i>	61 (96,83%)	10,95 (3,22-18,64)	115 (94,26%)	11,94 (2,06-21,71)	0,870
<i>Prevotella</i>	47 (74,6%)	0,91 (0,02-5,15)	101 (82,79%)	2,92 (0,44-12,35)	0,006
<i>Alloprevotella</i>	31 (49,21%)	<0,01 (<0,01-2,31)	77 (63,11%)	0,31 (<0,01-1,75)	0,330
<i>Bacteroides</i>	31 (49,21%)	<0,01 (<0,01-1,56)	46 (37,7%)	<0,01 (<0,01-0,41)	0,066
<i>Actinobacteria</i>	61 (96,83%)	3,39 (1,33-8,51)	91 (74,59%)	0,78 (0,01-4,35)	<0,001
<i>Rothia</i>	36 (57,14%)	0,1 (<0,01-0,79)	54 (44,26%)	<0,01 (<0,01-0,62)	0,180
<i>Actinomyces</i>	41 (65,08%)	0,57 (<0,01-2,87)	46 (37,7%)	<0,01 (<0,01-0,44)	<0,001
Minor phyla	54 (85,71%)	4,16 (0,54-20,95)	108 (88,52%)	2,67 (0,4-10,77)	0,170
<i>Fusobacterium</i>	37 (58,73%)	0,92 (<0,01-5,28)	76 (62,3%)	0,58 (<0,01-2,83)	0,670
<i>Leptotrichia</i>	33 (52,38%)	0,15 (<0,01-1,7)	70 (57,38%)	0,14 (<0,01-1,4)	0,920
<i>Campylobacter</i>	31 (49,21%)	<0,01 (<0,01-2,85)	45 (36,89%)	<0,01 (<0,01-0,31)	0,017
<i>Treponema</i>	24 (38,1%)	<0,01 (<0,01-1,05)	38 (31,15%)	<0,01 (<0,01-0,12)	0,180
Minor genera	63 (100%)	20,38 (15,54-33,3)	118 (96,72%)	21,47 (7,69-35,92)	0,260
Filo no clasificado	48 (76,19%)	0,3 (0,01-1,63)	93 (76,23%)	4,76 (0,04-30,95)	<0,001
Género no clasificado	63 (100%)	6,21 (2,49-10,88)	120 (98,36%)	10,93 (2,92-32,81)	0,006

#n (%): Número de muestras que detectaron el filo o género (frecuencia relativa de detección).

† Prueba Wilconxon de rangos con la corrección de Bonferroni.

Al eliminar el género *Helicobacter* de entre los 20 géneros mayoritarios se aprecia una “normalización” de las muestras de los pacientes positivos hacia los perfiles de los pacientes negativos. Múltiples géneros que originalmente presentaron diferencias significativas (*Streptococcus*, *Veillonella* o *Escherichia*) no mantuvieron esa tendencia con la eliminación de las secuencias de *Helicobacter* y solo uno que originalmente no presentó diferencias significativas, *Prevotella*,

Resultados

presentó diferencias significativas ($p = 0,006$) al aumentar su abundancia relativa en las muestras de pacientes *Helicobacter* positivos (Tabla 15).

Gráficamente podemos observar dicha tendencia al comparar los grupos de pacientes *H. pylori* positivos con los *H. pylori* negativos a pesar del aumento de la abundancia relativa de las secuencias de géneros no clasificados en el grupo de pacientes *H. pylori* positivos tras la extracción de las secuencias de las bacterias del género *Helicobacter* (Fig. 21).

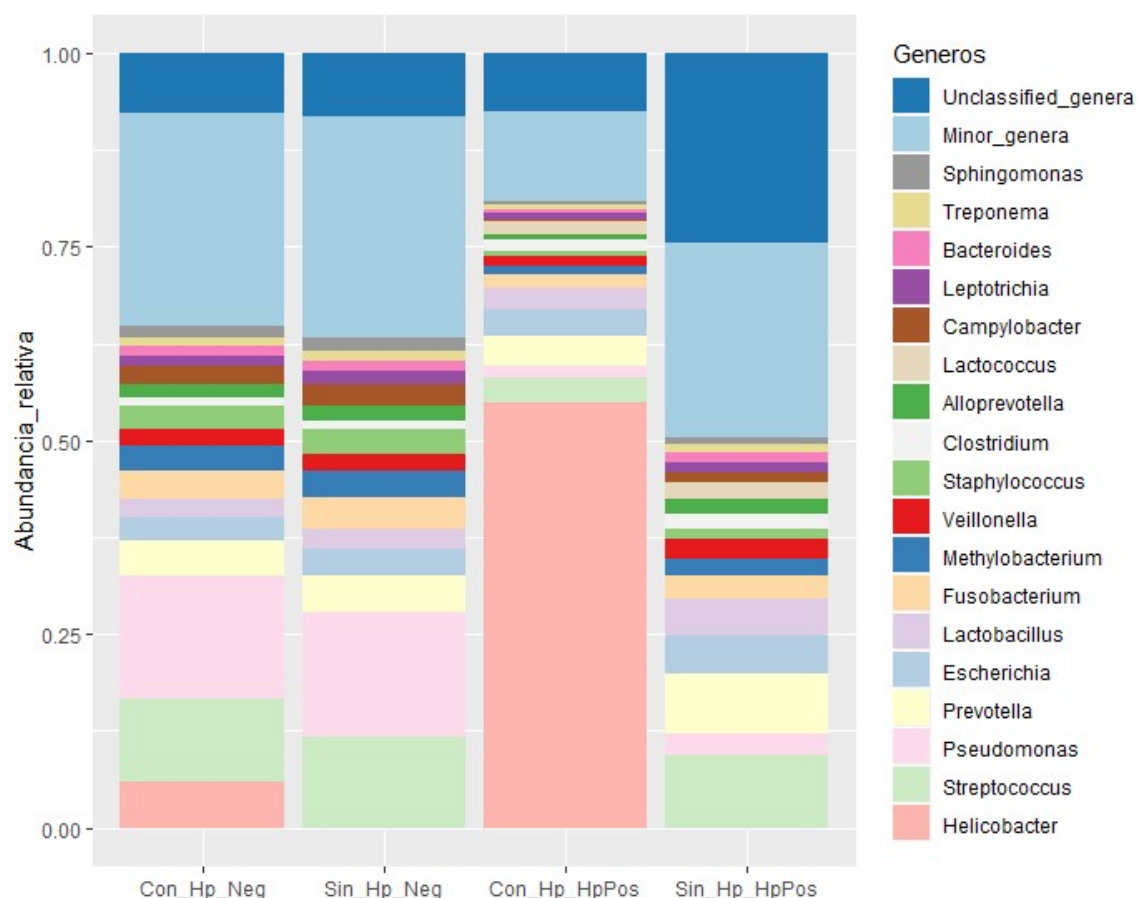


Figura 21: Comparación de la abundancia relativa de los géneros más abundantes de las muestras de biopsias en función de la extracción de las secuencias de *H. pylori* (Con secuencias = Con_; sin secuencias = Sin_) y el estado de infección por *H. pylori* (Negativo = Hp_Neg; Positivos = Hp_HpPos).

Con la sustracción de las secuencias del género *Helicobacter* y de las muestras cuyo número de secuencias no alcanzó el mínimo de secuencias (1.000 secuencias) no se aprecia una red de relaciones entre los géneros mayoritarios, más allá de una correlación positiva de los dos géneros del filo *Actinobacteria* (*Actinomyces* y *Rothia*) (Fig. 22-A).

Resultados

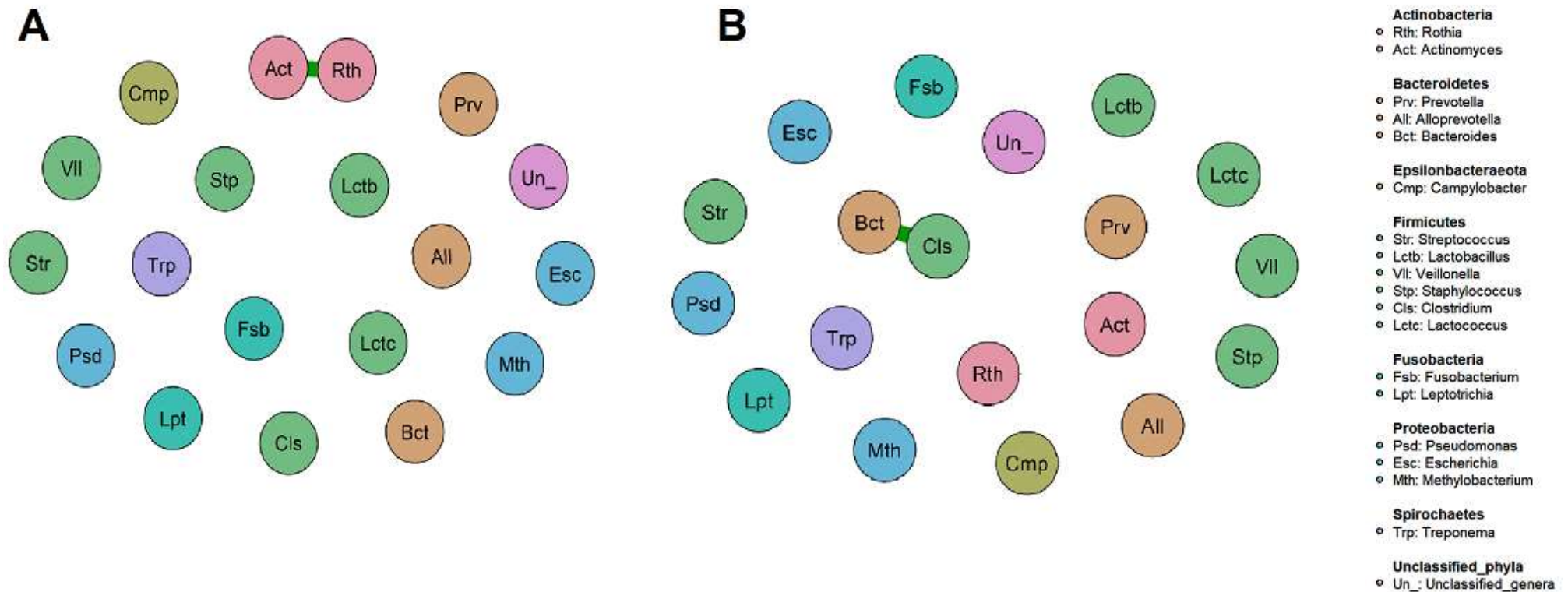


Figura 22: Estructura de relaciones estimadas basada en 185 muestras de biopsia gástrica: 122 muestras de pacientes diagnosticados positivos (A) y 63 muestras de pacientes diagnosticados negativos (B) tras eliminar las secuencias de las bacterias del género *Helicobacter* del análisis. Han sido representados los 19 géneros mayoritarios. Las líneas rojas indican correlaciones negativas y las líneas verdes correlaciones positivas...

Resultados

El mismo tipo de efecto se aprecia en las muestras procedentes de los pacientes diagnosticados como negativos. No se aprecia una red de correlaciones y solo destaca una correlación positiva entre los géneros *Bacteroides* y *Clostridium* (Fig.22-B).

5.7 Influencia de la patología

Para ver el efecto de las patologías más habituales relacionadas con la infección por *H. pylori* se valoró si el paciente tenía algún trastorno digestivo. Todos los pacientes fueron reclutados en diferentes hospitales por patología digestiva que precisaba de una biopsia gástrica, por lo que ningún paciente puede ser considerado sano.

Del total de 223 muestras analizadas, se formaron 8 grupos en función de su diagnóstico: pacientes diagnosticados como negativos para infección por *Helicobacter*, pero con patología de dispepsia, gastritis o cáncer gástrico y pacientes diagnosticados como positivos para infección por *Helicobacter* con las patologías dispepsia, gastritis, úlcera, metaplasia o cáncer (Tabla 16).

Tabla 16: Número de muestras de biopsia gástrica por diagnóstico para la infección por *H. pylori* (negativos= Neg; Positivos = HpPos) y patología gástrica asociada. Se muestra el número de secuencias e información sobre el número mínimo (Min), mediana, cuartil 1 (Q1), cuartil 3 (Q3) y el número máximo de secuencias en una muestra de ese grupo (Max).

Factor	Muestras	Nº secuencias	Min	Mediana	Q1	Q3	Max
Neg_Dispepsia	43	726.320	1.098	6.969	2.295	19.752	159.134
Neg_Gastritis	15	126.970	1.409	4.893	2.867	11.090	34.199
Neg_Cáncer	6	515.010	3.394	34.161	8.282	61.397	373.161
HpPos_Dispepsia	18	300.535	2.830	10.120	5.632	22.376	67.863
HpPos_Gastritis	58	4.025.828	1.980	53.293	17.320	101.220	225.024
HpPos_Ulcera	4	80.931	7.834	23.300	19.153	24.379	26.497
HpPos_Metaplasia	3	133.343	36.785	46.079	41.432	48.279	50.479
HpPos_Cáncer	76	2.153.478	1.961	20.789	15.094	29.776	182.297
Total	223	8.062.415	1.098	19.528	8.485	38.995	373.161

La separación de las muestras en diferentes grupos en función de la patología permitió el estudio en profundidad de la interacción microbiota-patología a expensas de bajar la significación estadística.

5.7.1 Diversidad microbiana por patología en pacientes negativos

No se obtuvieron diferencias significativas en los valores del índice de Shannon en función de la patología diagnosticada, pero se observó una tendencia hacia una menor diversidad en función de la gravedad de la patología gástrica, siendo la dispepsia la patología con un mayor índice de diversidad y el cáncer el de menor (Tabla 17).

Tabla 17: Biodiversidad, medida por el índice de Shannon y Simpson, en muestras de biopsia gástrica de pacientes diagnosticados como negativos a infección por *H. pylori* agrupados por patología gástrica.

Índice de diversidad	Pacientes Negativos			P Valor†
	Dispepsia	Gastritis	Cáncer	
Shannon	2,54 (1,89-2,78)	1,83 (1,4-2,84)	1,39 (0,87-2,25)	1,000
Simpson	0,86 (0,73-0,89)	0,67 (0,59-0,89)	0,60 (0,32-0,84)	1,000

† Prueba Kruskal–Wallis de rangos con la corrección de Bonferroni.

Entre los perfiles de los pacientes negativos, el grupo diagnosticado con gastritis es el que presenta un perfil de microorganismos (tanto por su abundancia (Fig. 23-A) como por la presencia y ausencia de microorganismos (Fig 23-B)) estadísticamente diferente con respecto a dispepsia (Bray Curtis $p = 0,028$ y Jaccard binario $p = 0,027$) o cáncer (Bray Curtis $p=0,028$ y Jaccard binario $p=0,028$).

Resultados

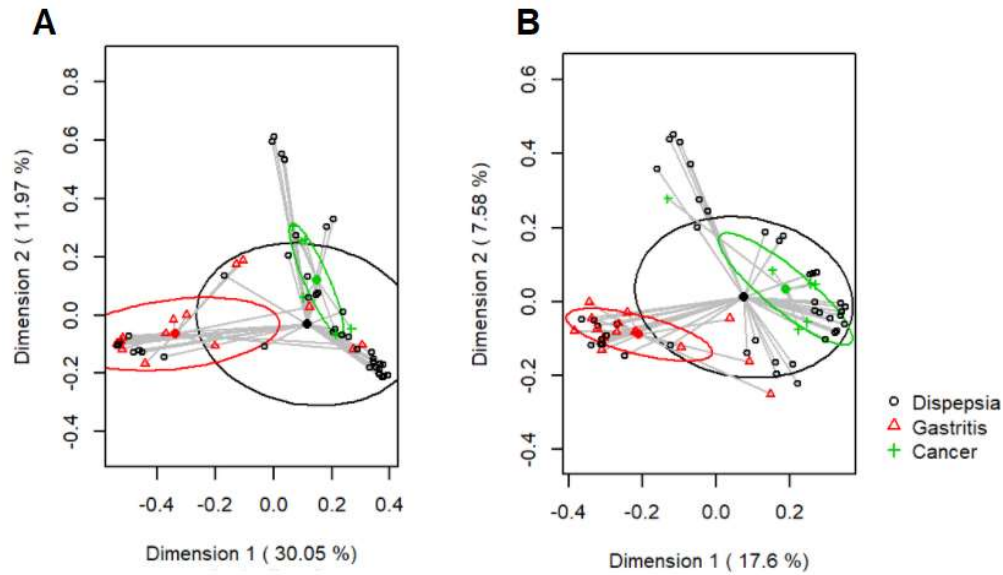


Figura 23: PCoAs de los perfiles bacterianos a nivel ASV basado en la matriz de distancia Bray-Curtis (A) y la matriz de distancia Jaccard binaria para el estudio por presencia/ausencia (B) de muestras de biopsias gástricas de pacientes negativos a infección por *H. pylori* agrupadas por patología gástrica (dispepsia, círculo negro; gastritis, triángulo rojo; cáncer, cruz verde). El valor en cada eje representa el porcentaje de variación explicado por ese eje.

Entre sus filos mayoritarios, el filo *Proteobacteria* fue el que presentó mayores diferencias entre los grupos de pacientes negativos. El grupo con gastritis fue el que presentó una mayor abundancia relativa 43.26 (28.35-69.09), frente a dispepsia 15.73 (9.48-58.53) ($p > 0.05$) y frente a cáncer gástrico 2,16 (0,64-13,29) ($p < 0.05$), representado mayoritariamente, por bacterias del género *Pseudomonas* siendo este el género que presentó mayores diferencias significativas (Fig. 24), entre el grupo gastritis con dispepsia ($p=0,014$) y entre gastritis y cáncer ($p=0,013$), del grupo de los pacientes negativos.

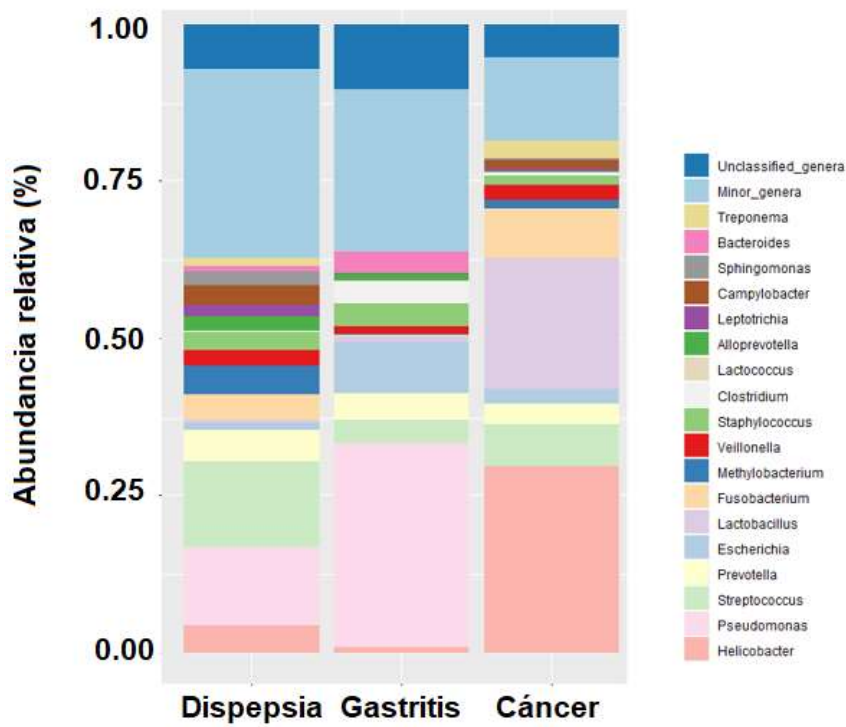


Figura 24: Comparación de la abundancia relativa de los géneros más abundantes de las muestras de biopsias en función de la patología de las de pacientes negativos a infección por *H. pylori*.

El género *Bacteroides* presentó una mayor abundancia y diferencias significativas en el grupo dispepsia con respecto a gastritis ($p=0,017$) (Tabla 18).

Resultados

Tabla 18: Abundancia relativa (%), expresada como la mediana y los rangos intercuartílicos (IQR), de los 20 géneros y 4 filios mayoritarios en función la patología en pacientes *H. pylori* negativos.

Filos/ géneros	Dispepsia		Gastritis		Cáncer		P- valor†
	N# (%)	Mediana (IQR)	N# (%)	Mediana (IQR)	N# (%)	Mediana (IQR)	
<i>Proteobacteria</i>	43 (100%)	15,73 (9,48-58,53)	15 (100%)	43,26 (28,35-69,09)	6 (100%)	2,16 (0,64-13,29)	0,002
<i>Pseudomonas</i>	33 (76,74%)	0,16 (0,01-5,97)	15 (100%)	25,2 (5,36-53,76)	1 (16,67%)	<0,01 (<0,01-<0,01)	<0,001
<i>Methylobacterium</i>	20 (46,51%)	<0,01 (<0,01-1,06)	2 (13,33%)	<0,01 (<0,01-<0,01)	3 (50%)	<0,01 (<0,01-0,67)	0,053
<i>Escherichia</i>	26 (60,47%)	0,05 (<0,01-0,57)	11 (73,33%)	0,46 (0,08-2,26)	4 (66,67%)	0,14 (<0,01-0,5)	0,210
<i>Sphingomonas</i>	13 (30,23%)	<0,01 (<0,01-0,02)	5 (33,33%)	<0,01 (<0,01-0,02)	2 (33,33%)	<0,01 (<0,01-0,01)	1,000
<i>Novosphingobium</i>	7 (16,28%)	<0,01 (<0,01-<0,01)	1 (6,67%)	<0,01 (<0,01-<0,01)	2 (33,33%)	<0,01 (<0,01-<0,01)	0,360
<i>Firmicutes</i>	42 (97,67%)	23,18 (12,34-37,8)	15 (100%)	15,89 (6,41-33,16)	6 (100%)	37,5 (13,14-53,28)	0,370
<i>Streptococcus</i>	38 (88,37%)	8,8 (1,82-23,88)	12 (80%)	0,49 (0,06-3,82)	6 (100%)	4,96 (3,17-8,65)	0,027
<i>Lactobacillus</i>	16 (37,21%)	<0,01 (<0,01-0,15)	11 (73,33%)	1,14 (0,04-1,56)	5 (83,33%)	2,15 (0,12-28,52)	0,002
<i>Staphylococcus</i>	30 (69,77%)	0,12 (<0,01-1,47)	9 (60%)	0,08 (<0,01-0,39)	5 (83,33%)	0,08 (0,04-1,13)	0,710
<i>Veillonella</i>	29 (67,44%)	0,33 (<0,01-3,77)	7 (46,67%)	<0,01 (<0,01-0,53)	6 (100%)	0,68 (0,58-0,77)	0,110
<i>Bacteroidetes</i>	40 (93,02%)	10,57 (2,93-18,56)	15 (100%)	13,29 (6,68-18,02)	6 (100%)	1,75 (0,93-14,1)	0,400
<i>Prevotella</i>	32 (74,42%)	1,02 (0,02-6,92)	9 (60%)	0,07 (<0,01-1,65)	6 (100%)	0,35 (0,09-0,87)	0,400
<i>Bacteroides</i>	16 (37,21%)	<0,01 (<0,01-0,96)	13 (86,67%)	2,31 (0,29-3,97)	2 (33,33%)	<0,01 (<0,01-0,01)	<0,001
<i>Epsilonbacteraeota</i>	34 (79,07%)	1,28 (0,18-9,74)	9 (60%)	0,29 (<0,01-2,4)	6 (100%)	8,74 (2,76-61,17)	0,020
<i>Helicobacter</i>	27 (62,79%)	0,06 (<0,01-1,54)	9 (60%)	0,29 (<0,01-1,93)	6 (100%)	8,33 (2,19-58,24)	0,024
<i>Campylobacter</i>	24 (55,81%)	0,36 (<0,01-4,81)	2 (13,33%)	<0,01 (<0,01-<0,01)	5 (83,33%)	0,79 (0,55-1,29)	0,006
<i>Actinobacteria</i>	42 (97,67%)	3,72 (1,67-8,2)	14 (93,33%)	1,92 (0,7-7,31)	6 (100%)	1,77 (0,53-3,67)	0,230
<i>Alloprevotella</i>	22 (51,16%)	0,05 (<0,01-3,41)	5 (33,33%)	<0,01 (<0,01-0,45)	4 (66,67%)	0,09 (<0,01-0,29)	0,430
<i>Rothia</i>	25 (58,14%)	0,06 (<0,01-0,52)	6 (40%)	<0,01 (<0,01-1,64)	5 (83,33%)	0,04 (0,02-0,23)	0,960
<i>Actinomyces</i>	29 (67,44%)	0,7 (<0,01-2,99)	7 (46,67%)	<0,01 (<0,01-0,42)	5 (83,33%)	0,44 (0,31-0,63)	0,130
Minor phyla	38 (88,37%)	6,19 (1,14-18,03)	15 (100%)	2 (0,92-4,97)	5 (83,33%)	2,75 (1,03-22,1)	0,500
<i>Fusobacterium</i>	27 (62,79%)	1,13 (<0,01-7,1)	5 (33,33%)	<0,01 (<0,01-0,67)	5 (83,33%)	1,33 (0,47-14,17)	0,015
<i>Leptotrichia</i>	25 (58,14%)	0,28 (<0,01-2,34)	3 (20%)	<0,01 (<0,01-<0,01)	5 (83,33%)	0,32 (0,21-0,81)	0,028
Minor genera	43 (100%)	21,8 (16,29-32,38)	15 (100%)	19,04 (17-28,9)	6 (100%)	4,07 (3,31-28,01)	0,300
Género no clasificado	43 (100%)	5,41 (2,39-9,26)	15 (100%)	9,81 (3,66-15,23)	6 (100%)	1,6 (0,66-7,15)	0,130

#n (%): Número de muestras que detectaron el filo o género (frecuencia relativa de detección).

† Prueba Kruskal–Wallis de rangos con la corrección de Bonferroni.

5.7.2 Diversidad microbiana por patología en pacientes positivos

En el caso de las muestras de pacientes positivos, se aprecia una evolución de la microbiota en función del estadio de la patología. En general, los índices de diversidad tienden a ser más bajos en los pacientes diagnosticados negativos, principalmente por la mayor abundancia relativa de bacterias del género *Helicobacter*, siendo éstas las que más modularán la diversidad de las muestras. Así pues, se aprecia un descenso de la diversidad entre los grupos de dispepsia a los de gastritis (Shannon $p=0,018$ y Simpson $p=0,006$) y un aumento progresivo en el grupo de cáncer gástrico (Shannon $p<0,001$ y Simpson $p<0,001$) (Fig. 25). En este estudio, los valores de diversidad de la úlcera y/o metaplasia hay que considerarlos con precaución debido al bajo número de muestras que cumplieron los requisitos de inclusión. No se encontraron diferencias significativas entre el grupo de gastritis y úlcera o metaplasia a pesar de poseer unos valores de diversidad elevados ($p=1,000$) en ambos casos.

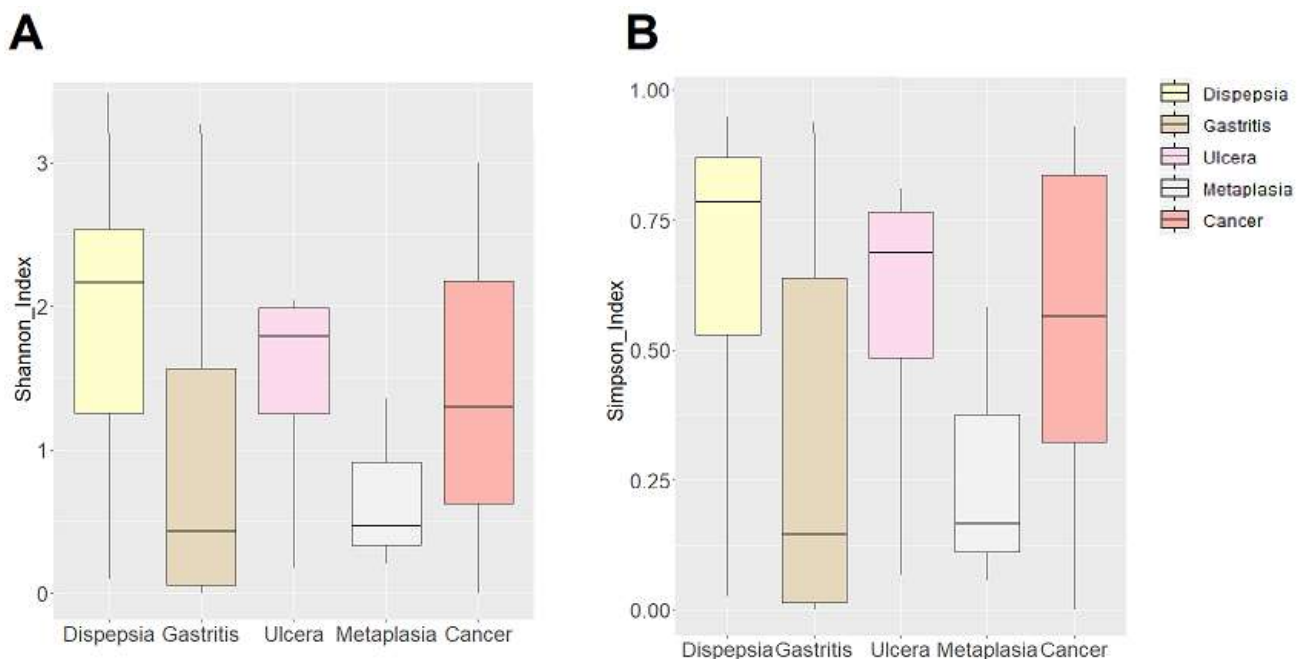


Figura 25: Diagrama de cajas mostrando la biodiversidad, medida por el índice de Shannon (A) y Simpson (B) en muestras de biopsia gástrica de pacientes diagnosticados como positivos a infección por *H. pylori* agrupados en función de su patología gástrica.

Resultados

En los estudios de diversidad beta se identificaron, diferentes perfiles microbianos en función de la patología. En los estudios por abundancia, los análisis indicaron una separación significativa del grupo de gastritis con respecto a los grupos de dispepsia ($p=0,034$), úlcera ($p=0,028$) y cáncer ($p=0,028$). Tampoco se observaron diferencias entre los grupos de dispepsia, úlcera, metaplasia o cáncer ($p>0,05$). Por otro lado, los estudios por presencia o ausencia indicaron diferencias significativas entre cada uno de los grupos de pacientes con dispepsia, gastritis o cáncer (comparación por pares $p = 0,028$ entre cada una de las 3 comparaciones).

Entre los filos mayoritarios, el filo *Epsilonbacteraeota* fue mayoritario en el grupo de dispepsia, metaplasia y, especialmente, en gastritis. Este grupo presentó diferencias significativas con los grupos de dispepsia ($p<0,01$) y cáncer ($p<0,001$). Por su parte, el grupo con patología de cáncer presentó una mayor abundancia relativa del filo *Firmicutes*, presentando diferencias significativas con el grupo con gastritis ($p<0,001$). Finalmente, el grupo con dispepsia presentó una mayor abundancia de Proteobacterias con respecto a gastritis ($p<0,01$). El filo *Bacteroidete* estuvo más presente en el grupo de dispepsia y cáncer, con una menor abundancia relativa en el grupo de gastritis ($p<0,05$ y $p<0,01$ respectivamente) (Tabla 19).

Estas diferencias en los filos vienen dadas por diferentes perfiles dentro de los géneros mayoritarios. Entre los 20 géneros mayoritarios, el género *Helicobacter* fue el que más abundancia relativa presentó en todos los grupos con la excepción del grupo úlcera, donde los géneros *Lactococcus* o *Prevotella* fueron más abundantes (Fig 26). Entre los 3 grupos con mayor número de muestras, 9 de los 20 géneros mayoritarios presentaron diferencias significativas: el género *Helicobacter* fue más abundante en el grupo con gastritis que en el grupo dispepsia ($p= 0,006$) y/o cáncer ($p<0,001$). El grupo con cáncer presentó una mayor abundancia y diferencias significativas en género *Prevotella* con el grupo gastritis ($p= 0,013$) y en el género *Bacteroides* con el grupo dispepsia ($p= 0,020$). El grupo con dispepsia mostró una mayor abundancia y diferencias significativas en género *Escherichia* con respecto al grupo gastritis ($p= 0,014$) junto a los géneros *Methylobacterium* o *Sphingomonas* que mostraron diferencias

significativas tanto en gastritis ($p= 0,001$ y $p<0,001$) como en cáncer ($p= 0,002$ y $p<0,001$) (Tabla 19).

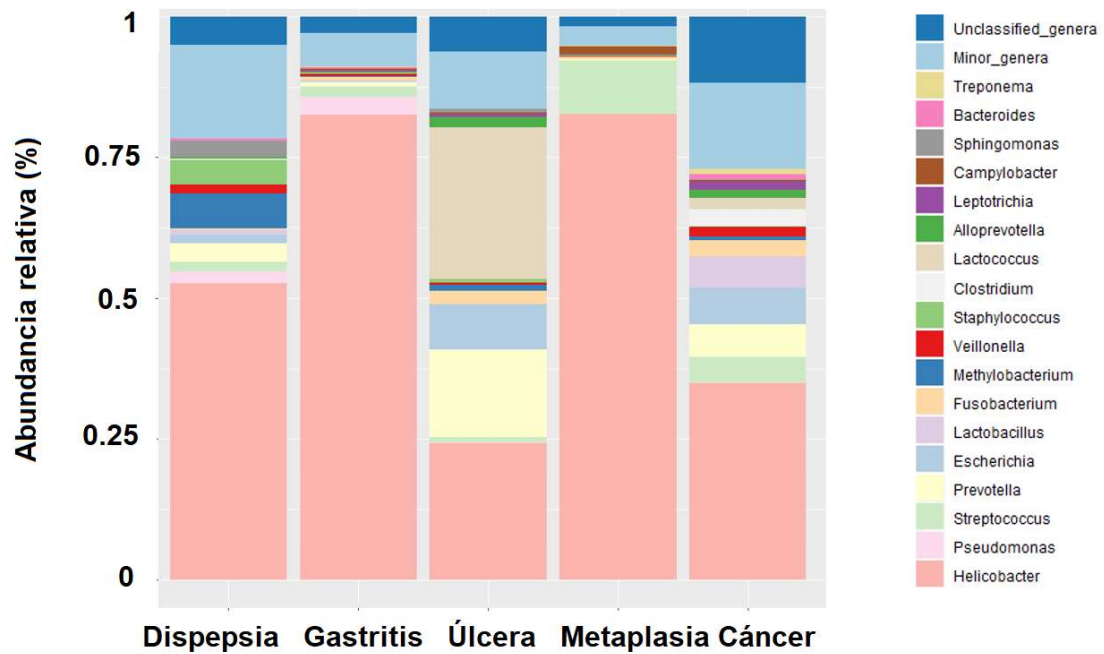


Figura 26: Comparación de la abundancia relativa de los 20 géneros más abundantes de las muestras de biopsias en función de la patología gástrica de pacientes infectados por *H. pylori*.

Resultados

Tabla 19: Abundancia relativa (%), expresada como la mediana y los rangos intercuartílicos (IQR), de los 20 géneros y 4 filos mayoritarios en función la patología en pacientes *H. pylori* positivos.

Filos/géneros	Dispepsia		Gastritis		Úlcera		Metaplasia		Cáncer		P-Valor †
	N (%)	Mediana (IQR)	N (%)	Mediana (IQR)	N (%)	Mediana (IQR)	N (%)	Mediana (IQR)	N (%)	Mediana (IQR)	
<i>Epsilonbacteraeota</i>	17 (94,44%)	67,62 (16,06-91,8)	58 (100%)	97,93 (80,14-99,68)	4 (100%)	0,65 (0,22-24,95)	3 (100%)	91,87 (77,39-94,56)	67 (88,16%)	9,76 (0,31-75,47)	<0,001
<i>Helicobacter</i>	17 (94,44%)	67,52 (16,06-91,8)	58 (100%)	97,93 (80,11-99,67)	2 (50%)	0,12 (<0,01-24,33)	3 (100%)	91,16 (75,5-94,16)	54 (71,05%)	9,18 (<0,01-75,47)	<0,001
<i>Campylobacter</i>	3 (16,67%)	<0,01 (<0,01-<0,01)	33 (56,9%)	0,01 (<0,01-0,07)	2 (50%)	0,07 (<0,01-0,38)	3 (100%)	0,71 (0,4-1,89)	23 (30,26%)	<0,01 (<0,01-0,06)	0,220
<i>Firmicutes</i>	18 (100%)	6,91 (2,2-12,46)	56 (96,55%)	0,46 (0,09-5,51)	4 (100%)	30,99 (10,6-51,81)	3 (100%)	6,42 (3,71-16,25)	72 (94,74%)	12,93 (0,59-39,04)	<0,001
<i>Streptococcus</i>	16 (88,89%)	0,38 (0,13-0,87)	54 (93,1%)	0,14 (0,05-0,91)	4 (100%)	0,24 (0,15-1,02)	3 (100%)	4,27 (2,47-13,85)	60 (78,95%)	0,52 (0,02-4,55)	0,160
<i>Lactobacillus</i>	8 (44,44%)	<0,01 (<0,01-0,55)	20 (34,48%)	<0,01 (<0,01-0,01)	1 (25%)	<0,01 (<0,01-0,05)	1 (33,33%)	<0,01 (<0,01-0,1)	30 (39,47%)	<0,01 (<0,01-0,99)	0,400
<i>Veillonella</i>	11 (61,11%)	0,04 (<0,01-0,67)	33 (56,9%)	0,02 (<0,01-0,17)	2 (50%)	0,02 (<0,01-0,53)	3 (100%)	0,21 (0,12-0,31)	43 (56,58%)	0,1 (<0,01-1,03)	0,540
<i>Staphylococcus</i>	9 (50%)	<0,01 (<0,01-0,04)	30 (51,72%)	0,01 (<0,01-0,07)	3 (75%)	0,03 (0,01-0,64)	3 (100%)	0,03 (0,03-0,06)	16 (21,05%)	<0,01 (<0,01-<0,01)	0,004
<i>Clostridium</i>	5 (27,78%)	<0,01 (<0,01-0,01)	4 (6,9%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	1 (33,33%)	<0,01 (<0,01-0,01)	10 (13,16%)	<0,01 (<0,01-<0,01)	0,150
<i>Lactococcus</i>	3 (16,67%)	<0,01 (<0,01-<0,01)	5 (8,62%)	<0,01 (<0,01-<0,01)	4 (100%)	23,82 (3,07-47,66)	1 (33,33%)	<0,01 (<0,01-<0,01)	8 (10,53%)	<0,01 (<0,01-<0,01)	<0,001
<i>Proteobacteria</i>	18 (100%)	8,39 (1,06-46,24)	52 (89,66%)	0,4 (0,02-2,01)	4 (100%)	18,94 (9,46-32,14)	3 (100%)	0,41 (0,33-2,07)	59 (77,63%)	1,09 (0,05-11,38)	<0,001
<i>Pseudomonas</i>	9 (50%)	0,01 (<0,01-0,06)	27 (46,55%)	<0,01 (<0,01-0,07)	2 (50%)	0,03 (<0,01-0,26)	1 (33,33%)	<0,01 (<0,01-<0,01)	15 (19,74%)	<0,01 (<0,01-<0,01)	0,018
<i>Escherichia</i>	15 (83,33%)	0,25 (0,03-0,91)	28 (48,28%)	<0,01 (<0,01-0,1)	2 (50%)	1,27 (<0,01-9,22)	2 (66,67%)	0,03 (0,02-0,05)	24 (31,58%)	<0,01 (<0,01-0,43)	0,016
<i>Methylobacterium</i>	11 (61,11%)	1,18 (<0,01-6,35)	12 (20,69%)	<0,01 (<0,01-<0,01)	4 (100%)	1,08 (0,37-1,74)	3 (100%)	0,02 (0,01-0,02)	10 (13,16%)	<0,01 (<0,01-<0,01)	<0,001
<i>Sphingomonas</i>	11 (61,11%)	0,45 (<0,01-4,63)	7 (12,07%)	<0,01 (<0,01-<0,01)	3 (75%)	0,44 (0,12-0,93)	2 (66,67%)	0,01 (<0,01-0,02)	10 (13,16%)	<0,01 (<0,01-<0,01)	0,008

Continuación Tabla 19

Resultados

Filos/ géneros	Dispepsia		Gastritis		Úlcera		Metaplasia		Cáncer		P-Valor †
	N (%)	Mediana (IQR)	N (%)	Mediana (IQR)	N (%)	Mediana (IQR)	N (%)	Mediana (IQR)	N (%)	Mediana (IQR)	
<i>Bacteroidetes</i>	18 (100%)	2,8 (0,96-7,49)	57 (98,28%)	0,26 (0,06-1,72)	4 (100%)	21,49 (15,93-23,27)	3 (100%)	0,33 (0,27-0,49)	65 (85,53%)	2,41 (0,22-18,46)	<0,001
<i>Prevotella</i>	16 (88,89%)	1,01 (0,16-2,78)	48 (82,76%)	0,05 (0,01-0,36)	4 (100%)	20,42 (14,81-21,15)	3 (100%)	0,07 (0,05-0,32)	57 (75%)	0,99 (0,02-6,49)	<0,001
<i>Alloprevotella</i>	10 (55,56%)	0,08 (<0,01-0,21)	40 (68,97%)	0,02 (<0,01-0,11)	3 (75%)	0,48 (0,31-2,03)	2 (66,67%)	0,03 (0,01-0,12)	41 (53,95%)	0,03 (<0,01-0,61)	0,570
<i>Bacteroides</i>	13 (72,22%)	0,08 (<0,01-0,19)	23 (39,66%)	<0,01 (<0,01-0,11)	1 (25%)	<0,01 (<0,01-0,02)	0 (0%)	<0,01 (<0,01-<0,01)	16 (21,05%)	<0,01 (<0,01-<0,01)	0,042
Minor phyla	18 (100%)	0,62 (0,26-4,3)	58 (100%)	0,42 (0,06-1,41)	4 (100%)	3,38 (1,67-5,01)	3 (100%)	1,16 (1,12-1,67)	65 (85,53%)	1,26 (0,12-8,37)	<0,001
<i>Fusobacterium</i>	11 (61,11%)	0,08 (<0,01-0,2)	35 (60,34%)	0,01 (<0,01-0,15)	2 (50%)	1,54 (<0,01-3,91)	3 (100%)	0,2 (0,18-0,38)	42 (55,26%)	0,12 (<0,01-1,16)	0,300
<i>Leptotrichia</i>	6 (33,33%)	<0,01 (<0,01-0,04)	33 (56,9%)	0,01 (<0,01-0,07)	1 (25%)	<0,01 (<0,01-0,48)	2 (66,67%)	0,12 (0,06-0,36)	39 (51,32%)	0,03 (<0,01-0,88)	0,140
<i>Treponema</i>	3 (16,67%)	<0,01 (<0,01-<0,01)	23 (39,66%)	<0,01 (<0,01-0,01)	0 (0%)	<0,01 (<0,01-<0,01)	2 (66,67%)	0,07 (0,04-0,12)	22 (28,95%)	<0,01 (<0,01-0,03)	0,200
Minor genera	18 (100%)	7,14 (2,13-25,28)	56 (96,55%)	1 (0,1-6,12)	4 (100%)	5,87 (1,24-14,71)	3 (100%)	2,63 (1,84-4,63)	72 (94,74%)	6,3 (0,76-28,61)	0,011
Filo no clasificado	6 (33,33%)	<0,01 (<0,01-0,28)	37 (63,79%)	0,02 (<0,01-0,85)	1 (25%)	<0,01 (<0,01-0,39)	3 (100%)	0,02 (0,01-2,22)	60 (78,95%)	9,05 (0,96-18,47)	0,310
Género no clasificado	17 (94,44%)	1,09 (0,28-4,22)	48 (82,76%)	0,1 (0,01-2,69)	4 (100%)	1,51 (1,13-6,66)	3 (100%)	0,26 (0,19-2,58)	72 (94,74%)	11,03 (3,97-20,06)	<0,001

N (%): Número de muestras que detectaron el filo o género (frecuencia relativa de detección).

† Prueba Kruskal–Wallis de rangos con la corrección de Bonferroni.

Resultados

5.7.3 Diversidad microbiana por patología eliminando las secuencias pertenecientes a bacterias del género *Helicobacter*

Siguiendo la metodología del apartado 5.6.6, se procedió a la eliminación de las secuencias de *Helicobacter* de las muestras de biopsia gástrica para determinar el efecto de *H. pylori* en los distintos grupos de pacientes divididos por patología. Con la sustracción de dichas secuencias de las muestras de estudio, las muestras y grupos quedaron conforme a los datos de la tabla 20. Del total de secuencias analizadas (8.062.415) un total de 5.942.636 secuencias fueron identificadas como provenientes de bacterias del género *Helicobacter* y fueron eliminadas. Con el filtrado de dichas secuencias de las muestras de estudio solo 185 (122 muestras provenientes de pacientes positivos y 63 de pacientes negativos) de las 223 muestras de biopsia continuaron con al menos 1.000 secuencias para una correcta comparación.

Tabla 20: Número de muestras aceptadas tras eliminar las secuencias de las bacterias del género *Helicobacter* de biopsias gástricas por diagnóstico para la infección por *H. pylori* (negativos= Neg; Positivos = HpPos) y patología gástrica asociada. Se muestra el número de secuencias e información sobre el número mínimo (Min), mediana, cuartil 1 (Q1), cuartil 3 (Q3) y el número máximo de secuencias en una muestra de ese grupo (Max).

Factor	Muestras	Seq_total	Min	Mediana	Q1	Q3	Max
Neg_Dispepsia	42	567.900	1098	7.063,5	2.425,25	15.828,5	69.106
Neg_Gastritis	15	126.096	1409	4.891	2.821	10.993,5	34.199
Neg_Cáncer	6	150.996	3222	16.680	6.111,75	44.232	58.409
HpPos_Dispepsia	16	71.375	1136	3.771	2.346	5.773,5	10.282
HpPos_Gastritis	29	214.944	1014	2.801	2.085	7.608	46.392
HpPos_Ulcera	3	54.415	7815	22.927	15.371	23.300	23.673
HpPos_Metaplasia	3	20.540	1306	4.463	2.884,5	9.617	14.771
HpPos_Cáncer	71	900.600	1764	12.377	6.736	17.605,5	33.660

Al sustraer las secuencias de las bacterias pertenecientes al género *Helicobacter* se aprecia una normalización de las muestras procedentes de pacientes positivos perdiéndose muchas de las diferencias que antes poseían. Así pues, el grupo de gastritis de los pacientes positivos normaliza pasando de una diversidad con un índice de Shannon de 0,16 (0,03-1,01) a 2,44 (2,13-2,8) mucho más próximo a los valores de diversidad de las muestras negativas. Por otro lado, las muestras procedentes de pacientes positivos con la patología cáncer fueron los que

mostraron una menor diversidad, mostrando diferencias significativas con otros grupos como el grupo de dispepsia de los pacientes negativos ($p < 0,001$) o los pacientes positivos con gastritis ($p < 0,001$). (Fig. 27)

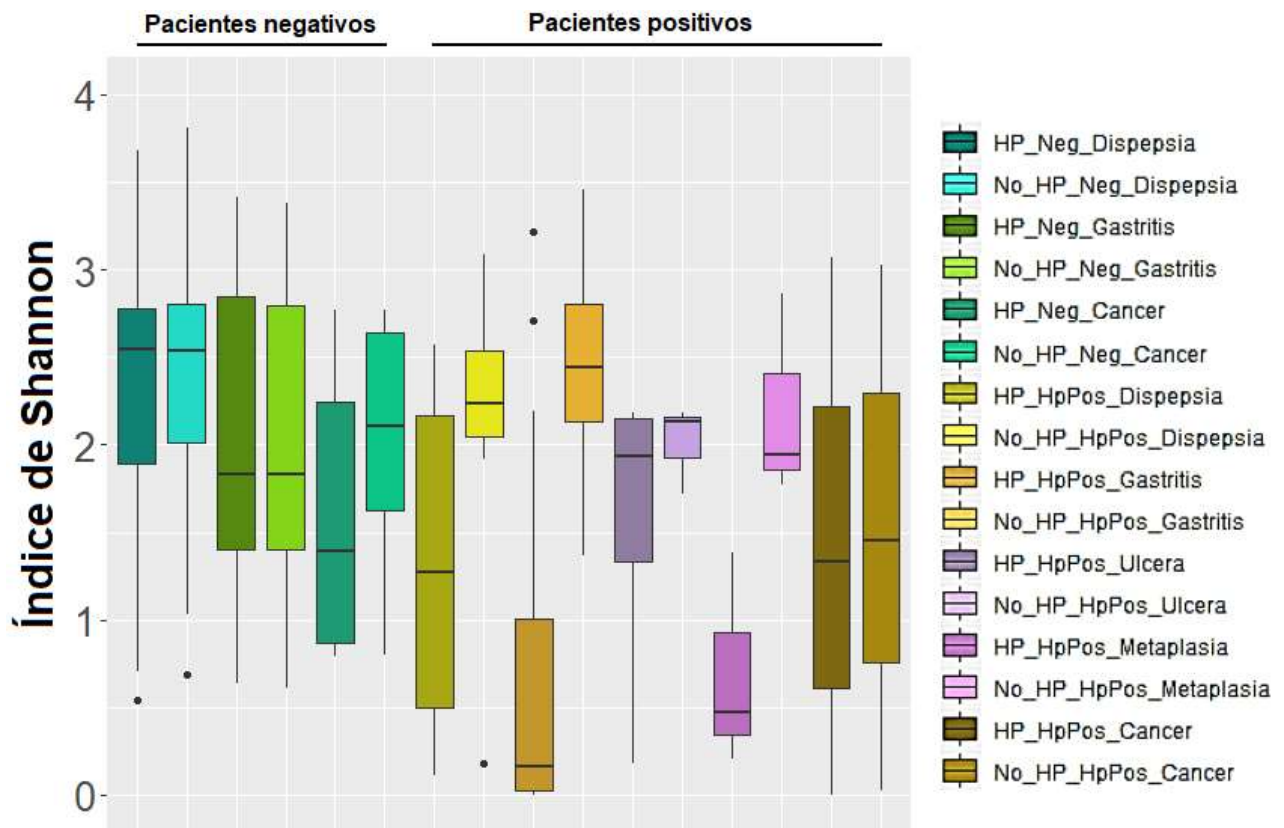


Figura 27: Diagrama de cajas mostrando la biodiversidad, medida por el índice de Shannon, de los pacientes negativos o positivos a la infección por *H. pylori* en función de la patología gástrica antes y después de sustraer las secuencias de las bacterias del género *Helicobacter*.

Los estudios de diversidad beta mostraron una normalización de los perfiles de los grupos positivos en casi todos los casos. En los estudios de abundancia, solo el grupo de los pacientes positivos con cáncer presentó diferencias con otros 2 grupos: positivos con gastritis ($p = 0,028$) y positivos con dispepsia ($p = 0,028$). En los estudios por presencia/ausencia solo se encontraron diferencias en la comparación por pares de pacientes negativos con dispepsia y pacientes positivos con cáncer ($p = 0,028$).

Resultados

Dentro de los filos mayoritarios se aprecia una normalización de todas las muestras, habiendo algunas diferencias aisladas como ocurre con el filo Proteobacteria entre los grupos de gastritis de los pacientes positivos y negativos ($p < 0.01$). También se aprecia un aumento de la abundancia relativa de las secuencias no clasificadas, especialmente en el grupo de cáncer positivos (Fig 28). Este aumento provoca diferencias que pueden no ser reales en las comparaciones con este grupo arrastrando este problema a las comparaciones de los géneros.



Figura 28: Comparación de la abundancia relativa de los 4 filos mayoritarios de las muestras de biopsias de los pacientes *H. pylori* negativos (A) y *H. pylori* positivos (B) en función de la patología gástrica con secuencias de bacterias del género *Helicobacter* (Hp) y sin las secuencias (Sin_Hp).

Entre los géneros más abundantes se encontraron diferencias, principalmente, en el grupo de pacientes positivos diagnosticados con cáncer gástrico debido a la gran abundancia relativa de secuencias no clasificadas dentro de algún género

17,31% (6,13-53,84) seguido, muy por abajo, del grupo de gastritis de los pacientes negativos 9,81% (3,66-15,44) (Fig. 29).

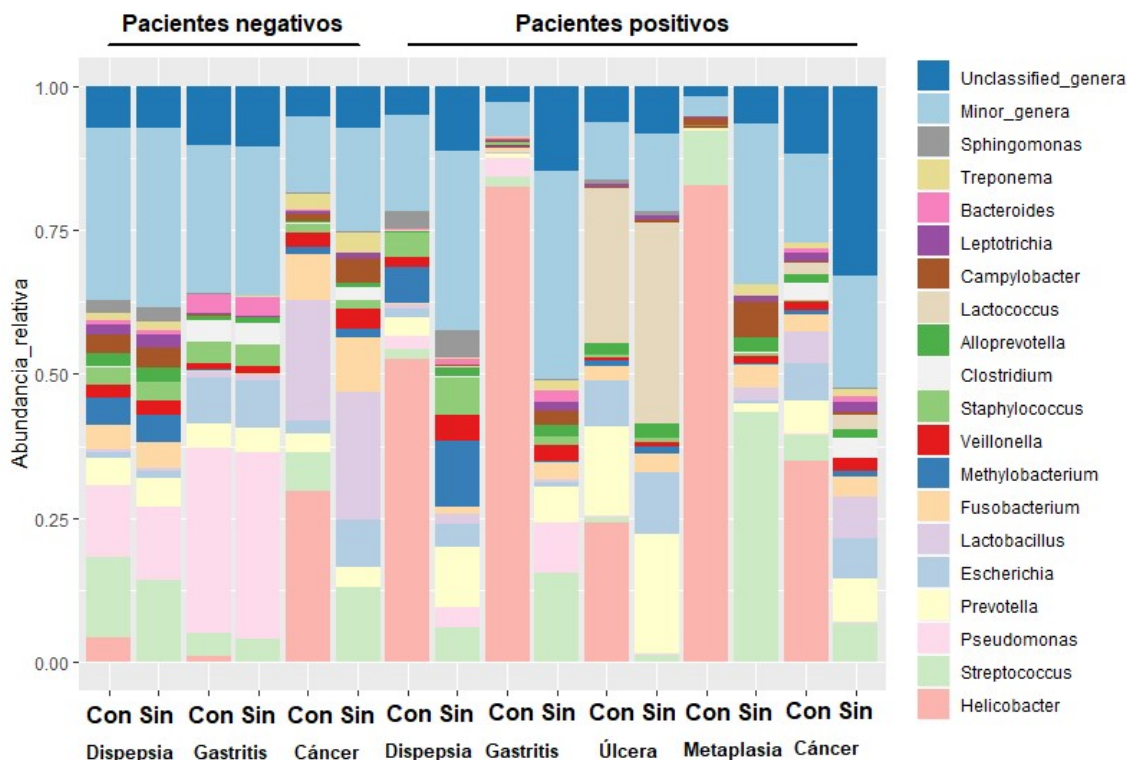


Figura 29: Comparación de la abundancia relativa de los 19 géneros mayoritarios de las muestras de biopsias de los pacientes *H. pylori* negativos y *H. pylori* positivos en función de la patología gástrica con secuencias de bacterias del género *Helicobacter* (Con) y sin las secuencias (Sin)

Entre los otros grupos se encontraron diferencias interesantes: el grupo de pacientes negativos con gastritis presentó una gran abundancia relativa del género *Pseudomonas* presentando diferencias significativas con el grupo de pacientes negativos con dispepsia o cáncer ($p= 0,019$ y $p= 0,018$ respectivamente) y con los pacientes positivos con dispepsia, gastritis o con cáncer gástrico ($p= 0,003$, $p= 0,010$ y $p <0,001$ respectivamente). Otros géneros con diferencias significativas fueron: el género *Campylobacter* con una abundancia relativa más elevada en el grupo de positivos con metaplasia 7.64 (5.43-7.82), el género *Leptotrichia* que presentó diferencias entre los grupos de gastritis de los pacientes negativos y positivos a *Helicobacter* ($p <0,05$) o el género *Lactococcus* con una abundancia relativa elevada solo en muestras de pacientes positivos con úlcera o cáncer (Fig. 29).

Resultados

Para analizar la microbiota *core* se sustrajeron las muestras de los grupos de pacientes positivos con úlcera y metaplasia al tratarse de grupos con muy poca representatividad (3 muestras).

Al analizar la microbiota *core* a nivel de género, 17 géneros cumplieron las condiciones para ser considerados dentro de esta categoría, siendo 13 de ellos exclusivos de alguna patología (Fig 30-A tabla 21). Al aumentar la abundancia relativa mínima solo cumplieron las condiciones los géneros *Streptococcus*, *Pseudomonas*, *Prevotella* y *Fusobacterium* junto a secuencias de géneros no identificados (Fig. 30-B y Tabla 22).

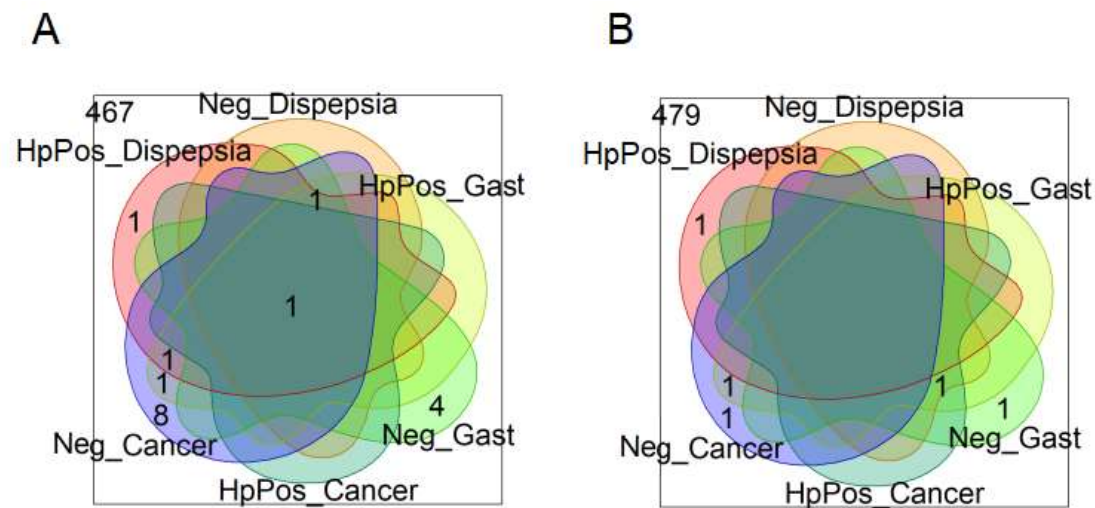


Figura 30: Diagrama de Venn de los géneros identificados en muestras de biopsia gástrica de pacientes diagnosticados positivos para infección por *H. pylori* (HpPos) y pacientes diagnosticados como negativos (Neg) agrupadas por patología gástrica (Gast=gastritis) con una abundancia relativa del 0,01% (A) y/o 1% (B). El número exterior indica el número total de géneros detectados.

Tabla 21: Géneros que cumplen la condición de estar presentes en más de un 80% de las muestras con una abundancia relativa del 0,01% en, al menos, 1 de los grupos.

	Neg Dispepsia	HpPos Dispepsia	Neg Gastritis	HpPos Gastritis	Neg Cáncer	HpPos Cáncer
Género no clasificado	+	+	+	+	+	+
<i>Streptococcus</i>	+	+	+	+	+	-
<i>Pseudomonas</i>	-	-	+	-	-	-
<i>Prevotella</i>	-	+	-	+	+	-
<i>Escherichia</i>	-	+	-	-	-	-
<i>Lactobacillus</i>	-	-	-	-	+	-
<i>Fusobacterium</i>	-	-	-	-	+	-
<i>Veillonella</i>	-	-	-	-	+	-
<i>Staphylococcus</i>	-	-	-	-	+	-
<i>Campylobacter</i>	-	-	-	-	+	-
<i>Leptotrichia</i>	-	-	-	-	+	-
<i>Bacteroides</i>	-	-	+	-	-	-
<i>Rothia</i>	-	-	-	-	+	-
<i>Actinomyces</i>	-	-	-	+	+	-
<i>Gemella</i>	-	-	-	-	+	-
<i>Acinetobacter</i>	-	-	+	-	-	-
<i>Herbaspirillum</i>	-	-	+	-	-	-

Tabla 22: Géneros que cumplen la condición de estar presentes en más de un 80% de las muestras con una abundancia relativa del 1% en, al menos, 1 de los grupos.

	Neg Dispepsia	HpPos Dispepsia	Neg Gastritis	HpPos Gastritis	Neg Cáncer	HpPos Cáncer
Género no clasificado	+	-	+	+	-	+
<i>Streptococcus</i>	-	-	-	+	+	-
<i>Pseudomonas</i>	-	-	+	-	-	-
<i>Prevotella</i>	-	+	-	-	-	-
<i>Fusobacterium</i>	-	-	-	-	+	-

5.8 Influencia del país de residencia en la patología por *H. pylori*

Para estudiar si diferentes poblaciones poseen diferentes perfiles microbianos a nivel gástrico y si estos, a su vez, pueden afectar a la evolución de las patologías asociadas a la infección por *H. pylori*, se seleccionaron los pacientes diagnosticados como positivos a infección por *H. pylori* de diferentes estudios regionales.

Se analizaron un total de 159 muestras de biopsias procedentes de pacientes diagnosticados como *Helicobacter* positivo de 6 países diferentes. En total se analizaron un total de 6.694.115 secuencias, con un rango desde 1.961 a 225.024 secuencias por muestra. La información del número de muestras por país se encuentra en la Tabla 23.

Tabla 23: Número de muestras y secuencias obtenidas de muestras de biopsia gástrica de pacientes diagnosticados positivos para infección por *H. pylori* categorizadas en función del país de residencia de los pacientes.

País	N.º muestras	Secuencias
España	16	115.985
Malasia	24	411.335
Austria	18	916.571
UK	42	3.973.836
México	34	621.996
China	25	654.392
Total	159	6.694.115

5.8.1 Diversidad alfa

El análisis de los índices de diversidad de las muestras agrupadas por países presentó diferencias estadísticamente significativas entre Reino Unido con un valor para el índice de Shannon de 0,04 (0,01-0,17), con cada uno del resto de países en comparaciones pareadas. Las muestras de Austria también presentaron diferencias significativas con respecto a las muestras de Malasia y México ($p = 0,040$ y $0,029$ respectivamente) (Fig. 31).

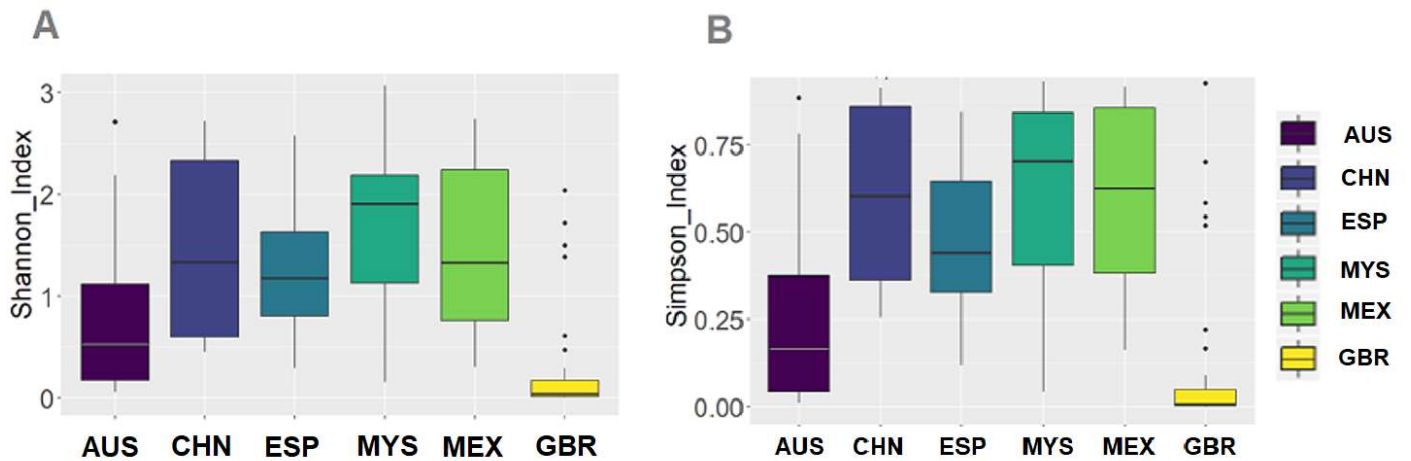


Figura 31: Diagrama de cajas mostrando la biodiversidad, medida por el índice de Shannon (A) y Simpson (B), en muestras de biopsia gástrica de los pacientes positivos a la infección por *H. pylori* en función del país de residencia de los pacientes. (AUS= Austria, CHN= China, ESP= España, MYS= Malasia, MEX=México, GBR=Reino Unido).

5.8.2 Diversidad beta

El análisis de la diversidad beta, calculada acorde a la abundancia relativa de las secuencias a nivel de ASVs (mediante el análisis de matrices de distancia de Bray-Curtis) indicaron una separación significativa de cada uno de los perfiles bacterianos dependiendo del tipo de país ($p = 0,015$; PERMANOVA), con la excepción de las muestras provenientes de Austria y Reino Unido que se agruparon sin detectarse diferencias significativas entre ambos grupos ($p = 0,051$). (Fig. 32-A). Al analizar la diversidad beta acorde a la presencia y/o ausencia de las secuencias a nivel de ASVs (mediante el análisis de las matrices de distancia de Jaccard binarias), se encontraron agrupaciones de las muestras en función del país encontrándose siempre diferencias significativas en las comparaciones pareadas ($p = 0,015$) (Fig. 32-B).

Resultados

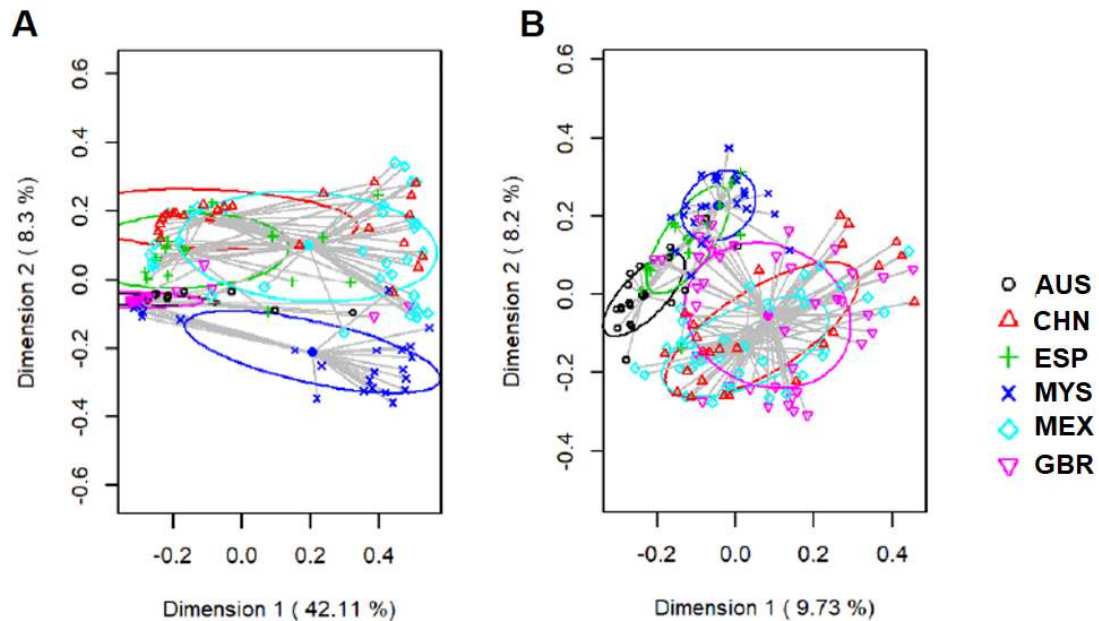


Figura 32: PCoAs de los perfiles bacterianos a nivel ASV basado en la matriz de distancia Bray-Curtis (A) y la matriz de distancia Jaccard binaria para el estudio por presencia/ausencia (B) de muestras de biopsias gástricas agrupadas en función del país de residencia de los pacientes positivos a infección por *H. pylori* (AUS= Austria, círculo negro; CHN= China, triángulo rojo; ESP= España, cruz verde; MYS= Malasia, equis azul oscuro; MEX = México, rombo azul; GBR= Reino Unido, triángulo rosa). El valor en cada eje representa el porcentaje de variación explicado por ese eje.

5.8.3 Taxonomía

Se identificaron un total de 20 filos diferentes, siendo los filos *Epsilonbacteraeota*, *Firmicutes*, *Proteobacteria* y *Bacteroidetes* los más abundantes. Hubo un efecto significativo del país de procedencia de los pacientes sobre los filos *Epsilonbacteraeota*, *Firmicutes*, *Proteobacteria* y *Bacteroidetes* y el grupo de filos minoritarios ($p < 0,001$ en todas las comparaciones) (Fig 33). Las muestras procedentes de Reino Unido y Austria presentaron, de media, las mayores abundancias relativas del filo *Epsilonbacteraeota*, presentando diferencias significativas con cada uno de los otros países ($p < 0,001$) en análisis pareados. Por otro lado, las muestras procedentes de los países de México y Malasia presentaron una abundancia relativa menor de bacterias procedentes del filo *Epsilonbacteraeota* ($p < 0,05$) y una abundancia relativa mayor del filo *Firmicutes*, siendo especialmente significativo las diferencias con Reino Unido ($p < 0,001$) (Tabla 24).

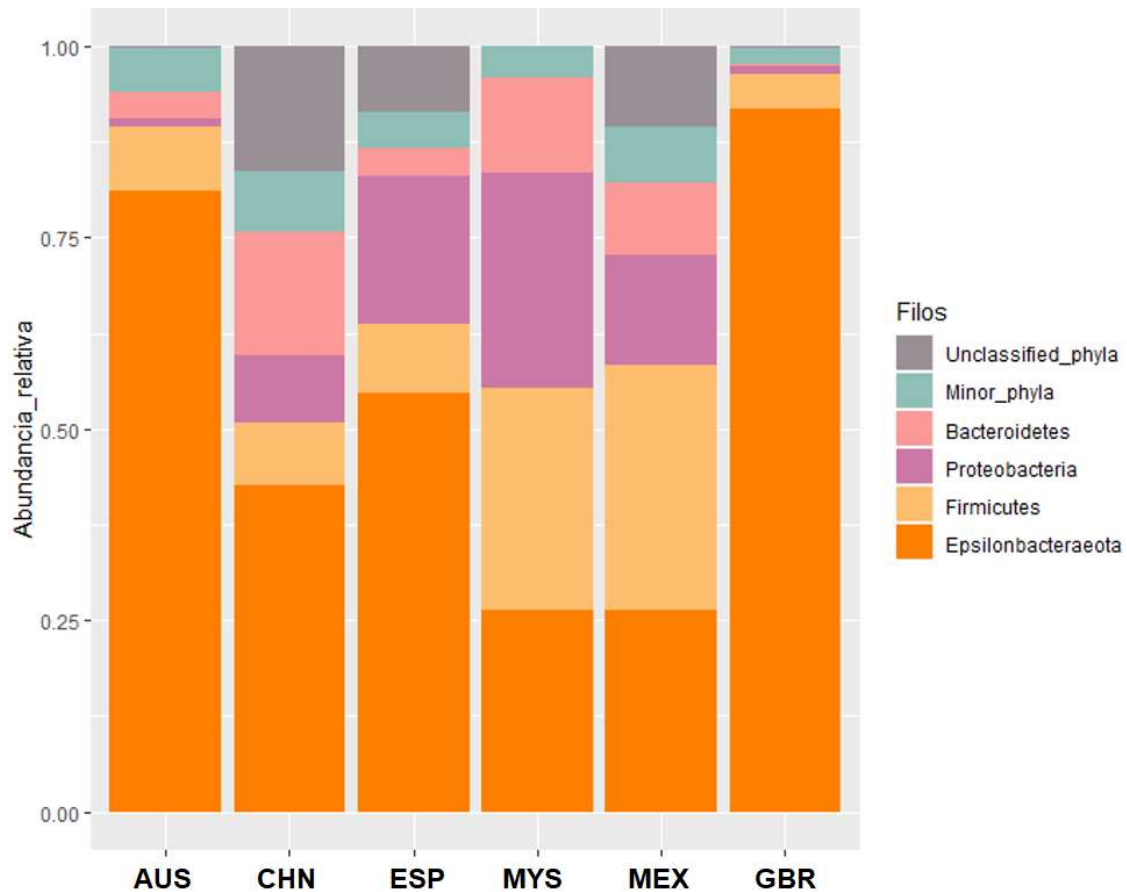


Figura 33: Comparación de la abundancia relativa de los 4 filos mayoritarios de las muestras de biopsias de los pacientes *H. pylori* positivos en función del país de residencia del paciente (AUS= Austria, CHN= China, ESP= España, MYS= Malasia, MEX=México, GBR=Reino Unido).

Al estudiar los 20 géneros mayoritarios llama la atención el rango de la mediana de la abundancia relativa del género *Helicobacter* entre las muestras de los distintos países. Por un lado, tenemos a Reino Unido y Austria con una abundancia relativa de 99.51 (97.42-99.85) y 91.3 (78.74-97.78) respectivamente, contrastando con las muestras de los países de Malasia 3.61 (<0.01-37.54) y México 2.07 (<0.01-54.03) (Tabla 24).

Resultados

Tabla 24: Abundancia relativa (%), expresada como la mediana y los rangos intercuartílicos (IQR), de los 20 géneros y 4 filios mayoritarios, de muestras de biopsia gástrica agrupadas en función al país de residencia de los pacientes diagnosticados como *H. pylori* positivos.

Filos/Géneros	Austria		China		España		Malasia		México		Reino Unido		p-valor†
	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	
<i>Epsilonbacteraeota</i>	18 (100%)	91,32 (78,91-97,78)	24 (96%)	55,03 (1,22-76,53)	16 (100%)	67,67 (22,31-81,48)	22 (91,67%)	4,67 (0,74-37,67)	27 (79,41%)	2,78 (0,11-54,03)	42 (100%)	99,59 (97,51-99,86)	<0,001
<i>Helicobacter</i>	18 (100%)	91,3 (78,74-97,78)	21 (84%)	54,83 (0,18-76,53)	16 (100%)	67,67 (22,31-81,48)	17 (70,83%)	3,61 (<0,01-37,54)	20 (58,82%)	2,07 (<0,01-54,03)	42 (100%)	99,51 (97,42-99,85)	<0,001
<i>Firmicutes</i>	18 (100%)	4,33 (0,67-12,98)	22 (88%)	4,41 (0,18-12,93)	16 (100%)	3,51 (2,43-8,3)	24 (100%)	17,47 (5,79-46,2)	33 (97,06%)	27,72 (6,06-47,26)	40 (95,24%)	0,1 (0,03-0,59)	<0,001
<i>Streptococcus</i>	18 (100%)	1,52 (0,16-4,21)	16 (64%)	0,18 (<0,01-1,84)	16 (100%)	0,28 (0,16-0,86)	21 (87,5%)	0,67 (0,19-1,85)	28 (82,35%)	1,22 (0,33-8,53)	38 (90,48%)	0,06 (0,02-0,23)	<0,001
<i>Lactobacillus</i>	4 (22,22%)	<0,01 (<0,01-<0,01)	4 (16%)	<0,01 (<0,01-<0,01)	12 (75%)	0,14 (0,01-0,51)	12 (50%)	0,01 (<0,01-3,45)	18 (52,94%)	0,27 (<0,01-5,87)	10 (23,81%)	<0,01 (<0,01-<0,01)	<0,001
<i>Lactococcus</i>	1 (5,56%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	15 (62,5%)	0,81 (<0,01-4,54)	0 (0%)	<0,01 (<0,01-<0,01)	5 (11,9%)	<0,01 (<0,01-<0,01)	<0,001
<i>Clostridium</i>	2 (11,11%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	4 (25%)	<0,01 (<0,01-<0,01)	4 (16,67%)	<0,01 (<0,01-<0,01)	8 (23,53%)	<0,01 (<0,01-<0,01)	2 (4,76%)	<0,01 (<0,01-<0,01)	0,022
<i>Veillonella</i>	16 (88,89%)	0,47 (0,05-0,92)	10 (40%)	<0,01 (<0,01-0,22)	12 (75%)	0,04 (0,01-0,26)	12 (50%)	0,01 (<0,01-2,49)	24 (70,59%)	0,34 (<0,01-2,59)	18 (42,86%)	<0,01 (<0,01-0,05)	<0,001
<i>Staphylococcus</i>	8 (44,44%)	<0,01 (<0,01-0,01)	0 (0%)	<0,01 (<0,01-<0,01)	15 (93,75%)	0,14 (0,05-0,66)	11 (45,83%)	<0,01 (<0,01-0,1)	7 (20,59%)	<0,01 (<0,01-<0,01)	20 (47,62%)	<0,01 (<0,01-0,06)	<0,001
<i>Selenomonas</i>	7 (38,89%)	<0,01 (<0,01-0,04)	11 (44%)	<0,01 (<0,01-0,32)	2 (12,5%)	<0,01 (<0,01-<0,01)	2 (8,33%)	<0,01 (<0,01-<0,01)	10 (29,41%)	<0,01 (<0,01-0,08)	5 (11,9%)	<0,01 (<0,01-<0,01)	0,002
<i>Proteobacteria</i>	18 (100%)	0,7 (0,5-1,53)	22 (88%)	0,8 (0,12-5,39)	16 (100%)	9,84 (2,88-21,42)	24 (100%)	20,53 (5,9-47,04)	24 (70,59%)	0,91 (<0,01-9,68)	32 (76,19%)	0,02 (<0,01-0,24)	<0,001
<i>Escherichia</i>	16 (88,89%)	0,1 (0,04-0,13)	3 (12%)	<0,01 (<0,01-<0,01)	14 (87,5%)	0,08 (0,02-0,42)	20 (83,33%)	0,69 (0,05-3,23)	11 (32,35%)	<0,01 (<0,01-1,98)	7 (16,67%)	<0,01 (<0,01-<0,01)	<0,001
<i>Pseudomonas</i>	13 (72,22%)	0,01 (<0,01-0,03)	11 (44%)	<0,01 (<0,01-0,15)	15 (93,75%)	4,68 (1,07-15,74)	7 (29,17%)	<0,01 (<0,01-0,03)	1 (2,94%)	<0,01 (<0,01-<0,01)	7 (16,67%)	<0,01 (<0,01-<0,01)	<0,001
<i>Methylobacterium</i>	1 (5,56%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	3 (18,75%)	<0,01 (<0,01-<0,01)	24 (100%)	2,34 (0,53-13,55)	0 (0%)	<0,01 (<0,01-<0,01)	12 (28,57%)	<0,01 (<0,01-<0,01)	<0,001
<i>Citrobacter</i>	0 (0%)	<0,01 (<0,01-<0,01)	4 (16%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	<0,001
<i>Haemophilus</i>	10 (55,56%)	0,01 (<0,01-0,12)	7 (28%)	<0,01 (<0,01-0,04)	12 (75%)	0,14 (0,01-0,67)	6 (25%)	<0,01 (<0,01-0,02)	16 (47,06%)	<0,01 (<0,01-1,04)	13 (30,95%)	<0,01 (<0,01-0,02)	0,006
<i>Bacteroidetes</i>	18 (100%)	1,64 (0,95-3,92)	21 (84%)	7,34 (0,9-25,72)	16 (100%)	1,85 (0,81-3,54)	24 (100%)	4,48 (1,11-22,43)	32 (94,12%)	2,12 (0,8-17,11)	36 (85,71%)	0,07 (0,01-0,16)	<0,001
<i>Prevotella</i>	17 (94,44%)	0,84 (0,16-3,07)	18 (72%)	1,91 (<0,01-6,14)	13 (81,25%)	0,28 (0,04-0,73)	20 (83,33%)	1,4 (0,12-13,92)	29 (85,29%)	1,07 (0,25-13,43)	31 (73,81%)	0,03 (<0,01-0,06)	<0,001
<i>Alloprevotella</i>	15 (83,33%)	0,24 (0,03-0,46)	14 (56%)	0,28 (<0,01-1)	13 (81,25%)	0,1 (0,04-0,2)	14 (58,33%)	0,03 (<0,01-0,18)	18 (52,94%)	0,06 (<0,01-0,4)	22 (52,38%)	<0,01 (<0,01-0,02)	0,007

Resultados

Continuación Tabla 24

Filos/Géneros	Austria		China		España		Malasia		México		Reino Unido		p-valor†
	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	n (%)#	Mediana (IQR)	
<i>Chryseobacterium</i>	5 (27,78%)	<0,01 (<0,01-<0,01)	6 (24%)	<0,01 (<0,01-<0,01)	5 (31,25%)	<0,01 (<0,01-0,02)	4 (16,67%)	<0,01 (<0,01-<0,01)	0 (0%)	<0,01 (<0,01-<0,01)	3 (7,14%)	<0,01 (<0,01-<0,01)	0,007
Minor_phyla	18 (100%)	0,64 (0,25-3,47)	19 (76%)	0,51 (<0,01-9,04)	16 (100%)	1,1 (0,51-1,79)	24 (100%)	1,77 (0,5-5,38)	32 (94,12%)	2,52 (0,4-9,51)	39 (92,86%)	0,11 (0,02-1,01)	0,004
<i>Fusobacterium</i>	13 (72,22%)	0,07 (<0,01-0,43)	12 (48%)	<0,01 (<0,01-2,51)	7 (43,75%)	<0,01 (<0,01-0,15)	16 (66,67%)	0,21 (<0,01-1,51)	20 (58,82%)	0,22 (<0,01-1,05)	25 (59,52%)	0,01 (<0,01-0,07)	0,11
<i>Leptotrichia</i>	11 (61,11%)	0,03 (<0,01-0,13)	11 (44%)	<0,01 (<0,01-0,2)	9 (56,25%)	0,01 (<0,01-0,08)	5 (20,83%)	<0,01 (<0,01-<0,01)	23 (67,65%)	0,17 (<0,01-3)	22 (52,38%)	<0,01 (<0,01-0,05)	0,002
Minor_genera	18 (100%)	3,21 (1,17-7,65)	25 (100%)	5,86 (0,29-24,21)	16 (100%)	5,8 (1,83-14,69)	24 (100%)	14,46 (3,85-31,58)	33 (97,06%)	8,4 (0,66-22,9)	39 (92,86%)	0,16 (0,04-1,25)	<0,001
Filo no clasificado	14 (77,78%)	0,02 (0,01-0,06)	25 (100%)	17,36 (12,34-21,32)	16 (100%)	7,59 (3,96-11,26)	2 (8,33%)	<0,01 (<0,01-<0,01)	33 (97,06%)	9,31 (4,36-16,41)	17 (40,48%)	<0,01 (<0,01-0,03)	<0,001
Género no clasificado	18 (100%)	0,18 (0,07-0,84)	25 (100%)	19,89 (14,47-22,24)	16 (100%)	11,62 (5,87-15,58)	23 (95,83%)	1,56 (0,42-4,5)	34 (100%)	10,72 (5,52-18,74)	28 (66,67%)	0,02 (<0,01-0,13)	<0,001

n (%): Número de muestras que detectaron el filo o género (frecuencia relativa de detección).

† Prueba Kruskal–Wallis de rangos con la corrección de Bonferroni.

Resultados

Otros géneros mayoritarios presentaron unos perfiles interesantes: el género *Lactococcus* fue detectado, mayoritariamente, en 15 muestras de pacientes de Malasia (62,5%) y solo en otras 6 muestras provenientes de diferentes países con una abundancia relativa muy baja. Otro género que presentó diferencias muy notables fue el género *Pseudomonas*, detectada en 15 muestras de pacientes españoles (93,75%) con una abundancia relativa de 4,68%, siendo la mediana del resto de países del 0,01% o menor.

Algo parecido pasó con el género *Methylobacterium*, muy presente en las muestras de pacientes de Malasia (se detectó en el 100% de las muestras) con una abundancia media de 2,34%, muy superiores al resto de países en los cuales, o no se detectó, o su abundancia relativa fue muy baja (Tabla 24).

Al analizar la microbiota *core* de las 159 muestras gástricas de los pacientes diagnosticados como positivos se determinaron 7 filos de 20 que cumplieron con la definición inicial de *core* (presentes en más del 80% de las muestras de alguno de los grupos con una representación mínima del 0,01% (Fig 34-A) y/o mínima del 1%(Fig. 34-B)) a parte de las secuencias no identificadas (Tabla 25).

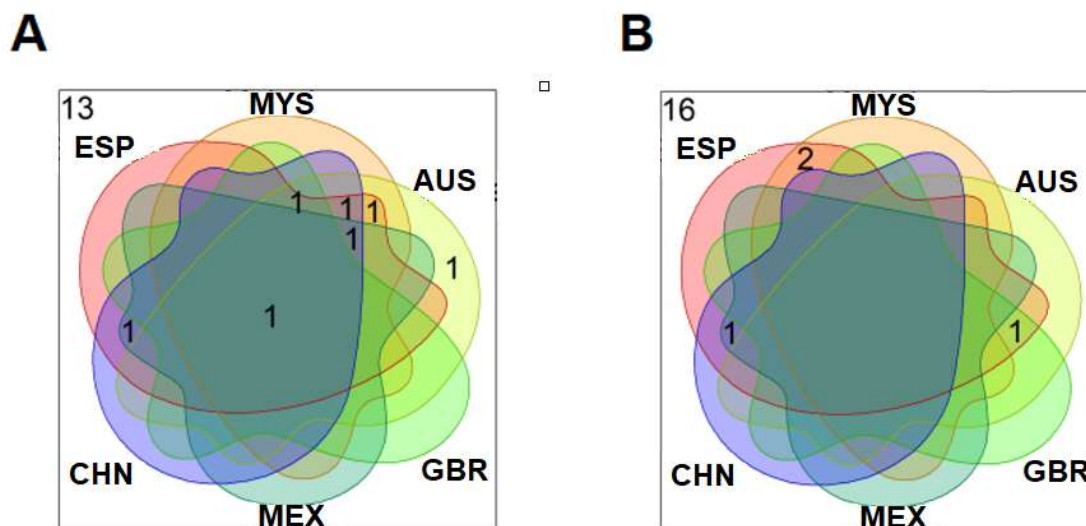


Figura 34: Filos que cumplen la condición de estar presentes en más de un 80% de las muestras con una abundancia relativa del 0,01% (A) y/o 1% (B) por cada en las muestras de biopsias de pacientes positivos agrupados por el país de residencia del paciente (AUS= Austria, CHN= China, ESP= España, MYS= Malasia, MEX=México, GBR=Reino Unido).

Tabla 25: Filos que cumplen la condición de estar presentes en más de un 80% de las muestras con una abundancia relativa del 0,01% por cada grupo.

Filos	España	Malasia	Austria	Reino Unido	Mexico	China
<i>Epsilonbacteraeota</i>	+	+	+	+	-	+
<i>Firmicutes</i>	+	+	+	+	+	+
<i>Proteobacteria</i>	+	+	+	-	-	+
<i>Bacteroidetes</i>	+	+	+	-	+	+
Filo no clasificado	+	-	-	-	+	+
<i>Fusobacteria</i>	-	-	+	-	-	-
<i>Actinobacteria</i>	+	+	+	-	-	-

Conforme aumentamos los requisitos de abundancia relativa, el número de filos va reduciéndose hasta quedar en última instancia los filos *Epsilonbacteraeota* en los grupos España, Austria y Reino Unido y *Firmicutes* y *Proteobacteria* en España y Malasia (Tabla 26).

Tabla 26: Filos que cumplen la condición de estar presentes en más de un 80% de las muestras con una abundancia relativa del 1% por cada grupo.

Filos	España	Malasia	Austria	Reino Unido	México	China
<i>Epsilonbacteraeota</i>	+	-	+	+	-	-
<i>Firmicutes</i>	+	+	-	-	-	-
<i>Proteobacteria</i>	+	+	-	-	-	-
Filo no clasificado	+	-	-	-	+	+

Al analizar la microbiota *core* al nivel de género, 20 géneros fueron determinados como géneros *core* (Tabla 27), siendo 15 de ellos exclusivos de alguno de los países (Fig. 33-A)

Al aumentar los requisitos de inclusión se ve reducida la presencia de los géneros *core* para terminar quedando solo el género *Helicobacter* de los países España, Austria y Reino Unido (Fig. 33-B).

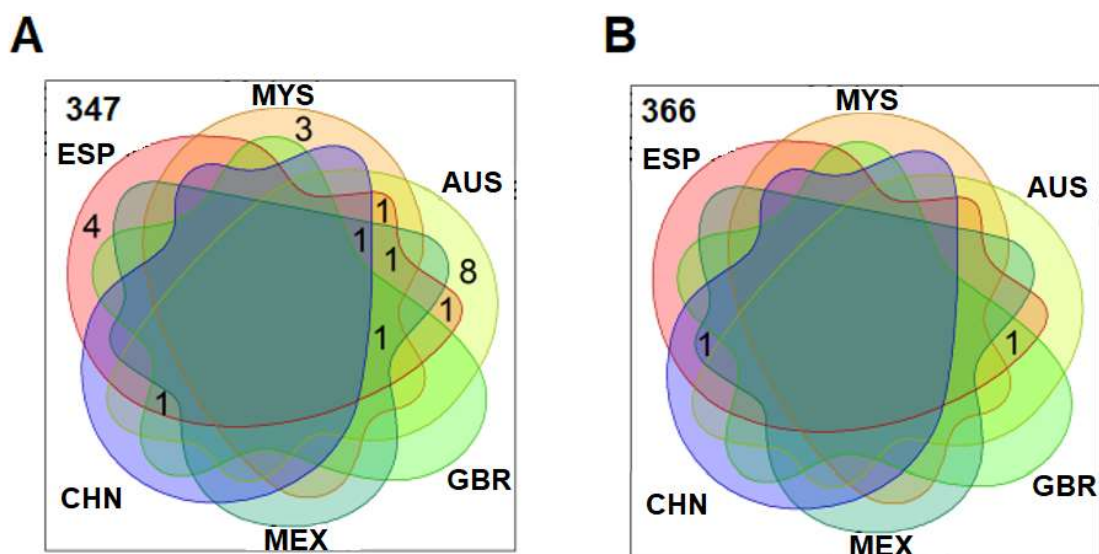


Figura 35: Géneros que cumplen la condición de estar presentes en más de un 80% de las muestras con una abundancia relativa del 0,01% (A) y/o 1% (B) en las muestras de biopsias de pacientes positivos agrupados por el país de residencia del paciente (AUS= Austria, CHN= China, ESP= España, MYS= Malasia, MEX=México, GBR=Reino Unido).

Tabla 27: Géneros que cumplen la condición de estar presentes en más de un 80% de las muestras con una abundancia relativa del 0,01% por cada grupo.

Géneros	España	Malasia	Austria	Reino Unido	México	China
<i>Helicobacter</i>	+	-	+	+	-	+
Género no clasificado	+	+	+	-	+	+
<i>Prevotella</i>	+	+	+	-	+	-
<i>Escherichia</i>	+	+	+	-	-	-
<i>Streptococcus</i>	+	+	+	+	+	-
<i>Pseudomonas</i>	+	-	-	-	-	-
<i>Veillonella</i>	-	-	+	-	-	-
<i>Methylobacterium</i>	-	+	-	-	-	-
<i>Alloprevotella</i>	+	-	+	-	-	-
<i>Staphylococcus</i>	+	-	-	-	-	-
<i>Bacteroides</i>	-	-	+	-	-	-
<i>Sphingomonas</i>	-	+	-	-	-	-
<i>Novosphingobium</i>	-	+	-	-	-	-
<i>Acinetobacter</i>	+	-	-	-	-	-
<i>Actinomyces</i>	-	-	+	-	-	-
<i>Herbaspirillum</i>	+	-	-	-	-	-
<i>Granulicatella</i>	-	-	+	-	-	-
<i>Pedobacter</i>	-	-	+	-	-	-
<i>Stenotrophomonas</i>	-	-	+	-	-	-
<i>Alkalibacterium</i>	-	-	+	-	-	-
<i>Rhodanobacter</i>	-	-	+	-	-	-

Discusión

6. Discusión

Con el aumento de la información y la mayor facilidad de acceso a los datos, hay una tendencia a la realización de estudios comparativos en los que se prima la reproducibilidad y comparación de los resultados.

En ellos, disponer de un estudio previo en el cual todas las condiciones, desde la toma de muestra, pasando por la selección de los pacientes hasta llegar al análisis final, sean conocidas, aporta mucha información de los procesos y de las posibles secciones donde se pudo introducir más variabilidad.

Por todo ello, disponer de una base de datos lo más completa y fiel posible a las muestras de estudio, contar con herramientas útiles y fácilmente aplicables para el análisis de los datos y poder distinguir qué información es de calidad e importante para nuestro estudio, adquiere un papel fundamental de cara a obtener unos resultados exactos y reproducibles.

6.1 Aplicaciones bioinformáticas a estudios proteómicos

El uso de técnicas proteicas, entre las que se encuentra el uso del MALDI TOF, no es nuevo en el estudio o diagnóstico de *H. pylori* (Winkler *et al.*, 1999). En muchos estudios se ha detectado una gran plasticidad dentro de las cepas de *H. pylori*, lo que dificulta su identificación (Ilina *et al.*, 2010).

Según las indicaciones del fabricante, los *scores* obtenidos se clasifican en 3 categorías: alta confianza de identificación (*score* > 2,00), baja confianza de identificación (*score* entre 1,99-1,70) e identificación no posible del organismo (*score* < 1,70) (Schulthess *et al.*, 2013). Con experiencia en la técnica y en el uso de la máquina se puede realizar una segunda división en los *scores*: identificación excelente (*score* > 2,00), identificación muy buena (*score* entre 1,99-1,80), identificación buena (*score* entre 1,79-1,6), identificación pobre (*score* entre 1,59-1,40) e identificación no válida (*score* < 1,4) recomendándose la identificación por otros métodos en aquellas identificaciones con valores por debajo de 1,70 (Anneloes *et al.*, 2014; Khot, Couturier, Wilson, Croft, & Fisher, 2012).

Discusión

El uso de la base de datos Bruker presenta dificultades a la hora de identificar las cepas de *H. pylori* locales. Estas dificultades ya fueron detectadas por Iliina 2009 donde se realizó un estudio por Ultraflex II MALDI-TOF/TOF (Bruker Daltonics, Germany) de diferentes cepas de *Helicobacter* con 2 cepas de referencia y 17 cepas clínicas proponiendo la necesidad de ampliar la base de datos de Bruker (E N Iliina, 2009).

En la fecha de realización del estudio actual, la base de datos Bruker incorporaba 7 cepas de *H. pylori* pertenecientes a colecciones internacionales. Los resultados obtenidos en 60 cepas aisladas de muestras clínicas obtuvieron un valor medio de 1,640, siendo un valor con el cual no podríamos realizar una identificación final.

Las ventajas de incorporar cepas locales a la base de datos Biotyper OC 3.1 es clara ya que, al utilizar solo esta base de datos, mejora el *score* hasta una media de 1,818, aumentando mucho la confianza en su identificación. Finalmente, unir las 2 bases de datos mejoró aún más la identificación, especialmente en los valores más bajos indicando que es preciso aumentar no solo el número sino la diversidad.

Finalmente, los resultados del estudio de la nueva base de datos con muestras clínicas reales para validar su implementación para el diagnóstico de infección por *H. pylori* demostró una mejor identificación coincidiendo con los valores de referencia.

La alta plasticidad que presenta *H. pylori* ha sido tema de estudio durante años y los estudios proteómicos pueden resultar clave en este tema. En el dendograma se aprecia una agrupación de las cepas locales con respecto a las cepas incorporadas por la empresa Bruker. Un mayor número de cepas de diferentes localizaciones y programas bioinformáticos que aumenten la resolución de la prueba podrían contribuir enormemente en la realización de clasificaciones intraespecie con un claro enfoque epidemiológico.

6.2 Desarrollo de un documento con instrucciones

En este trabajo se realizó y probó un documento de instrucciones con diferentes lenguajes de programación para agilizar los análisis más solicitados por diferentes grupos de investigación con el fin de llevar a cabo los estudios metataxonómicos.

Uno de los criterios para determinar qué comandos incluir ha sido reunir y predecir cuales pueden ser las necesidades de cada grupo y preparar unos modelos de tablas que sean fácilmente reproducibles. Conocer esos modelos facilitará enormemente la labor de los grupos, agilizando y flexibilizando los análisis que se podrán llevar a cabo. El uso de los formatos predefinidos de los datos es ampliamente utilizado con las plataformas semiabiertas y fundamental en las plataformas cerradas (Zakrzewski *et al.*, 2017). Es tentador para el programador/analista acotar lo máximo posible el formato de los datos de entrada para evitar posibles errores en los análisis.

En este caso se ha optado por tablas lo más simples posibles y ampliamente utilizadas, con el fin de facilitar su implementación y uso desde la toma de la muestra hasta su análisis final.

Se ha desarrollado un documento de comandos en lenguaje de programación “*Script*” con el objetivo de facilitar el análisis de grandes cantidades de información, especialmente dirigido al análisis de datos metataxonómicos de una manera modular y fácilmente reutilizable. La metodología desarrollada en esta tesis permite realizar:

- a) Normalización de los primeros resultados de tablas metataxonómicas.
- b) Resumen de los resultados por variable a analizar.
- c) Estudio de la diversidad alfa mediante los análisis de los índices de diversidad más ampliamente utilizados como son el índice de Shannon y Simpson con su representación gráfica y estadística.

Discusión

d) Estudio de la diversidad beta mediante la realización de matrices de distancia de Bray-Curtis y Jaccson binario mediante análisis PERMANOVA, estudios pareados y representación mediante PCoAs.

e) Resumen de los filos y géneros detectados, selección de los filos y géneros mayoritarios, resumen y representación gráfica y análisis estadístico mediante diagramas de cajas, mapas de calor o *heatmaps* y diagramas de Venn.

f) Representación por correlación y preparación de tablas para una correcta disposición de los resultados.

g) Módulos adicionales para análisis en profundidad basados en variables intermedias para representación de diagramas de barras apiladas, correlaciones, PCAs y cambios de colores de los gráficos.

En múltiples publicaciones se han utilizado diferentes versiones de este script (Castro *et al.* 2019, 2020; Pérez *et al.* 2019; Browne *et al.* 2019; Aparicio *et al.* 2020) y su utilización ha facilitado y agilizado el análisis de los resultados para su posterior publicación.

Uno de los inconvenientes de este documento podría radicar en la existencia de plataformas semiabiertas (Calypso, Qiime2) que aportan unas herramientas con un uso ampliamente extendido por la comunidad científica con funcionalidades muy parecidas a éste, sin la necesidad de conocer nada de programación. Estudios como el de Balvočiute y Huson (2017) demostraron que el uso de diferentes herramientas puede influir en los resultados finales (Balvočiute & Huson, 2017). Estas diferencias pueden ser mitigadas, al menos en parte, con el uso de bases de datos de referencia lo más actualizadas posibles (López-García *et al.*, 2018). A pesar de la existencia de dichas herramientas, el poder personalizar las tablas de resultados resaltando las secciones que más favorezcan a la investigación, la “facilidad” de realizar gráficos o figuras con el mismo formato y diseño, comprender cada paso en el flujo de los datos y ser uno mismo el que seleccione los filtros o estadísticos que sean oportunos en función del tipo de dato que se quiere analizar (paramétricos o no paramétricos, factores, vectores, cualitativos o cuantitativos, etc.), es su mayor fortaleza.

6.3 Estudio de la microbiota gástrica

6.3.1 Descripción de las muestras y estudios seleccionados

La creación de bases de datos públicas ha permitido que cada vez más personas puedan acceder de una manera relativamente sencilla a multitud de recursos y posibilidades que eran impensables hace un par de décadas. Esto ha favorecido que puedan llevarse a cabo nuevos estudios con un mayor tamaño muestral, que sería inviable para la mayoría de los grupos de investigación por su alto precio o intenso trabajo, permitiendo que los datos puedan ser reanalizados y estudiados bajo un nuevo ángulo o perspectiva.

A pesar de brindar múltiples posibilidades, el investigador tiene que valorar el hecho de no disponer de un control total de la adquisición, conservación o procesamiento de la muestra. Para minimizar este factor, el investigador tiene que conocer muy bien el procesamiento al que se han sometido las muestras para poder discernir si la comparación entre diferentes estudios es realmente posible y no introducir posibles sesgos a su estudio.

Por este motivo, la selección de los estudios públicos y, en particular, la selección de muestras de calidad, adquiere un papel fundamental.

En nuestro caso, el estudio partió con un total de 631 muestras gástricas de las cuales solo 331, un 52,46% del total, fueron incluidas. Los criterios elegidos no fueron triviales:

- El primer criterio a considerar fue el número mínimo de secuencias. Este criterio facilita la comparación al proporcionar unos resultados “más reales”. En todos los análisis metataxonómicos son necesarias un número mínimo de secuencias que sea representativo de la misma. En caso de que el número de secuencias sea pequeño podemos encontrar aberraciones en los resultados al normalizar el número de secuencias a su abundancia relativa, siendo las secuencias de los géneros microbianos minoritarios los que más desviaciones pueden introducir. Para arreglar este hecho, hay técnicas computacionales, como las curvas de rarefacción, que permiten trabajar con un menor número de secuencias a expensas de una pérdida

Discusión

de información reduciendo el número de secuencias utilizadas hasta un número mínimo que el investigador considere aceptable. Por lo tanto, con el objetivo de no perder información y que toda ella sea de calidad se dispusieron los criterios del número mínimo de secuencias y que aportaran la mayor información posible.

- El segundo criterio es la correcta identificación taxonómica, dejando que no haya más de un 25% de indeterminaciones en cada una de las muestras de estudio.

Entre los estudios incluidos se analizó especialmente el estudio PRJNA313391 por presentar unos perfiles microbiológicos particulares. Al realizar los estudios taxonómicos se apreciaron secuencias de géneros que, aun pudiendo ser patógenos oportunistas para pacientes inmunocomprometidos, no es habitual detectarlos.

El uso de las correlaciones para detectar contaminantes es una técnica habitual en el procesamiento de datos de secuencias (Davis, Proctor, Holmes, Relman, & Callahan, 2018).

Las contaminaciones se encuentran en una concentración inicial de ADN que, al ser amplificada y secuenciada, mantendrá la proporción entre las diferentes secuencias en cada muestra del estudio.

Inferir qué secuencias pueden ser fruto de contaminaciones es un problema complejo en los estudios metataxonómicos y nunca se puede estar seguro al 100%. En este estudio se realizó la comparación con muestras procedentes de la misma región geográfica para confirmar que esos géneros no fueran géneros propios de una región particular y se realizó un estudio de correlaciones para finalmente no poder asegurar que ese estudio haya sufrido una fuente de contaminación. Es interesante ver que muchos estudios comparativos no tengan en consideración la posibilidad de que entre los estudios comparados, posiblemente, secuencias contaminantes (de Leeuw & Duval, 2020).

6.3.2 Cálculo del tamaño muestral.

A la hora de evaluar el tamaño muestral siempre nos encontraremos con una serie de factores que condicionarán el tamaño muestral final. Estos factores pueden ser englobados en 2 grandes grupos: factores estadísticos y factores logísticos. En este trabajo los factores estadísticos fueron los más ampliamente estudiados y en los que más esfuerzos se dedicó, siendo estos factores: la selección de la hipótesis, el estudio del error tipo I o alfa, el error tipo II o beta, el poder estadístico o la variabilidad de los resultados. También se valoraron, aunque en menor medida, los factores logísticos como las limitaciones financieras o la disponibilidad de estudios que cumplieran los criterios de inclusión.

La formulación de las hipótesis para cada uno de los apartados vino basado en estudios previos y en los objetivos de esta tesis, así como el valor del error tipo I o alfa aceptando un riesgo de $\alpha < 0,05$.

Un factor más complejo de determinar es la variabilidad o dispersión esperada de los datos. Su determinación debe ser realizada previo al estudio, lo que obliga a recurrir a la literatura publicada, a realizar un estudio piloto o a la estimación de los expertos. En este trabajo se utilizaron las muestras de Llorca *et al.* (2017) ya que disponíamos de todo el conocimiento sobre la recogida, almacenamiento, procesamiento y análisis de las secuencias para el estudio de variabilidad. Tradicionalmente, el factor humano es considerado un factor de variabilidad. En aquellos estudios en los que se utilicen datos de diferentes grupos de investigación con diferentes metodologías será necesario exigir un poder estadístico más riguroso. Tradicionalmente se considera aceptable un valor de 0,80%, pero para este estudio se solicitó el 0,95% para ser más exigentes y reducir los riesgos calculados para los errores alfa y beta.

Al aplicar estas condiciones, observamos que un tamaño muestral de entre 95 y 100 muestras son un número suficiente para paliar una posible variabilidad entre las muestras.

El tamaño muestral mínimo es un indicativo de la variabilidad de los datos, pero nunca una limitación del número máximo de muestras a incluir. Contar con un

mayor número de muestras nos permitirá aportar un mayor poder estadístico en aquellas variables con una mayor desviación de los datos.

6.4 Utilización de diferentes bases de datos

La asignación taxonómica es uno de los componentes principales de los análisis de comunidades bacterianas. Esto hace que la elección de la base de datos a utilizar pueda afectar a los análisis de los resultados y a la correcta interpretación de la composición bacteriana. En esta tesis hemos evaluado los resultados obtenidos entre 2 bases de datos ampliamente utilizadas SILVA 132 (Quast *et al.*, 2013) y Greengenes 13.8 (DeSantis *et al.*, 2006) con las mismas secuencias obtenidas del estudio de Llorca *et al.* (2017) y nos hemos centrado en la categorización de los perfiles bacterianos en función de la identificación taxonómica.

Se asignaron 345 identificaciones taxonómicas de las cuales 49 fueron comunes a ambas bases de datos, 223 exclusivas de la base de datos SILVA y 171 exclusivas de la base de datos Greengenes. Este tipo de resultados concuerda con los resultados de Balvočiute y Huson (2017), en los que se percibió un aumento de asignaciones taxonómicas en la base de datos SILVA basado en un mayor número de secuencias de referencia (1.861.569 de SILVA frente a 92.684 secuencias de Greengenes) (Balvočiute & Huson, 2017).

Esta menor diversidad en la base de datos de Greengenes no se hizo patente en el estudio de la diversidad alfa mediante los índices de Shannon y Simpson. Este hecho se corresponde con los resultados obtenidos en la tabla 10 en el que se aprecia como las asignaciones taxonómicas fueron compartidas o intercambiadas, al menos, en las ASVs mayoritarias. En la tabla 11, tras la normalización a nivel de género mayoritarios, se observa que las diferencias desaparecen en la mayoría de los casos.

Este tipo de datos es especialmente llamativo en los gráficos de PCoAs en los que se aprecia la separación de las muestras en función de la base de datos utilizada, pero la distribución de las muestras en cada agrupación es parecida. De ello se podrían extraer que el mismo estudio con diferente base de datos aporta las

mismas conclusiones sin llegar a ser comparables en el mismo análisis. Es decir, las conclusiones extraídas de los diferentes estudios son válidas independientemente de la base de datos utilizada siempre y cuando se valore la evolución de la taxonomía y las bases de datos no se combinen en los estudios.

6.5 Influencia de *H. pylori*

Utilizar el estómago y en particular a *H. pylori* como modelo para el estudio de cambios en el microbioma gástrico no es algo fortuito. El estómago es un órgano en el cual, por su bajo pH, se pensaba que no podía sobrevivir ningún microorganismo. El descubrimiento de *H. pylori* supuso la rotura de este dogma abriendo la puerta a la posibilidad de un ecosistema con diferentes bacterias colonizándolo. Con la llegada de nuevas técnicas de secuenciación de ADN de alto rendimiento no solo se ha confirmado la predominancia de las bacterias del género *Helicobacter*, sino también la presencia de una gran diversidad de bacterias.

Estudios previos como el de Ana Blanco (Blanco, 2018) determinó el número normalizado de bacterias por célula humana en función del diagnóstico de infección de *H. pylori*. Detectando siempre células bacterianas, pero aumentando en 43 veces el ratio células bacterianas/ células humanas en caso de infección por *H. pylori*.

Actualmente, se sabe que el estómago es un ecosistema diverso dominado principalmente por 5 filos tanto en población pediátrica como en población adulta: *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Proteobacterias* (Llorca *et al.*, 2017; Nardone & Compare, 2015) y el reciente filo *Epsilonbacteraeota*.

De todos estos filos, *Epsilonbacteraeota* tienen una gran importancia aumentando su abundancia relativa cuando hay una infección por *H. pylori*, convirtiéndose éste en el género mayoritario de la microbiota gástrica. El problema y/o la belleza de este hecho es la interacción que tiene *H. pylori* con su medio, la pared gástrica, que, por medio de sus antígenos, va a “micromodular” el ambiente que la rodea. Esto abre un sinfín de posibilidades resumidas claramente en 2 líneas de pensamiento opuestas sin que una haya prevalecido frente a la otra: si *H. pylori*

Discusión

afecta o no afecta al microbioma gástrico (Brawner *et al.*, 2017b; Llorca *et al.*, 2017; I. Yang *et al.*, 2016).

En nuestro estudio, al comparar la diversidad alfa, se aprecia una disminución en los índices de Shannon y Simpson en los pacientes diagnosticados como *H. pylori* positivos, coincidiendo con lo obtenido en otros estudios (Bik *et al.*, 2006; Klymiuk *et al.*, 2017; Parsons *et al.*, 2017; Schulz *et al.*, 2015). Múltiples estudios correlacionan el estado de salud con la diversidad microbiana, dependiendo de la región anatómica estudiada (Ceccarani *et al.*, 2019; del Campo *et al.*, 2014; Matsuoka & Kanai, 2015; Nardone & Compare, 2015).

Mediante los estudios de la diversidad beta podemos apreciar un efecto muy significativo en los perfiles de microorganismos de las muestras de biopsia gástrica de los diferentes pacientes, en función del estado de infección por *H. pylori*. Mediante los estudios de las matrices de distancia y su representación mediante PCoAs se aprecian 2 grandes agrupaciones de las muestras de las matrices de distancia de Bray-Curtis. En estas 2 agrupaciones es interesante el papel modulador que ejerce *H. pylori* haciendo que el perfil de las muestras negativas no se junte a la agrupación de las muestras positivas, pero sí que haya muestras positivas junto a las muestras negativas. Este fenómeno se detectó en estudios como el de Llorca *et al.* (2017) en el cual se atribuyó a una baja abundancia relativa de *H. pylori* o a una baja virulencia del microorganismo. Este doble perfil se difumina al realizar los estudios de los perfiles de microorganismos mediante la presencia/ausencia.

A pesar de la clara predominancia de *H. pylori* en el ambiente gástrico, se aprecia una gran diversidad de microorganismos. En general, se detecta una gran abundancia de las bacterias pertenecientes a los filos Epsilonbacteraeota, Firmicutes, Proteobacteria, Bacteroidetes y Actinobacteriota. Estos filos siguen siendo los más abundantes tanto en la edad pediátrica como en pacientes adultos independientemente del estado de infección por *H. pylori* (Miao *et al.*, 2020). Sin embargo, se observaron diferencias significativas en su abundancia relativa en función del estado de infección por *H. pylori*. En primer lugar se aprecia el claro dominio del filo *Epsilonbacteraeota* en los pacientes *H. pylori* positivos (71,61%)

frente al de los pacientes *H. pylori* negativos (1,34%). Esta gran diferencia en la abundancia relativa de las bacterias pertenecientes a dicho filo (especialmente en las bacterias pertenecientes al género *Helicobacter*) afectará a la abundancia relativa de los otros filos, siendo siempre menor en los pacientes *H. pylori* positivos.

Ese mismo efecto se aprecia en los perfiles bacterianos a nivel de género. El género mayoritario en los pacientes *H. pylori* positivos fue *Helicobacter* detectado en el 84,28% de las muestras con una abundancia relativa de 71,49%, lo que deja al resto de la microbiota el 28,51% restante. En el caso de las muestras procedentes de pacientes *H. pylori* negativos, la distribución de los géneros fue mayor siendo *Streptococcus* el género mayoritario con una abundancia relativa de 5,43% seguido del género *Prevotella* (0,74%) y *Fusobacterium* (0,68%). Se detectaron secuencias de bacterias pertenecientes al género *Helicobacter* en el 65,62% de las muestras de los pacientes negativos con una abundancia relativa de 0,11%.

Al realizar un estudio de red de correlaciones se aprecia el entramado de relaciones de la microbiota gástrica. Si se comparan los resultados obtenidos entre las muestras de los pacientes *H. pylori* positivos y los pacientes negativos, se aprecia el efecto modulador del género *Helicobacter* en los pacientes positivos, siendo el género principal con el mayor número de correlaciones significativas. Por otro lado, en los pacientes negativos, el género *Helicobacter* queda desplazado siendo los géneros *Campylobacter*, *Fusobacterium*, *Pseudomonas* y *Prevotella* los principales nexos en estas redes de correlaciones. Estos resultados concuerdan con la hipótesis de la interacción entre *H. pylori* y el resto de la microbiota teniendo un papel fundamental en la modulación de la misma (Alarcón, Llorca, & Perez-Perez, 2017; Bik et al., 2006; Schulz et al., 2015; L. Yang et al., 2019).

El efecto que se aprecia de *H. pylori* como agente modulador del ecosistema gástrico podría deberse a un enmascaramiento dada su gran abundancia relativa, en relación con los otros géneros bacterianos presentes en las muestras gástricas. Es decir, al realizarse un estudio semicuantitativo de las secuencias del gen 16S ARNr solo podemos conocer la proporción de secuencias de cada uno de los

Discusión

géneros, por lo que si uno de ellos es muy mayoritario, se apreciará un desplazamiento de los otros géneros (disminuyendo su abundancia relativa). Este efecto puede deberse, principalmente, a 2 motivos: un desplazamiento real con alteración de los otros géneros (por ocupación del nicho ecológico, producción de compuestos biocidas, etc.) o por una sobrerrepresentación del género mayoritario, ya que al aumentar de manera muy significativa el número de bacterias de dicho género, se apreciará un efecto de disminución de los otros géneros presentes, sin necesidad de mediar ninguna otra causa (desplazamiento por masa).

Para determinar si el efecto de las bacterias del género *Helicobacter* tienen un efecto más allá de la modulación por medio de la gran abundancia relativa de sus secuencias, se realizó el mismo estudio mediante una sustracción computacional de las secuencias cuya asignación taxonómica incluía el género *Helicobacter*.

Las secuencias de *Helicobacter* comprendían el 73,71% de las secuencias teniendo un efecto directo tanto en la diversidad alfa como en la diversidad beta.

Se aprecia una tendencia a la normalización de los resultados provenientes de pacientes *H. pylori* positivos aun manteniéndose las diferencias estadísticamente significativas tanto en los índices de diversidad como en la diversidad beta al estudiar los perfiles microbianos por la presencia/ausencia de los microorganismos. Sin embargo, al estudiar los perfiles microbianos en función de la abundancia de los microorganismos se observa un solapamiento de los perfiles sin poder encontrarse diferencias significativas entre los pacientes negativos y los pacientes positivos a *H. pylori* (Fig. 20).

Esta tendencia también se aprecia cuando se realiza la erradicación de *H. pylori* por medio de tratamiento antibiótico. Estudios como el de Guo *et al* (2020) han estudiado el efecto de los antibióticos en los microorganismos gástricos en pacientes antes y después del tratamiento. En este estudio aprecian como la microbiota gástrica presenta una disbiosis en pacientes diagnosticados con infección por *H. pylori* y como, tras un tratamiento eficaz con antibióticos, la microbiota se normaliza a la de pacientes sanos (Guo *et al.*, 2020).

En este estudio solo se apreció un efecto significativo en 5 de los 19 géneros mayoritarios (*Staphylococcus*, *Pseudomonas*, *Prevotella*, *Actinomyces* y *Campylobacter*) al eliminar las secuencias identificadas como *Helicobacter* (Tabla 14 y Tabla 15).

6.6 Influencia de la patología

En este estudio hemos analizado el perfil microbiológico de muestras de pacientes diagnosticados o no con infección por *H. pylori* en función de su patología gástrica.

El papel que juegan las bacterias en el desarrollo de las patologías gástricas como la gastritis, la úlcera o el cáncer gástrico sigue siendo un tema de debate en la comunidad científica. Se ha observado que hay un sobrecrecimiento bacteriano en estómagos que sufren de hipoclorhidria, siendo *H. pylori*, junto con la toma de medicamentos, la principal causa de esta hipoclorhidria (P. R. Harris et al., 2013). Se ha investigado el papel de *H. pylori* en la progresión de las patologías gástricas, desde la gastritis hasta las úlceras o cáncer gástrico (Mégraud, Floch, Labenz, & Lehours, 2016). Sin embargo, otras bacterias gástricas también pueden interactuar con *H. pylori*, afectando tanto a su colonización inicial, por el nuevo nicho que ha producido *H. pylori* y afectando a la evolución clínica del paciente (Gunathilake et al., 2019).

Dentro de las patologías estudiadas el cáncer gástrico es el que mayor relevancia clínica presenta, siendo el sexto cáncer más común en el mundo. La mayor parte de los casos se localizan en la zona este de Asia siendo allí el segundo cáncer con mayor causa de muerte después del cáncer de pulmón (Rahman, Asombang, & Ibdah, 2014).

Estudiar el perfil microbiológico de las diferentes etapas de las patologías digestivas ha sido el objetivo de numerosos artículos, con las limitaciones de ser estudios regionales con un número limitado de muestras por cada uno de los grupos. Esto, junto a la gran variabilidad individual que nos encontramos en los perfiles microbianos, especialmente a nivel de género, debido a causas externas al estudio como son: el nivel de inflamación gástrica, los hábitos dietéticos, la toma de antibióticos, el tipo de vida rural o urbano o los estilos de vida sedentario o activo dificultan mucho determinar los efectos exclusivos de la patología en grupos de pacientes con pocos individuos (Nardone & Compare, 2015).

Estas últimas limitaciones también las encontramos en nuestro estudio, ya que los grupos de pacientes diagnosticados con infección por *H. pylori* y diagnóstico de

úlceras y metaplasia solo está compuesto por 4 y 3 pacientes respectivamente. Esto, junto a la variabilidad individual de cada paciente, hace difícil encontrar diferencias significativas en la microbiota de estos 2 grupos a nivel de género.

El estudio de las patologías gástricas separando los pacientes en función del estado de infección por *H. pylori* se relaciona con estudios como el de Li *et al.* (2009) en el que estudiaban la microbiota gástrica en pacientes con gastritis sin infección por *H. pylori* para determinar el efecto en las patologías gástricas de la microbiota gástrica no-*H.pylori* (Li *et al.*, 2009).

Aunque el ambiente gástrico de pacientes sin infección por *H. pylori* es mucho más desconocido que el ambiente gástrico de los pacientes infectados, hay varios artículos que tratan sobre la importancia de las bacterias no-*H.pylori* en las patologías gástricas, por lo que su estudio por separado merece especial atención (Yuan Hu *et al.*, 2012).

El principal resultado al estudiar estos grupos de muestras es que, en las muestras diagnosticadas como *H. pylori* negativos, se detectaron secuencias de *H. pylori* en 42 de las 64 muestras, siendo *Helicobacter* el género mayoritario dentro de las muestras de pacientes con cáncer gástrico.

El aumento de la diversidad alfa en las muestras con dispepsia frente a las muestras con gastritis, y este a su vez frente al grupo de pacientes con cáncer, coincide con la hipótesis de que una mayor diversidad bacteriana está, generalmente y dependiendo de la región anatómica (Ceccarani *et al.*, 2019), asociada a un mejor estado de salud (Clemente, Ursell, Parfrey, & Knight, 2012).

Los perfiles microbiológicos de los pacientes infectados con *H. pylori* con una patología asociada presentaron diferencias significativas en función de la patología. Se detectó una disminución de la diversidad alfa en el grupo de pacientes con patología de gastritis y del grupo de pacientes con metaplasia con respecto a los grupos con dispepsia, úlcera o cáncer. Este tipo de resultados difiere con los encontrados en los estudios de los pacientes no diagnosticados con infección por *H. pylori* y son los esperados en un proceso de infección, siendo el

Discusión

pico de mayor abundancia relativa el estado de gastritis activa y disminuyendo conforme aumentan las patologías gástricas.

El grupo de pacientes con gastritis presentó diferencias significativas en su perfil microbiano con respecto a los grupos de disbiosis y cáncer, debido a la gran abundancia relativa de bacterias pertenecientes al género *Helicobacter*. En estas muestras se aprecia una tendencia en la abundancia relativa de las bacterias del género *Helicobacter* que varía en función de la patología: aumenta su abundancia relativa desde disbiosis a gastritis, donde alcanza su abundancia relativa máxima, y disminuye otra vez en el grupo de pacientes con metaplasia y aún más en el grupo de cáncer gástrico. Esta disminución fue también detectada en el estudio de Eun *et al.* (2014) donde se vió un aumento en las bacterias del género *Streptococcus* (Eun *et al.*, 2014) que en nuestro caso no fue significativo. En nuestro estudio la disminución de las bacterias pertenecientes al género *Helicobacter* está asociada a un aumento significativo de las bacterias del género *Prevotella* y *Escherichia* en las muestras de los grupos de úlcera y cáncer gástrico.

El aumento de la abundancia relativa de bacterias pertenecientes al género *Prevotella* en las muestras con cáncer gástrico concuerda con el estudio de Karpiński (2019), en el que se las vincula con la producción de compuestos proinflamatorios y daño endotelial en la superficie oral (Karpiński, 2019). Otros estudios relacionan la presencia de *Prevotella copri* como un factor de riesgo para el desarrollo de cáncer gástrico (Gunathilake *et al.*, 2019). En el estudio de Dong *et al.* (2017) se relaciona la producción de ácido acético por las bacterias del género *Prevotella* con la inducción de la secreción de óxido nítrico (NO), molécula señalizadora para diferentes procesos fisiológicos que, en exceso o no regulada, está vinculada a diversos procesos patológicos incluidos el cáncer (Dong *et al.*, 2017).

Al quitar las secuencias procedentes de bacterias del género *Helicobacter* y realizar los estudios de diversidad alfa entre los pacientes con diferentes patologías, tanto negativos como positivos para *H. pylori*, se aprecia una normalización general de todos los grupos de pacientes positivos. Se obtuvieron valores para el índice de Shannon de aproximadamente 2,00 excepto para el grupo *H. pylori* positivos con cáncer que, aunque subió de 1,34 a 1,54, debido a

la gran abundancia relativa de secuencias no clasificadas tras la extracción de las secuencias de *H. pylori*, no se obtuvo un valor real de la diversidad de dicho grupo.

Se encontraron diferencias aisladas dentro de los filos mayoritarios, principalmente en el filo Proteobacteria entre los pacientes de los grupos de gastritis positivos y negativos a infección por *H. pylori* ($p < 0.01$). Esta diferencia viene dada por la mayor abundancia relativa de las secuencias provenientes de bacterias del género *Pseudomonas* en el grupo de gastritis negativos. Esto coincide con el estudio de Kachuei *et al.* (2020), en el cual se ha descrito que los pacientes no colonizados por *H. pylori* son más propensos a la colonización por *Pseudomonas aeruginosa* y *Staphylococcus aureus* (Kachuei, Abadi, Rahimi, & Forootan, 2020), aunque las diferencias de las bacterias del género *Staphylococcus* no fueron significativas en nuestro estudio.

6.7 Efecto de la región geográfica en la microbiota gástrica de pacientes *H. pylori* positivos

H. pylori es un factor clave para el desarrollo de las patologías gástricas siendo el responsable del 89% de los cánceres gástricos que no afectan al cardias. Sin embargo, esta patología no se distribuye de manera uniforme entre las poblaciones con un porcentaje similar de infecciones por *H. pylori*. En los estudios de Yang *et al.* (2016) detectaron que 2 poblaciones con una tasa similar de infectados por *H. pylori* tenían tasas de cáncer gástrico muy diferentes a pesar de ser poblaciones, ambas, de Colombia (Yang *et al.*, 2016). Bajo este supuesto se podría alegar que existen diferencias genéticas/raciales. Parkin (2004) investigó si éste era el único factor asociado al cáncer gástrico y encontró que las personas que procedían de zonas con alta tasa de riesgo de cáncer gástrico y emigraban a una de menor riesgo disminuían la probabilidad de sufrir este tipo de cáncer (Parkin, 2004).

El estudio de Lertpiriyapong *et al.* (2014) en ratones libres de gérmenes demostró que la microbiota gástrica inicial es fundamental para la progresión hacia cáncer gástrico. Aquellos ratones libres de microbiota gástrica demostraron ser menos propensos que los ratones con una microbiota controlada, con menor expresión de genes ligados a rutas proinflamatorias y una gastritis más leve (Lertpiriyapong *et al.*, 2014).

Para estudiar si diferentes poblaciones poseen una microbiota diferente, se seleccionaron los pacientes diagnosticados como positivos a infección por *H. pylori* de diferentes estudios regionales PRJEB21497 (Castaño-Rodríguez *et al.*, 2017), PRJEB22107 (Klymiuk *et al.*, 2017), PRJEB21104 (Parsons *et al.*, 2017) y PRJNA310127 (Yu *et al.*, 2017).

Estudios como el de Delgado *et al.* (2016) señalan una gran variabilidad en la microbiota gástrica de los pacientes de una misma región geográfica y como hemos visto, diferentes patologías gástricas presentan perfiles microbiológicos diferentes.

En este apartado se ha buscado focalizar en las diferencias de los perfiles microbianos de cada país con el fin de encontrar una microbiota gástrica *core* y poder determinar diferencias regionales en la microbiota gástrica.

En nuestro estudio se ha observado una agrupación de las muestras procedentes de pacientes *H. pylori* positivos en función del país de residencia de los pacientes. Se encontraron diferencias significativas entre las muestras de cada país, con la excepción de las muestras de los países de Reino Unido y Austria donde la abundancia relativa de bacterias del género *Helicobacter* era muy elevada y su variabilidad microbiana era menor.

Más claras fueron las diferencias encontradas al analizar las muestras en función de la presencia/ausencia de las diferentes ASVs. Por medio de este análisis se apreció una mayor agrupación de las muestras en función del país de origen, indicando que los géneros mayoritarios permanecen constantes mientras que se encuentran mayores diferencias en los géneros minoritarios.

Al estudiar más en detalle los géneros característicos de cada país, llaman la atención especialmente 3 géneros: *Lactococcus*, *Pseudomonas* y *Methylobacterium*.

Los 3 géneros fueron detectados en otros estudios y es especialmente relevante el género *Lactococcus* (Castaño-Rodríguez et al., 2017; Gunathilake et al., 2019). Este género, ha sido vinculado a un mayor riesgo de padecer cáncer gástrico (Stewart, Wu, & Chen, 2020) y en este estudio coincide que 8 de 15 muestras con *Lactococcus* tienen cáncer y 4 úlcera gástrica en la región de Malasia.

El género *Methylobacterium* ha sido descrito como un posible género contaminante durante la endoscopia (Kovaleva, Degener, & Van Der Mei, 2014) y como microbiota gástrica (Delgado et al., 2013). En nuestro estudio ha aparecido en muestras de diferentes países y estudios principalmente de muestras de pacientes con dispepsia, coincidiendo con otros estudios que lo relacionan a una microbiota gástrica sana (Liu et al., 2019).

Las bacterias del género *Pseudomonas* son comúnmente encontradas al realizar estudios y diagnósticos con muestras de biopsia gástrica, pudiendo afectar a los

Discusión

métodos diagnósticos por producir respuestas cruzadas con *H. pylori* (Johansen *et al.*, 1995) o inhibiendo su crecimiento en los cultivos *in vitro* (Lacey, Mehme, & Taylor, 1995). En el estudio de Kachuei *et al.* (2020) detectaron bacterias del género *Pseudomonas* en un porcentaje elevado de muestras (aproximadamente un 80%) siendo mayor su detección en los pacientes no infectados por *H. pylori*. En su estudio comentan que las bacterias *Pseudomonas aeruginosa* junto a bacterias de la especie *Staphylococcus aureus* podrían producir una inflamación gástrica en las primeras etapas de gastritis (Kachuei *et al.*, 2020). Las muestras procedentes de España en las que se detectaron secuencias de *Pseudomonas* proceden de pacientes pediátricos en los que la infección por *H. pylori* no llevaría muchos años. En estas muestras nos encontramos, principalmente, con infecciones tempranas de poco tiempo de desarrollo, pudiendo coincidir con las etapas que remarca el estudio de Kachuei *et al.* (2020).

6.8 Limitaciones del estudio

A lo largo del presente estudio se ha intentado dar una visión lo más neutral posible tanto de la metodología como de sus limitaciones técnicas. Una vez discutidos todos los puntos y resultados de esta tesis podría ser interesante remarcar aquellas limitaciones que consideramos más importantes para intentar resaltar los problemas más comunes que nos encontramos durante la investigación.

Además de las limitaciones comunes a la mayoría de los estudios metataxonómicos previstas en su diseño remarcadas en el punto “2.3.4 Etapas en el estudio de la microbiota y sus limitaciones”, han ido surgiendo otras derivadas de los estudios seleccionados y de los análisis utilizados:

- Bajo tamaño muestral: a pesar de medir y calcular el tamaño muestral mínimo para el presente estudio, el objetivo inicial era contar con un tamaño muestral mucho mayor. De un total de 631 muestras procedentes de 6 proyectos diferentes solo 331 cumplieron los requisitos de inclusión marcados en el presente estudio. Finalmente, y tras el filtrado del proyecto PRJNA313391 solo 223 muestras fueron finalmente incluidas en los análisis finales.
- Falta de datos disponibles y fiables: esta limitación enlaza con la anterior. El número de estudios metataxonómicos en muestras de biopsias gástricas es limitado y, debido a las limitaciones generales de este tipo de estudios, los datos generan poca confianza. Es necesario ser muy estricto con la selección de las muestras a incluir en este tipo de estudios.
- Falta de un consenso en los análisis de la microbiota: a pesar del gran número de artículos que se están publicando sobre análisis metataxonómicos aún queda un gran camino en la normalización de los términos. Un claro ejemplo de ello es la definición de *core*, quedando a la discreción del investigador determinar unos límites para su definición, lo que puede generar confusión al comparar diferentes estudios metataxonómicos.

Discusión

- No poder profundizar más en los análisis de factores: esta limitación es una limitación metodológica general pero se hace más patente cuanto más se profundiza en los análisis. Pueden surgir preguntas y posibles líneas de investigación interesantes que, aunque pueden ser valoradas, tienen que partir de un nuevo estudio, repensado y con un nuevo planteamiento de la idea inicial para evitar posibles sesgos.

- Acceso a los estudios limitado: actualmente la mayoría de las revistas están solicitando a los investigadores que faciliten el acceso a sus secuencias a toda la comunidad científica. A pesar de ello, no todos los estudios aportan dicho acceso lo que limita al investigador a posibles estudios.

- Por último, es importante remarcar que todos los pacientes incorporados a este estudio son pacientes con alguna patología, cuyo diagnóstico requería de una endoscopia gástrica. El grupo ideal de pacientes control sería aquél compuesto por personas sanas sin ninguna patología digestiva. A pesar de ello, el hecho de tomar una biopsia gástrica de pacientes sanos, sin causa médica, para el estudio de la microbiota gástrica, choca frontalmente con la ética ligada a cualquier estudio.

Todas estas limitaciones pueden ser resumidas en una: incertidumbre. La falta de control de las muestras, la dificultad de trabajar con biopsias gástricas y la falta de consenso para el análisis de la microbiota hace imprescindible tomar los datos con cautela y conocimiento de dichas limitaciones.

6.9 Discusión final

El estudio de la microbiota humana se ha vuelto muy popular en los últimos años conforme más sabemos del papel fundamental que tiene en nuestra salud. Con el aumento de la información de la que disponemos la necesidad de procesar todos esos datos nos ha llevado a diseñar una gran colección de herramientas. Muchas de ellas han sido creadas con el objetivo de poder normalizar los estudios y disminuir la variabilidad asociada a utilizar diferentes herramientas, pero hoy en día no disponemos de un modelo estandarizado para poder llevar a cabo dichos estudios. Así pues, es el propio investigador el que debe conocer todas las ventajas y limitaciones de las herramientas que planea utilizar. Esta labor es, desde luego, ardua y laboriosa y no es de extrañar que muchos investigadores contraten los servicios de diferentes empresas o realicen colaboraciones con grupos que aporten su experiencia. Con todo ello, y a pesar de que el objetivo ideal es aportar el conocimiento a la comunidad por medio de la divulgación científica, están apareciendo intereses comerciales en torno a los análisis metataxonómicos, que junto a su publicidad y, por supuesto, legítimo interés comercial pueden confundir a los investigadores o presuponer unos resultados imposibles.

Ser capaz de discernir qué herramientas son las precisas en cada momento de la investigación y poder modificarlas para poder obtener unos resultados reales se ha vuelto fundamental para todos los investigadores. Con ello se reducen costes asociados a los análisis y preparan a los grupos para realizar los estudios de manera correcta. Este estudio se ha centrado en la importancia de los análisis metataxonómicos que tratan sobre la salud humana, pero no hay que olvidar que otras muchas áreas están también beneficiándose de este tipo de tecnología. Así pues, múltiples áreas de carácter industrial como son los alimentos, el cuidado animal o incluso en el cuidado medio ambiental son también partícipes necesarios del desarrollo de estas técnicas por lo que no es de extrañar que esta relación entre la bioinformática y el análisis de grandes cantidades de datos o “big data” sea día a día más estrecha.

Discusión

En nuestro estudio, y gracias a la herramienta que hemos preparado, se ha visto la presencia de una microbiota gástrica compleja y variada. Es indudable la importancia que tiene *H. pylori*, siendo la principal colonizadora del ambiente gástrico y también su papel en el desarrollo de patologías gástricas. A lo largo de este estudio hemos visto que la microbiota gástrica está compuesta, a su vez, de bacterias de otros filos bacterianos como son el filo Proteobacteria, Firmicutes, Bacteroidetes, Actinobacteria y/o Fusobacteria, junto a unos géneros representativos de cada uno de esos filos: *Streptococcus*, *Pseudomonas*, *Prevotella*, *Fusobacterium*, *Veillonella*, *Campylobacter*, *Leptotrichia*, *Bacteroides*, *Actinomyces*, etc.

Con las técnicas utilizadas en este estudio no se ha podido determinar si las bacterias que han sido identificadas eran bacterias que realmente habían colonizado el estómago o eran bacterias en transición provenientes del ambiente oral, del tracto respiratorio o formaban parte de la microbiota intestinal. En este tipo de estudios presenta un mayor interés discernir si esas bacterias, independientemente de si son residentes o transitorias, tienen un efecto en la protección o en la progresión de las patologías digestivas. Hemos visto que *H. pylori* está muy presente en casi todas las muestras gástricas, incluso en pacientes diagnosticados como negativos para infección por *H. pylori*. pero con una abundancia relativa menor a la de la mayoría de las muestras de los pacientes positivos. Con ello se abre la posibilidad de que otras bacterias no *H. pylori* puedan modular el ambiente gástrico y puedan inducir o facilitar diferentes patologías mediante la producción de compuestos que medien en la inflamación de la pared del estómago, produzcan compuestos carcinogénicos o, al contrario, protejan de dichos compuestos mediante su metabolismo o por competitividad. Para poder realizar todas estas pruebas sería necesario estudiar todos las rutas metabólicas de las bacterias implicadas, realizar estudios genéticos, estudiar la dinámica de las bacterias frente al ambiente gástrico y ver la respuesta de las células humanas frente a esas bacterias. El papel de los estudios metataxonómicos es dar a conocer las posibles bacterias que pueden tener un papel relevante en dicho proceso y poner la luz sobre que bacterias precisan de una mayor atención.

Este tipo de estudios nos aportarán nuevos tipos de tratamientos, ya sea en forma de tratamientos con antibióticos más dirigidos a grupos bacterianos o mediante la modulación de la microbiota por medio de otros microorganismos con carácter probiótico.

Bajo esta última premisa se están llevando a cabo numerosos estudios en animales (Nagano et al., 2019; Zhu & Liu, 2017) donde los probióticos demostraron ciertos efectos positivos en la prevención de la carcinogénesis influyendo en la colonización de esta bacteria y atenuando la inflamación gástrica inducida por *H. pylori* (Isobe et al., 2012; Kabir et al., 1997). Sin embargo, sería necesario realizar más investigaciones, especialmente en humanos, para confirmarlo.

Conclusiones

7. Conclusiones

1. Las técnicas proteómicas como el MALDI-TOF son buenas herramientas para la identificación de colonias de *H. pylori* en la práctica clínica si se tiene en consideración la alta plasticidad de esta especie, pudiendo ser necesario preparar una base de datos de cepas de cada área geográfica.
2. La creación de bases de datos locales mejora la identificación de cepas de *H. pylori* en la rutina clínica aumentando significativamente el número de cepas identificadas con un valor de *score* superior a 1,7.
3. Con el desarrollo de la herramienta bioinformática, en esta tesis hemos podido reducir tiempos y gastos asociados a toda investigación metataxonómica facilitando el análisis de grandes volúmenes de datos, de una manera modular y reutilizable.
4. Las análisis metataxonómicos muestran una gran variabilidad en los perfiles de microbiota gástrica dentro de los grupos de pacientes, siendo necesarias al menos 100 muestras para obtener un poder estadístico elevado y aumentar, en la medida de lo posible, la resolución de las pruebas estadísticas y reducir la probabilidad de error tipo II.
5. El uso de diferentes bases de datos de referencia afecta de manera directa sobre los resultados obtenidos. La incorporación de nuevas secuencias de referencia y la actualización de las asignaciones taxonómicas hacen que la mejor opción sea el uso de las bases de datos más actualizadas.
6. La comparación con estudios previos que hayan utilizado otra base de datos es posible, siempre y cuando se conozca las limitaciones de ambas bases de datos, ya que los resultados podrían estar sesgados debido a reasignaciones taxonómicas o mejoras en las secuencias de referencia.
7. No todo lo que está publicado en los repositorios de datos está libre de errores. Es fundamental comprobar el material que se ha obtenido por estas fuentes previamente al análisis y a sacar conclusiones.

Conclusiones

8. El análisis de correlaciones mostró que *H. pylori* tiene un efecto directo en la diversidad de la microbiota gástrica, teniendo un papel modulador principal en aquellos pacientes que hayan sido infectados por esta bacteria. En los pacientes negativos *H. pylori* pierde el papel modulador, obteniéndolo bacterias de los géneros *Prevotella* y *Pseudomonas*.

9. Al eliminar las secuencias de las bacterias del género *Helicobacter* de los análisis, se aprecia una normalización significativa de los perfiles bacterianos entre los pacientes positivos y negativos para la infección por *H. pylori*.

10. No podemos determinar si las patologías gástricas tienen un efecto en la microbiota o la microbiota tiene efecto en las patologías, pero ambas están claramente relacionadas y la modulación de una puede tener efecto sobre la otra.

11. *H. pylori* es el principal agente patológico en las patologías digestivas, pero también se han observado otras bacterias asociadas a diferentes patologías como: bacterias del género *Pseudomonas* en gastritis, los géneros *Escherichia* y *Fusobacterium* en pacientes de cáncer gástrico (todos ellos negativos a *H. pylori*) y/o bacterias del género *Prevotella* en pacientes con cáncer positivos a *H. pylori*.

12. Los perfiles microbianos de las muestras gástricas encontradas en este estudio parecen tener un comportamiento ligeramente diferente en función de la región. A pesar de ello se aprecia una microbiota común caracterizada, especialmente, por bacterias del género *Helicobacter*, *Prevotella*, *Escherichia* y *Streptococcus*.

Bibliografía

8. Bibliografía

- Alarcón, T., Llorca, L., & Perez-Perez, G. (2017). Impact of the Microbiota and Gastric Disease Development by *Helicobacter pylori*. In N. Tegtmeier & S. Backert (Eds.), *Curr Top Microbiol Immunol* (Vol. 400, pp. 253–275). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-50520-6_11
- Alba, C., Blanco, A., & Alarcón, T. (2017). Antibiotic resistance in *Helicobacter pylori*. *Current Opinion in Infectious Diseases*, 30(5), 489–497. <https://doi.org/10.1097/QCO.0000000000000396>
- Allali, I., Arnold, J. W., Roach, J., Cadenas, M. B., Butz, N., Hassan, H. M., ... Azcarate-Peril, M. A. (2017). A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiology*, 17(1), 1–16. <https://doi.org/10.1186/s12866-017-1101-8>
- Alvarado Capó, Y., Portal González, N., García Aguila, L., Freire Seijo, M., Martínez, Y., & Kosky, R. (2007). *Ochrobactrum anthropi*, contaminants of in vitro culture of sugarcane cells and tissues. *Biotecnología Vegetal*, 7(4), 211–214.
- Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. (2012). {FastQC}. Babraham, UK.
- Anneloes, V., Kolecka, A., Khayhan, K., Theelen, B., Groenewald, M., Boel, E., Wunderink, H. F. (2014). Interlaboratory comparison of sample preparation methods, database expansions, and cutoff values for identification of yeasts by matrix-assisted laser desorption ionization-time of flight mass spectrometry using a yeast test panel. *Journal of Clinical Microbiology*, 52(8), 3023–3029. <https://doi.org/10.1128/JCM.00563-14>
- Aparicio, M., Alba, C., Rodríguez, J. M., & Fernández, L. (2020). Microbiological and immunological markers in milk and infant feces for common gastrointestinal disorders: A pilot study. *Nutrients*, 12(3). <https://doi.org/10.3390/nu12030634>
- Aras, R. A., Kang, J., Tschumi, A. I., Harasaki, Y., & Blaser, M. J. (2003). Extensive repetitive DNA facilitates prokaryotic genome plasticity.

Bibliografía

- Proceedings of the National Academy of Sciences*, 100(23), 13579–13584.
<https://doi.org/10.1073/pnas.1735481100>
- Atherton, J. C., & Blaser, M. J. (2009). Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. *Journal of Clinical Investigation*, 119(9), 2475–2487. <https://doi.org/10.1172/JCI38605>
- Bache, S. M., & Wickham, H. (2014). magrittr: A Forward-Pipe Operator for R. Retrieved from <https://cran.r-project.org/package=magrittr>
- Balvočiute, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics*, 18(Suppl 2), 1–8. <https://doi.org/10.1186/s12864-017-3501-4>
- Banerjee, S., Shen, B., Nelson, D. B., Lichtenstein, D. R., Baron, T. H., Anderson, M. A., ... Stewart, L. E. (2008). Infection control during GI endoscopy. *Gastrointestinal Endoscopy*, 67(6), 781–790. <https://doi.org/10.1016/j.gie.2008.01.027>
- Beausejour, C. M., & Campisi, J. (2006). Ageing: Balancing regeneration and cancer. *Nature*, 443(7110), 404–405. <https://doi.org/10.1038/nature05221>
- Belkaid, Y. and T. H. (2015). Role of the Microbiota in Immunity and inflammation Yasmine. *Cell*, 157(1), 121–141. <https://doi.org/10.1016/j.cell.2014.03.011>.Role
- Benavides-Ward, A., Vasquez-Achaya, F., Silva-Caso, W., Aguilar-Luis, M. A., Mazulis, F., Urteaga, N., & Del Valle-Mendoza, J. (2018). *Helicobacter pylori* and its relationship with variations of gut microbiota in asymptomatic children between 6 and 12 years. *BMC Research Notes*, 11(1), 1–7. <https://doi.org/10.1186/s13104-018-3565-5>
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., ... Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53–59. <https://doi.org/10.1038/nature07517>
- Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F., ... Relman, D. A. (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proceedings of the National Academy of Sciences of*

- the United States of America*, 103(3), 732–737.
<https://doi.org/10.1073/pnas.0506655103>
- Blanco, A. (2018). APLICACIÓN DE LAS TÉCNICAS DE SECUENCIACIÓN MASIVA AL ESTUDIO DE LA INFECCIÓN POR *Helicobacter pylori*. Madrid, España.
- Blaser, M. J. (2005, February). An endangered species in the stomach. *Scientific American*. <https://doi.org/10.1038/scientificamerican0205-38>
- Blaser, M. J., & Atherton, J. C. (2004). *Helicobacter pylori* persistence: Biology and disease. *Journal of Clinical Investigation*, 113(3), 321–333.
<https://doi.org/10.1172/JCI20925>
- Bokulich, N. A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., ... Blaser, M. J. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med*, 8(343), 343ra82.
<https://doi.org/10.1126/scitranslmed.aad7121>
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- Brawner, K. M., Kumar, R., Serrano, C. A., Ptacek, T., Lefkowitz, E., Morrow, C. D., ... Smith, P. D. (2017a). *Helicobacter pylori* infection is associated with an altered gastric microbiota in children. *Mucosal Immunology*, 10(5), 1169–1177. <https://doi.org/10.1038/mi.2016.131>
- Brawner, K. M., Kumar, R., Serrano, C. A., Ptacek, T., Lefkowitz, E., Morrow, C. D., ... Smith, P. D. (2017b). *Helicobacter pylori* infection is associated with an altered gastric microbiota in children. *Mucosal Immunology*, 10(5), 1169–1177. <https://doi.org/10.1038/mi.2016.131>
- Browne, P. D., Aparicio, M., Alba, C., Hechler, C., Beijers, R., Rodríguez, J. M., ... de Weerth, C. (2019). Human milk microbiome and maternal postnatal psychosocial distress. *Frontiers in Microbiology*, 10(OCT), 1–16.
<https://doi.org/10.3389/fmicb.2019.02333>
- Buffie, C. G., & Pamer, E. G. (2013). Microbiota-mediated colonization

Bibliografía

- resistance against intestinal pathogens. *Nature Reviews Immunology*, 13(11), 790–801. <https://doi.org/10.1038/nri3535>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Castaño-Rodríguez, N., Goh, K. L., Fock, K. M., Mitchell, H. M., & Kaakoush, N. O. (2017). Dysbiosis of the microbiome in gastric carcinogenesis. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-16289-2>
- Castro, I., Alba, C., Aparicio, M., Arroyo, R., Jiménez, L., Fernández, L., ... Rodríguez, J. M. (2019). Metataxonomic and immunological analysis of milk from ewes with or without a history of mastitis. *Journal of Dairy Science*, 102(10), 9298–9311. <https://doi.org/10.3168/jds.2019-16403>
- Castro, I., Alba, C., Fernández, L., García, A. J., & Rodríguez, J. M. (2020). Culture-dependent and metataxonomic analysis of milk from red deer (*Cervus elaphus*). *International Dairy Journal*, 102. <https://doi.org/10.1016/j.idairyj.2019.104610>
- Ceccarani, C., Foschi, C., Parolin, C., D'Antuono, A., Gaspari, V., Consolandi, C., ... Marangoni, A. (2019). Diversity of vaginal microbiome and metabolome during genital infections. *Scientific Reports*, 9(1), 1–12. <https://doi.org/10.1038/s41598-019-50410-x>
- Chey, W. D., Leontiadis, G. I., Howden, C. W., & Moss, S. F. (2017). ACG Clinical Guideline: Treatment of *Helicobacter pylori* Infection. *American Journal of Gastroenterology*, 112(2), 212–238. <https://doi.org/10.1038/ajg.2016.563>
- Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2016). GenBank. *Nucleic Acids Research*, 44(D1), D67–D72. <https://doi.org/10.1093/nar/gkv1276>
- Clemente, J. C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell*, 148(6), 1258–1270. <https://doi.org/10.1016/j.cell.2012.01.035>

- Clooney, A. G., Fouhy, F., Sleator, R. D., O'Driscoll, A., Stanton, C., Cotter, P. D., & Claesson, M. J. (2016). Comparing apples and oranges?: Next generation sequencing and its impact on microbiome analysis. *PLoS ONE*, *11*(2), 1–16. <https://doi.org/10.1371/journal.pone.0148028>
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., ... Tiedje, J. M. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, *42*(D1), 633–642. <https://doi.org/10.1093/nar/gkt1244>
- Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M., & Relman, D. A. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science*, *336*(6086), 1255–1262. <https://doi.org/10.1126/science.1224203>
- Cowan, S. (1971). Sense and nonsense in bacterial taxonomy. *Journal of General Microbiology*, *67*(1), 1–8. <https://doi.org/10.1007/s00213-008-1291-8>
- Dahlberg, J., Sun, L., Waller, K. P., Ostensson, K., McGuire, M., Agenas, S., & Dicksved, J. (2019). Microbiota data from low biomass milk samples is markedly affected by laboratory and reagent contamination. *PLoS ONE*, *14*(6), 1–17. <https://doi.org/10.1371/journal.pone.0218257>
- Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*(1), 1–14. <https://doi.org/10.1186/s40168-018-0605-2>
- de Leeuw, M. A., & Duval, M. X. (2020). The Presence of Periodontal Pathogens in Gastric Cancer. *Exploratory Research and Hypothesis in Medicine*, *000*(000), 1–10. <https://doi.org/10.14218/ERHM.2020.00024>
- del Campo-Moreno, R., Alarcón-Cavero, T., D'Auria, G., Delgado-Palacio, S., & Ferrer-Martínez, M. (2018). Microbiota and Human Health: characterization techniques and transference. *Enfermedades Infecciosas y Microbiología Clínica*, *36*(4), 241–245. <https://doi.org/10.1016/j.eimc.2017.02.007>
- del Campo, R., Garriga, M., Pérez-Aragón, A., Guallarte, P., Lamas, A., Máiz, L., ... Suárez, L. (2014). Improvement of digestive health and reduction in

Bibliografía

- proteobacterial populations in the gut microbiota of cystic fibrosis patients using a *Lactobacillus reuteri* probiotic preparation: A double blind prospective study. *Journal of Cystic Fibrosis*, 13(6), 716–722.
<https://doi.org/10.1016/j.jcf.2014.02.007>
- Delgado, S., Cabrera-Rubio, R., Mira, A., Suárez, A., & Mayo, B. (2013). Microbiological Survey of the Human Gastric Ecosystem Using Culturing and Pyrosequencing Methods. *Microbial Ecology*, 65(3), 763–772.
<https://doi.org/10.1007/s00248-013-0192-5>
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7), 5069–5072.
<https://doi.org/10.1128/AEM.03006-05>
- Dong, T., Feng, Q., Liu, F., Chang, L. K., Zhou, X., Han, M., ... Liu, S. (2017). Alteration of stomach microbiota compositions in the progression of gastritis induces nitric oxide in gastric cell. *Experimental and Therapeutic Medicine*, 13(6), 2793–2800. <https://doi.org/10.3892/etm.2017.4373>
- Dusa, A. (2018). venn: Draw Venn Diagrams. Retrieved from <https://cran.r-project.org/package=venn>
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., ... Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science (New York, N.Y.)*, 308(5728), 1635–1638.
<https://doi.org/10.1126/science.11110591>
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). {qgraph}: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, 48(4), 1–18. Retrieved from <http://www.jstatsoft.org/v48/i04/>
- Eun, C. S., Kim, B. K., Han, D. S., Kim, S. Y., Kim, K. M., Choi, B. Y., ... Kim, J. F. (2014). Differences in gastric mucosal microbiota profiling in patients with chronic gastritis, intestinal metaplasia, and gastric cancer using pyrosequencing methods. *Helicobacter*, 19(6), 407–416.
<https://doi.org/10.1111/hel.12145>

- Favier, C. F., De Vos, W. M., & Akkermans, A. D. L. (2003). Development of bacterial and bifidobacterial communities in feces of newborn babies. *Anaerobe*, 9(5), 219–229. <https://doi.org/10.1016/j.anaerobe.2003.07.001>
- Ferreira, R. M., Pereira-Marques, J., Pinto-Ribeiro, I., Costa, J. L., Carneiro, F., Machado, J. C., & Figueiredo, C. (2018). Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. *Gut*, 67(2), 226–236. <https://doi.org/10.1136/gutjnl-2017-314205>
- Garnier, S. (2018). viridis: Default Color Maps from “matplotlib.” Retrieved from <https://cran.r-project.org/package=viridis>
- Garrity, G., Bell, J., & Lilburn, T. (2006). “Family II. *Helicobacteraceae* fam. nov,” in *Bergey’s Manual of Systematic Bacteriology, (The Proteobacteria), Part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria)*. (N. S. New York, Ed.) (2nd ed.). New York, NY: D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity.
- Garza-González, E., Perez-Perez, G. I., Maldonado-Garza, H. J., & Bosques-Padilla, F. J. (2014). A review of *Helicobacter pylori* diagnosis, treatment, and methods to detect eradication. *World Journal of Gastroenterology*, 20(6), 1438–1449. <https://doi.org/10.3748/wjg.v20.i6.1438>
- Gomez-Arango, L. F., Barrett, H. L., McIntyre, H. D., Callaway, L. K., Morrison, M., & Nitert, M. D. (2017). Contributions of the maternal oral and gut microbiome to placental microbial colonization in overweight and obese pregnant women. *Scientific Reports*, 7(1), 1–10. <https://doi.org/10.1038/s41598-017-03066-4>
- Goodwin, C. S., Armstrong, J. A., & Marshall, B. J. (1986). *Campylobacter pyloridis*, gastritis, and peptic ulceration. *Journal of Clinical Pathology*, 39(4), 353–365. <https://doi.org/10.1136/jcp.39.4.353>
- Graham, D. Y., & Miftahussurur, M. (2018). *Helicobacter pylori* urease for diagnosis of *Helicobacter pylori* infection: A mini review. *Journal of Advanced Research*, 13, 51–57. <https://doi.org/10.1016/j.jare.2018.01.006>
- Gressmann, H., Linz, B., Ghai, R., Pleissner, K. P., Schlapbach, R., Yamaoka, Y., ... Achtman, M. (2005). Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genetics*, 1(4), 0419–0428.

Bibliografía

<https://doi.org/10.1371/journal.pgen.0010043>

- Gunathilake, M. N., Lee, J., Choi, I. J., Kim, Y. Il, Ahn, Y., Park, C., & Kim, J. (2019). Association between the relative abundance of gastric microbiota and the risk of gastric cancer: a case-control study. *Scientific Reports*, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-50054-x>
- Guo, Y., Zhang, Y., Gerhard, M., Gao, J.-J., Mejias-Luque, R., Zhang, L., ... Pan, K.-F. (2020). Effect of *Helicobacter pylori* on gastrointestinal microbiota: a population-based study in Linqu, a high-risk area of gastric cancer. *Gut*, 69(9), 1598–1607. <https://doi.org/10.1136/gutjnl-2019-319696>
- Harris, P. R., Serrano, C. A., Villagrán, A., Walker, M. M., Thomson, M., Duarte, I., ... Crabtree, J. E. (2013). *Helicobacter pylori*-associated hypochlorhydria in children, and development of iron deficiency. *Journal of Clinical Pathology*, 66(4), 343–347. <https://doi.org/10.1136/jclinpath-2012-201243>
- Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., ... Xie, Z. (2008). Single-molecule DNA sequencing of a viral genome. *Science*, 320(5872), 106–109. <https://doi.org/10.1126/science.1150427>
- Hatakeyama, M. (2014). Review *Helicobacter pylori* CagA and Gastric Cancer : *Cell Host and Microbe*, 15(3), 306–316. <https://doi.org/10.1016/j.chom.2014.02.008>
- Hernandez, C., Serrano, C., Einisman, H., Villagran, A., Pena, A., Duarte, I., ... Harris, P. R. (2014). Peptic ulcer disease in *Helicobacter pylori*-infected children: clinical findings and mucosal immune response. *Journal of Pediatric Gastroenterology and Nutrition*, 59(6), 773–778. <https://doi.org/10.1097/MPG.0000000000000500>
- Heuermann, D., & Haas, R. (1995). Genetic organization of a small cryptic plasmid of *Helicobacter pylori*. *Gene*, 165(1), 17–24. [https://doi.org/10.1016/0378-1119\(95\)00469-M](https://doi.org/10.1016/0378-1119(95)00469-M)
- Hu, Yuan, He, L. H., Xiao, D., Liu, G. D., Gu, Y. X., Tao, X. X., & Zhang, J. Z. (2012). Bacterial flora concurrent with helicobacter pylori in the stomach of patients with upper gastrointestinal diseases. *World Journal of Gastroenterology*, 18(11), 1257–1261. <https://doi.org/10.3748/wjg.v18.i11.1257>

- Hu, Yue, Zhang, M., Lu, B., & Dai, J. (2016). *Helicobacter pylori* and Antibiotic Resistance, A Continuing and Intractable Problem. *Helicobacter*, 21(5), 349–363. <https://doi.org/10.1111/hel.12299>
- Ilianov Iliev, H., Dimitrova Kovacheva-Slavova, M., Asenov Angelov, T., Yankov Valkov, H., Bedran, A., & Georgiev Vladimirov, B. (2019). Gastric Microbiota: Between Health and Disease. *Gastrointestinal Stomas*. <https://doi.org/10.5772/intechopen.86926>
- Ilina, E N. (2009). Direct Matrix-Assisted Laser Desorption – Ionisation (MALDI) Bacteria Profiling for Identifying and Characterizing Pathogens, 115–120.
- Ilina, Elena N, Borovskaya, A. D., Serebryakova, M. V, Chelysheva, V. V, Momynaliev, K. T., Maier, T., ... Govorun, V. M. (2010). Application of matrix-assisted laser desorption / ionization time-of-flight mass spectrometry for the study of *Helicobacter pylori*, (Table 1), 328–334. <https://doi.org/10.1002/rcm>
- Illumina Inc. (2015). MiSeq Reporter Metagenomics Workflow Reference Guide, (February 2014).
- Imirzalioglu, C., Hain, T., Hossain, H., Chakraborty, T., & Domann, E. (2010). Erythema caused by a localised skin infection with *Arthrobacter mysorens*. *BMC Infectious Diseases*, 10, 2–5. <https://doi.org/10.1186/1471-2334-10-352>
- Isobe, H., Nishiyama, A., Takano, T., Higuchi, W., Nakagawa, S., Taneike, I., ... Yamamoto, T. (2012). Reduction of overall *Helicobacter pylori* colonization levels in the stomach of mongolian gerbil by *Lactobacillus johnsonii* La1 (LC1) and its in Vitro activities against *H. pylori* motility and adherence. *Bioscience, Biotechnology and Biochemistry*, 76(4), 850–852. <https://doi.org/10.1271/bbb.110921>
- Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., & Reddy, D. N. (2015). Role of the normal gut microbiota. *World Journal of Gastroenterology*, 21(29), 8836–8847. <https://doi.org/10.3748/wjg.v21.i29.8787>
- Jo, H. J., Kim, J., Kim, N., Park, J. H., Nam, R. H., Seok, Y. J., ... Jung, H. C. (2016). Analysis of Gastric Microbiota by Pyrosequencing: Minor Role of

Bibliografía

- Bacteria Other Than *Helicobacter pylori* in the Gastric Carcinogenesis. *Helicobacter*, 21(5), 364–374. <https://doi.org/10.1111/hel.12293>
- Johansen, H. K., Norgaard, A., Andersen, L. P., Jensen, P., Nielsen, H., & Hoiby, N. (1995). Cross-reactive antigens shared by *Pseudomonas aeruginosa*, *Helicobacter pylori*, *Campylobacter jejuni* and *Haemophilus influenzae* may cause false-positive titers of antibody to H. pylori. *Clinical and Diagnostic Laboratory Immunology*, 2(2), 149–155. <https://doi.org/10.1128/cdli.2.2.149-155.1995>
- Kabir, A. M. A., Aiba, Y., Takagi, A., Kamiya, S., Miwa, T., & Koga, Y. (1997). Prevention of *Helicobacter pylori* infection by lactobacilli in a gnotobiotic murine model. *Gut*, 41(1), 49–55. <https://doi.org/10.1136/gut.41.1.49>
- Kachuei, V., Abadi, A. T. B., Rahimi, F., & Forootan, M. (2020). Colonization by *Pseudomonas aeruginosa* and *Staphylococcus aureus* of antral biopsy specimens from gastritis patients uninfected with *Helicobacter pylori*. *Infection and Drug Resistance*, 13, 1411–1417. <https://doi.org/10.2147/IDR.S254967>
- Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., ... Apweiler, R. (2005). The EMBL nucleotide sequence database. *Nucleic Acids Research*, 33(DATABASE ISS.), 29–33. <https://doi.org/10.1093/nar/gki098>
- Karnholz, A., Hoefler, C., Odenbreit, S., Fischer, W., Hofreuter, D., & Haas, R. (2006). Functional and Topological Characterization of Novel Components of the, 188(3), 882–893. <https://doi.org/10.1128/JB.188.3.882>
- Karpiński, T. M. (2019). Role of oral microbiota in cancer development. *Microorganisms*, 7(1). <https://doi.org/10.3390/microorganisms7010020>
- Kassambara, A. (2019). ggpubr: “ggplot2” Based Publication Ready Plots. Retrieved from <https://cran.r-project.org/package=ggpubr>
- Kassambara, A., & Mundt, F. (2017). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Retrieved from <http://www.sthda.com/english/rpkgs/factoextra>
- Kavitt, R. T., & Cifu, A. S. (2017). Management of *Helicobacter pylori* infection.

- JAMA - Journal of the American Medical Association*, 317(15), 1572–1573.
<https://doi.org/10.1001/jama.2017.1949>
- Khot, P. D., Couturier, M. R., Wilson, A., Croft, A., & Fisher, M. A. (2012). Optimization of matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis for bacterial identification. *Journal of Clinical Microbiology*, 50(12), 3845–3852. <https://doi.org/10.1128/JCM.00626-12>
- Klymiuk, I., Bilgilier, C., Stadlmann, A., Thannesberger, J., Kastner, M. T., Högenauer, C., ... Steininger, C. (2017). The human gastric microbiome is predicated upon infection with *Helicobacter pylori*. *Frontiers in Microbiology*, 8(DEC), 1–12. <https://doi.org/10.3389/fmicb.2017.02508>
- Koga, Y., Ohtsu, T., Kimura, K., & Asami, Y. (2019). Probiotic *L. gasseri* strain (LG21) for the upper gastrointestinal tract acting through improvement of indigenous microbiota. *BMJ Open Gastroenterology*, 6(1).
<https://doi.org/10.1136/bmjgast-2019-000314>
- Konturek, J. W. (2003). Discovery by Jaworski of *Helicobacter pylori* and its pathogenetic role in peptic ulcer, gastritis and gastric cancer. *Journal of Physiology and Pharmacology*, 54(SUPPL. 3), 23–41.
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., ... Phillippy, A. M. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology*, 30(7), 693–700. <https://doi.org/10.1038/nbt.2280>
- Kovaleva, J., Degener, J. E., & Van Der Mei, H. C. (2014). Methylobacterium and its role in health care-associated infection. *Journal of Clinical Microbiology*, 52(5), 1317–1321. <https://doi.org/10.1128/JCM.03561-13>
- Kuipers, E. J., Uytterlinde, A. M., Pena, A. S., Hazenberg, H. J. A., Bloemena, E., Lindeman, J., ... Meuwissen, S. G. M. (1995). Increase of *Helicobacter pylori*-associated corpus gastritis during acid suppressive therapy: Implications for long-term safety. *American Journal of Gastroenterology*, 90(9), 1401–1406.
- Kusters, J. G., Gerrits, M. M., Van Strijp, J. A. G., & Vandenbroucke-Grauls, C. M. J. E. (1997). Coccoid forms of *Helicobacter pylori* are the morphologic manifestation of cell death. *Infection and Immunity*, 65(9), 3672–3679.

Bibliografía

<https://doi.org/10.1128/iai.65.9.3672-3679.1997>

- Kusters, Johannes G., Van Vliet, A. H. M., & Kuipers, E. J. (2006). Pathogenesis of *Helicobacter pylori* infection. *Clinical Microbiology Reviews*, 19(3), 449–490. <https://doi.org/10.1128/CMR.00054-05>
- Lacey, S. L., Mehme, S., & Taylor, G. W. (1995). Inhibition of *Helicobacter pylori* growth by 4-hydroxy-2-alkyl-quinolines produced by *Pseudomonas aeruginosa*. *Journal of Antimicrobial Chemotherapy*, 36(5), 827–831. <https://doi.org/10.1093/jac/36.5.827>
- Lagier, J. C., Dubourg, G., Million, M., Cadoret, F., Bilen, M., Fenollar, F., ... Raoult, D. (2018). Culturing the human microbiota and culturomics. *Nature Reviews Microbiology*, 16(9), 540–550. <https://doi.org/10.1038/s41579-018-0041-0>
- Lan, H. C., Chen, T. S., Li, A. F. Y., Chang, F. Y., & Lin, H. C. (2012). Additional corpus biopsy enhances the detection of *Helicobacter pylori* infection in a background of gastritis with atrophy. *BMC Gastroenterology*, 12(1), 1. <https://doi.org/10.1186/1471-230X-12-182>
- Lara-Ramírez, E. E., Segura-Cabrera, A., Guo, X., Yu, G., García-Pérez, C. A., & Rodríguez-Pérez, M. A. (2011). New implications on genomic adaptation derived from the *Helicobacter pylori* genome comparison. *PLoS ONE*, 6(2). <https://doi.org/10.1371/journal.pone.0017300>
- Layfield, L. J., Witt, B. L., Metzger, K. G., & Anderson, G. M. (2011). Extraneous tissue: A potential source for diagnostic error in surgical pathology. *American Journal of Clinical Pathology*, 136(5), 767–772. <https://doi.org/10.1309/AJCP4FFSBPHAU8IU>
- Lê, S., Josse, J., & Husson, F. (2008). {FactoMineR}: A Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1), 1–18. <https://doi.org/10.18637/jss.v025.i01>
- Lertpiriyapong, K., Whary, M. T., Muthupalani, S., Lofgren, J. L., Gamazon, E. R., Feng, Y., ... Fox, J. G. (2014). Gastric colonisation with a restricted commensal microbiota replicates the promotion of neoplastic lesions by diverse intestinal microbiota in the *Helicobacter pylori* INS-GAS mouse model of gastric carcinogenesis. *Gut*, 63(1), 54–63.

- <https://doi.org/10.1136/gutjnl-2013-305178>
- Li, X. X., Wong, G. L. H., To, K. F., Wong, V. W. S., Lai, L. H., Chow, D. K. L., ... Ding, C. (2009). Bacterial microbiota profiling in gastritis without *Helicobacter pylori* infection or non-steroidal anti-inflammatory drug use. *PLoS ONE*, 4(11), 1–9. <https://doi.org/10.1371/journal.pone.0007985>
- Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., ... Achtman, M. (2007). An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*, 445(7130), 915–918. <https://doi.org/10.1038/nature05562>
- Liu, X., Shao, L., Liu, X., Ji, F., Mei, Y., Cheng, Y., ... Ling, Z. (2019). Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer. *EBioMedicine*, 40, 336–348. <https://doi.org/10.1016/j.ebiom.2018.12.034>
- Llorca, L., Perez-Perez, G., Urruzuno, P., Martinez, M. J., Iizumi, T., Gao, Z., ... Alarcon, T. (2016). Characterization of the Gastric Microbiota in a Pediatric Population According to *Helicobacter pylori* Status. *The Pediatric Infectious Disease Journal*. <https://doi.org/10.1097/INF.0000000000001383>
- Llorca, L., Pérez-Pérez, G., Urruzuno, P., Martinez, M. J., Iizumi, T., Gao, Z., ... Alarcón, T. (2017). Characterization of the gastric microbiota in a pediatric population according to *Helicobacter pylori* status. *Pediatric Infectious Disease Journal*, 36(2), 173–178. <https://doi.org/10.1097/INF.0000000000001383>
- Loayza, M. F., Villavicencio, F. X., Santander, S. C., Baldeón, M., Ponce, L. K., Salvador, I., & Díaz, N. V. (2015). Improved method for extraction and detection of *Helicobacter pylori* DNA in formalin-fixed paraffin embedded gastric biopsies using laser micro-dissection. *MethodsX*, 2, 1–7. <https://doi.org/10.1016/j.mex.2014.11.003>
- Lofgren, J. L., Whary, M. T., Ge, Z., Muthupalani, S., Taylor, N. S., Mobley, M., ... Fox, J. G. (2011). Lack of commensal flora in *Helicobacter pylori*-infected INS-GAS mice reduces gastritis and delays intraepithelial neoplasia. *Gastroenterology*, 140(1), 210-220.e4. <https://doi.org/10.1053/j.gastro.2010.09.048>

Bibliografía

- Lopes, A. I., Vale, F. F., & Oleastro, M. (2014). *Helicobacter pylori* infection - Recent developments in diagnosis. *World Journal of Gastroenterology*, 20(28), 9299–9313. <https://doi.org/10.3748/wjg.v20.i28.9299>
- López-Brea, M., Alarcón, T., Baquero, M., Domingo, D., & Royo, G. (2004). Diagnóstico microbiológico de la infección por *Helicobacter pylori*. *Procedimientos En Microbiología Clínica*, 1–25. Retrieved from www.seimc.org
- López-García, A., Pineda-Quiroga, C., Atxaerandio, R., Pérez, A., Hernández, I., García-Rodríguez, A., & González-Recio, O. (2018). Comparison of mothur and QIIME for the analysis of rumen microbiota composition based on 16S rRNA amplicon sequences. *Frontiers in Microbiology*, 9(DEC), 1–11. <https://doi.org/10.3389/fmicb.2018.03010>
- Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology*, 73(5), 1576–1585. <https://doi.org/10.1128/AEM.01996-06>
- Makki, K., Deehan, E. C., Walter, J., & Bäckhed, F. (2018). The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host and Microbe*, 23(6), 705–715. <https://doi.org/10.1016/j.chom.2018.05.012>
- Malfertheiner, P., Megraud, F., O'Morain, C. A., Gisbert, J. P., Kuipers, E. J., Axon, A. T., ... You, W. (2017). Management of *Helicobacter pylori* infection-the Maastricht V/Florence consensus report. *Gut*, 66(1), 6–30. <https://doi.org/10.1136/gutjnl-2016-312288>
- Malfertheiner, P., Megraud, F., O'Morain, C., Bazzoli, F., El-Omar, E., Graham, D., ... Xiao, S. (2007). Current concepts in the management of *Helicobacter pylori* infection: The maastricht III consensus report. *Gut*, 56(6), 772–781. <https://doi.org/10.1136/gut.2006.101634>
- Marchandin, H., Teyssier, C., Siméon de Buochberg, M., Jean-Pierre, H., Carriere, C., & Jumas-Bilak, E. (2003). Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus *Veillonella*: Implications for phylogeny and taxonomy. *Microbiology*, 149(6), 1493–1501. <https://doi.org/10.1099/mic.0.26132-0>

- Marine, R. L., Magaña, L. C., Castro, C. J., Zhao, K., Montmayeur, A. M., Schmidt, A., ... Oberste, M. S. (2020). Comparison of Illumina MiSeq and the Ion Torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses. *Journal of Virological Methods*, 280, 1–36. <https://doi.org/10.1016/j.jviromet.2020.113865>
- Marshall, B. J., & Warren, J. R. (1984). Unidentified Curved Bacilli in the Stomach of Patients With Gastritis and Peptic Ulceration. *The Lancet*, 323(8390), 1311–1315. [https://doi.org/10.1016/S0140-6736\(84\)91816-6](https://doi.org/10.1016/S0140-6736(84)91816-6)
- Martinez, K. B., Leone, V., & Chang, E. B. (2017). Microbial metabolites in health and disease: Navigating the unknown in search of function. *Journal of Biological Chemistry*, 292(21), 8553–8559. <https://doi.org/10.1074/jbc.R116.752899>
- Martinsen, T. C., Bergh, K., & Waldum, H. L. (2005). Gastric juice: A barrier against infectious diseases. *Basic and Clinical Pharmacology and Toxicology*, 96(2), 94–102. <https://doi.org/10.1111/j.1742-7843.2005.pto960202.x>
- Mashima, J., Kodama, Y., Fujisawa, T., Katayama, T., Okuda, Y., Kaminuma, E., ... Takagi, T. (2017). DNA Data Bank of Japan. *Nucleic Acids Research*, 45(D1), D25–D31. <https://doi.org/10.1093/nar/gkw1001>
- Matsuoka, K., & Kanai, T. (2015). The gut microbiota and inflammatory bowel disease. *Seminars in Immunopathology*, 37(1), 47–55. <https://doi.org/10.1007/s00281-014-0454-4>
- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., Desantis, T. Z., Probst, A., ... Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME Journal*, 6(3), 610–618. <https://doi.org/10.1038/ismej.2011.139>
- McNulty, R., Ulmschneider, J. P., Luecke, H., & Ulmschneider, M. B. (2013). Mechanisms of molecular transport through the urea channel of *Helicobacter pylori*. *Nature Communications*, 4, 1–10. <https://doi.org/10.1038/ncomms3900>
- Mégraud, F., Floch, P., Labenz, J., & Lehours, P. (2016). Diagnostic of

Bibliografía

- Helicobacter pylori* infection. *Helicobacter*, 21, 8–13.
<https://doi.org/10.1111/hel.12333>
- Metzger, J., Styger, S., Sieber, C., Von Fliie, M., Vogelbach, P., & Harder, F. (2001). Prevalence of *Helicobacter pylori* infection in peptic ulcer perforations. *Swiss Medical Weekly*, 131(7–8), 99–103.
<https://doi.org/10.4414/smw.2001.06118>
- Miao, R., Wan, C., & Wang, Z. (2020). The relationship of gastric microbiota and *Helicobacter pylori* infection in pediatrics population. *Helicobacter*, 25(1), 13–22. <https://doi.org/10.1111/hel.12676>
- Microbiology by numbers. (2011). *Nature Reviews Microbiology*, 9(9), 628–628.
<https://doi.org/10.1038/nrmicro2644>
- Miqueleiz-Zapatero, A., Alba-Rubio, C., Domingo-García, D., Cantón, R., Gómez-García de la Pedrosa, E., Aznar-Cano, E., ... Alarcón, T. (2020). First national survey of the diagnosis of *Helicobacter pylori* infection in Clinical Microbiology Laboratories in Spain. *Enfermedades Infecciosas y Microbiología Clínica*. <https://doi.org/10.1016/j.eimc.2019.11.008>
- Mobley, H. L., Mendz, G. L., & Hazell, S. L. (2001). *Helicobacter pylori: Physiology and Genetics* (H. L. T. Mobley, G. L. Mendz, & S. L. Hazell, Eds.). Washington (DC).
- Moller, H., Heseltine, E., & Vainio, H. (1995). Working group report on schistosomes, liver flukes and *Helicobacter pylori*. *International Journal of Cancer*, 60(5), 587–589. <https://doi.org/10.1002/ijc.2910600502>
- Mowat, C., Williams, C., Gillen, D., Hossack, M., Gilmour, D., Carswell, A., ... McColl, K. E. L. (2000). Omeprazole, *Helicobacter pylori* status, and alterations in the intragastric milieu facilitating bacterial N-nitrosation. *Gastroenterology*, 119(2), 339–347. <https://doi.org/10.1053/gast.2000.9367>
- Muñoz-Peña, C., Ocaña-Cano, M. J., Amores-Antequera, C., & Cantudo-Muñoz, P. (2016). Infección cutánea por *Gordonia aarii*. *Enfermedades Infecciosas y Microbiología Clínica*, 34(10), 685–686.
<https://doi.org/10.1016/j.eimc.2016.02.006>
- Muñoz-Ramírez, Z. Y., Mendez-Tenorio, A., Kato, I., Bravo, M. M., Rizzato, C.,

- Thorell, K., ... Torres, J. (2017). Whole genome sequence and phylogenetic analysis show *Helicobacter pylori* strains from Latin America have followed a unique evolution pathway. *Frontiers in Cellular and Infection Microbiology*, 7(FEB). <https://doi.org/10.3389/fcimb.2017.00050>
- Nagano, T., Otoshi, T., Hazama, D., Kiri, T., Umezawa, K., Katsurada, N., & Nishimura, Y. (2019). Novel cancer therapy targeting microbiome. *OncoTargets and Therapy*. <https://doi.org/10.2147/OTT.S207546>
- Nardone, G., & Compare, D. (2015). The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases? *United European Gastroenterology Journal*, 3(3), 255–260. <https://doi.org/10.1177/2050640614566846>
- Naumov, I., & Fenjvesi, A. (2011). [Correlation between rapid urease test and pathohistological gastrobiopsy finding with positive immunological test in detecting *Helicobacter pylori* infection]. *Medicinski Pregled*, 64(7–8), 413–417. <https://doi.org/10.2298/MPNS1108413N>
- Ndip, R. N., MacKay, W. G., Farthing, M. J. G., & Weaver, L. T. (2003). Culturing *Helicobacter pylori* from Clinical Specimens: Review of Microbiologic Methods. *Journal of Pediatric Gastroenterology and Nutrition*, 36(5), 616–622. <https://doi.org/10.1097/00005176-200305000-00005>
- Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. Retrieved from <https://cran.r-project.org/package=RColorBrewer>
- O'Hara, A. M., & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, 7(7), 688–693. <https://doi.org/10.1038/sj.embor.7400731>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2019). vegan: Community Ecology Package. Retrieved from <https://cran.r-project.org/package=vegan>
- Parkin, D. M. (2004). International variation. *Oncogene*, 23(38), 6329–6340. <https://doi.org/10.1038/sj.onc.1207726>
- Parsons, B. N., Ijaz, U. Z., D'Amore, R., Burkitt, M. D., Eccles, R., Lenzi, L., ... Pritchard, D. M. (2017). Comparison of the human gastric microbiota in hypochlorhydric states arising as a result of *Helicobacter pylori*-induced

Bibliografía

- atrophic gastritis, autoimmune atrophic gastritis and proton pump inhibitor use. *PLoS Pathogens*, 13(11), 1–18.
<https://doi.org/10.1371/journal.ppat.1006653>
- Patel, J. (2001). 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis*, 6(4), 313–321.
<https://doi.org/10.1054/modi.2001.29158>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Louppe, G. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12.
- Peng, C., Li, N. S., Hu, Y., & Lu, N. H. (2019). Impact factors that modulate gastric cancer risk in *Helicobacter pylori*-infected rodent models. *Helicobacter*, 24(4), 1–11. <https://doi.org/10.1111/hel.12580>
- Pérez, T., Alba, C., Aparicio, M., de Andrés, J., Ruiz Santa Quiteria, J. A., Rodríguez, J. M., & Gibello, A. (2019). Abundant bacteria in the proximal and distal intestine of healthy Siberian sturgeons (*Acipenser baerii*). *Aquaculture*, 506(3), 325–336.
<https://doi.org/10.1016/j.aquaculture.2019.03.055>
- Perri, F., Quitadamo, M., Ricciardi, R., Piepoli, A., Cotugno, R., Gentile, A., ... Andriulli, A. (2005). Comparison of a monoclonal antigen stool test (Hp StAR) with the 13 C-urea breath test in monitoring *Helicobacter pylori* eradication therapy. *World Journal of Gastroenterology*, 11(37), 5878–5881. <https://doi.org/10.3748/wjg.v11.i37.5878>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), 590–596. <https://doi.org/10.1093/nar/gks1219>
- R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria.
- Rahman, R., Asombang, A. W., & Ibdah, J. A. (2014). Characteristics of gastric cancer in Asia. *World Journal of Gastroenterology*, 20(16), 4483–4490.
<https://doi.org/10.3748/wjg.v20.i16.4483>

- Ribeiro, M. L., Godoy, A. P. O., Benvengo, Y. H. B., Ecclissato, C. C., Mendonça, S., & Pedrazzoli, J. (2004). The influence of endoscopic procedures upon the contamination of *Helicobacter pylori* cultures. *Arquivos de Gastroenterologia*, *41*(2), 100–103. <https://doi.org/10.1590/s0004-28032004000200006>
- Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 19126–19131. <https://doi.org/10.1073/pnas.0906412106>
- Ritchie, B., Brewster, D., Tran, C. D., McNeil, Y., Zacharakis, B., Davidson, G. P., & Butler, R. N. (2009). Lack of diagnostic accuracy of the monoclonal stool antigen test for detection of *Helicobacter pylori* infection in young Australian aboriginal children. *Pediatric Infectious Disease Journal*, *28*(4), 287–289. <https://doi.org/10.1097/INF.0b013e31818e039b>
- Rivas, F., & Hernandez, F. (2012). *Helicobacter pylori*: Factores de virulencia, patología y diag-. *Rev Biomed*, *11*(3), 187–205. Retrieved from www.revbiomed.uady.mx/pdf/rb001136.pdf
- Roma, E., Kafritsa, Y., Panayiotou, J., Liakou, R., & Constantopoulos, A. (2001). Is peptic ulcer a common cause of upper gastrointestinal symptoms? *European Journal of Pediatrics*, *160*(8), 497–500.
- Rosselló-Móra, R. (2012). Towards a taxonomy of Bacteria and Archaea based on interactive and cumulative data repositories. *Environmental Microbiology*, *14*(2), 318–334. <https://doi.org/10.1111/j.1462-2920.2011.02599.x>
- Salama, N. R., Hartung, M. L., & Müller, A. (2013). Life in the stomach: persistence strategies *Helicobacter Pylori*. *National Institute Health*, *11*(6), 385–399. <https://doi.org/10.1038/nrmicro3016.Life>
- Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogestraat, D. R., Cummings, L. A., Sengupta, D. J., ... Hoffman, N. G. (2014). Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Applied and Environmental Microbiology*, *80*(24), 7583–7591. <https://doi.org/10.1128/AEM.02206-14>

Bibliografía

- Salonen, A., Salojärvi, J., Lahti, L., & de Vos, W. M. (2012). The adult intestinal core microbiota is determined by analysis depth and health status. *Clinical Microbiology and Infection*, *18*(SUPPL. 4), 16–20.
<https://doi.org/10.1111/j.1469-0691.2012.03855.x>
- Sanger, F. (1959). Chemistry of Insulin: Determination of the structure of insulin opens the way to greater understanding of life processes. *Science*, *129*(3359), 1340–1344. <https://doi.org/10.1126/science.129.3359.1340>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Sauer, S., Freiwald, A., Maier, T., Kube, M., Reinhardt, R., Kostrzewa, M., & Geider, K. (2008). Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS ONE*, *3*(7).
<https://doi.org/10.1371/journal.pone.0002843>
- Savage, D. C. (1977). Microbial Ecology of the Gastrointestinal Tract. *Annual Review of Microbiology*, *31*(1), 107–133.
<https://doi.org/10.1146/annurev.mi.31.100177.000543>
- Schloss, P. D. (2019). Reintroducing mothur: 10 Years Later. *Applied and Environmental Microbiology*, *86*(2), 1–13.
<https://doi.org/10.1128/AEM.02343-19>
- Schulthess, B., Brodner, K., Bloemberg, G. V., Zbinden, R., Böttger, E. C., & Hombach, M. (2013). Identification of Gram-positive cocci by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry: Comparison of different preparation methods and implementation of a practical algorithm for routine diagnostics. *Journal of Clinical Microbiology*, *51*(6), 1834–1840. <https://doi.org/10.1128/JCM.02654-12>
- Schulz, C., Koch, N., Schütte, K., Pieper, D. H., & Malfertheiner, P. (2015). *H.pylori* and its modulation of gastrointestinal microbiota. *Journal of Digestive Diseases*, *16*(3), 109–117. <https://doi.org/10.1111/1751-2980.12233>
- Sender, R., Fuchs, S., & Milo, R. (2016a). Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*, *164*(3), 337–

340. <https://doi.org/10.1016/j.cell.2016.01.013>
- Sender, R., Fuchs, S., & Milo, R. (2016b). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), 1–21. <https://doi.org/10.1371/journal.pbio.1002533>
- Sherr, C. J., & Weber, J. D. (2000). The ARF/p53 pathway. *Current Opinion in Genetics and Development*, *10*(1), 94–99. [https://doi.org/10.1016/S0959-437X\(99\)00038-6](https://doi.org/10.1016/S0959-437X(99)00038-6)
- Shimoyama, T. (2013). Stool antigen tests for the management of *Helicobacter pylori* infection. *World Journal of Gastroenterology*, *19*(45), 8188–8191. <https://doi.org/10.3748/wjg.v19.i45.8188>
- Simala-Grant, J. L., & Taylor, D. E. (2004). Molecular biology methods for the characterization of *Helicobacter pylori* infections and their diagnosis. *Apmis*, *112*(11–12), 886–897. <https://doi.org/10.1111/j.1600-0463.2004.apm11211-1211.x>
- Sisto, F., Brenciaglia, M. I., Scaltrito, M. M., & Dubini, F. (2000). *Helicobacter pylori*: ureA, cagA and vacA expression during conversion to the coccoid form. *International Journal of Antimicrobial Agents*, *15*(4), 277–282. [https://doi.org/10.1016/S0924-8579\(00\)00188-6](https://doi.org/10.1016/S0924-8579(00)00188-6)
- Sipponen, P., Kekki, M., Haapakoski, J., Ihamäki, T., & Siurala, M. (1985). Gastric cancer risk in chronic atrophic gastritis: Statistical calculations of cross-sectional data. *International Journal of Cancer*, *35*(2), 173–177. <https://doi.org/10.1002/ijc.2910350206>
- Solnick, J. V., O'rourke, J. L., Vandamme, P., & Lee, A. (2006). The Genus *Helicobacter*. In *The Prokaryotes* (pp. 139–177). New York, NY: Springer New York. https://doi.org/10.1007/0-387-30747-8_5
- Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., Kämpfer, P., Maiden, M. C. J., ... Whitman, W. B. (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, *52*(3), 1043–1047. <https://doi.org/10.1099/00207713-52-3-1043>
- STACKEBRANDT, E., & GOEBEL, B. M. (1994). Taxonomic Note: A Place for

Bibliografía

- DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, 44(4), 846–849. [https://doi.org/10.1016/S0140-6736\(01\)43317-4](https://doi.org/10.1016/S0140-6736(01)43317-4)
- Stewart, O. A., Wu, F., & Chen, Y. (2020). The role of gastric microbiota in gastric cancer. *Gut Microbes*, 11(5), 1220–1230. <https://doi.org/10.1080/19490976.2020.1762520>
- Sung, J., Kim, N., Kim, J., Jo, H. J., Park, J. H., Nam, R. H., ... Jung, H. C. (2016). Comparison of Gastric Microbiota Between Gastric Juice and Mucosa by Next Generation Sequencing Method. *Journal of Cancer Prevention*, 21(1), 60–65. <https://doi.org/10.15430/jcp.2016.21.1.60>
- Tawfik, S. A., Azab, M. M., Ahmed, A. A. A., & Fayyad, D. M. (2018). Illumina MiSeq Sequencing for Preliminary Analysis of Microbiome Causing Primary Endodontic Infections in Egypt. *International Journal of Microbiology*, 2018. <https://doi.org/10.1155/2018/2837328>
- Taylor, N. S., Fox, J. G., Akopyants, N. S., Berg, D. E., Thompson, N., Shames, B., ... Correa, P. (1995). Long-term colonization with single and multiple strains of *Helicobacter pylori* assessed by DNA fingerprinting. *Journal of Clinical Microbiology*, 33(4), 918–923. <https://doi.org/10.1128/jcm.33.4.918-923.1995>
- Tenover, F. C., Fennel, C. L., Tanner, A., & Paster, B. J. (1992). “Subdivision C5: *Epsilon* subclass,” in *The Prokaryotes* (2nd Edn). Springer-Verlag Berlin.
- Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., ... Craig Venter, J. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, 388(6642), 539–547. <https://doi.org/10.1038/41483>
- Tsuda, A., Suda, W., Morita, H., Takanashi, K., Takagi, A., Koga, Y., & Hattori, M. (2015). Influence of proton-pump inhibitors on the luminal microbiota in the gastrointestinal tract. *Clinical and Translational Gastroenterology*, 6(6), e89-9. <https://doi.org/10.1038/ctg.2015.20>
- Ubeda, C., Djukovic, A., & Isaac, S. (2017). Roles of the intestinal microbiota in

- pathogen protection. *Clinical & Translational Immunology*, 6(2), e128.
<https://doi.org/10.1038/cti.2017.2>
- Vu, V. Q. (2011). ggbiplot: A ggplot2 based biplot. Retrieved from
<http://github.com/vqv/ggbiplot>
- Waite, D. W., Chuvochina, M. S., & Hugenholtz, P. (2019). Road Map of the Phylum Campylobacterota. *Bergey's Manual of Systematics of Archaea and Bacteria*, (i), 1–11. <https://doi.org/10.1002/9781118960608.bm00040>
- Waite, D. W., Vanwonterghem, I., Rinke, C., Parks, D. H., Zhang, Y., Takai, K., ... Hugenholtz, P. (2017). Comparative genomic analysis of the class Epsilonproteobacteria and proposed reclassification to epsilonbacteraeota (phyl. nov.). *Frontiers in Microbiology*, 8(APR).
<https://doi.org/10.3389/fmicb.2017.00682>
- Waite, D. W., Vanwonterghem, I., Rinke, C., Parks, D. H., Zhang, Y., Takai, K., ... Hugenholtz, P. (2018). Addendum: Comparative genomic analysis of the class Epsilonproteobacteria and proposed reclassification to Epsilonbacteraeota (phyl. nov.) [Front. Microbiol., 8, (2017), (682)] doi: 10.3389/fmicb.2017.00682. *Frontiers in Microbiology*, 9(APR), 1–2.
<https://doi.org/10.3389/fmicb.2018.00772>
- Wang, B., Yao, M., Lv, L., Ling, Z., & Li, L. (2017). The Human Microbiota in Health and Disease. *Engineering*, 3(1), 71–82.
<https://doi.org/10.1016/J.ENG.2017.01.008>
- Wang, L., Zhou, J., Xin, Y., Geng, C., Tian, Z., Yu, X., & Dong, Q. (2016). Bacterial overgrowth and diversification of microbiota in gastric cancer. *European Journal of Gastroenterology and Hepatology*, 28(3), 261–266.
<https://doi.org/10.1097/MEG.0000000000000542>
- Wang, M., Zhang, X., Jiang, T., Hu, S., Yi, Z., Zhou, Y., ... Chen, S. (2017). Liver abscess caused by Pannonibacter phragmitetus: Case report and literature review. *Frontiers in Medicine*, 4(APR), 14–17.
<https://doi.org/10.3389/fmed.2017.00048>
- Wang, Y. K., Kuo, F. C., Liu, C. J., Wu, M. C., Shih, H. Y., Wang, S. S. W., ... Wu, D. C. (2015). Diagnosis of *Helicobacter pylori* infection: Current options and developments. *World Journal of Gastroenterology*, 21(40),

Bibliografía

- 11221–11235. <https://doi.org/10.3748/wjg.v21.i40.11221>
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., ... Venables, B. (2019). gplots: Various R Programming Tools for Plotting Data. Retrieved from <https://cran.r-project.org/package=gplots>
- Warnes, G. R., Bolker, B., & Lumley, T. (2018). gtools: Various R Programming Tools. Retrieved from <https://cran.r-project.org/package=gtools>
- Wayne, L. G., Moore, W. E. C., Stackebrandt, E., Kandler, O., Colwell, R. R., Krichevsky, M. I., ... Moore, L. H. (1987). Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *International Journal of Systematic and Evolutionary Microbiology*, 37(4), 463–464. <https://doi.org/10.1099/00207713-37-4-463>
- Whittaker, A. R. H., & Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity Published by : International Association for Plant Taxonomy (IAPT) Stable URL : <http://www.jstor.org/stable/1218190> . extend access to Taxon . EVOLUTION AND MEASUREMENT OF SPECIES DIVERSITY ". *Diversity*, 21(2), 213–251. <https://doi.org/10.2307/1218190>
- Wickham, H. (2007). Reshaping Data with the {reshape} Package. *Journal of Statistical Software*, 21(12), 1–20. Retrieved from <http://www.jstatsoft.org/v21/i12/>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H. (2017). tidyverse: Easily Install and Load the “Tidyverse.” Retrieved from <https://cran.r-project.org/package=tidyverse>
- Wickham, H. (2019). forcats: Tools for Working with Categorical Variables (Factors). Retrieved from <https://cran.r-project.org/package=forcats>
- Williams, C., & McColl, K. E. L. (2006). Review article: Proton pump inhibitors and bacterial overgrowth. *Alimentary Pharmacology and Therapeutics*, 23(1), 3–10. <https://doi.org/10.1111/j.1365-2036.2006.02707.x>
- Winkler, M. A., Uher, J., Cepa, S., Division, D., Division, P. P., Laboratories, A., & Park, A. (1999). Direct Analysis and Identification of Helicobacter and

- Campylobacter Species by MALDI-TOF Mass Spectrometry, *71*(16), 3416–3419.
- Xie, Y. (2019). knitr: A General-Purpose Package for Dynamic Report Generation in R. Retrieved from <https://yihui.name/knitr/>
- Yang, I., Woltemate, S., Piazuolo, M. B., Bravo, L. E., Yopez, M. C., Romero-Gallo, J., ... Suerbaum, S. (2016). Different gastric microbiota compositions in two human populations with high and low gastric cancer risk in Colombia. *Scientific Reports*, *6*(September 2015), 1–10. <https://doi.org/10.1038/srep18594>
- Yang, L., Zhang, J., Xu, J., Wei, X., Yang, J., Liu, Y., ... Gai, Z. (2019). *Helicobacter pylori* Infection Aggravates Dysbiosis of Gut Microbiome in Children With Gastritis. *Frontiers in Cellular and Infection Microbiology*, *9*(November), 1–14. <https://doi.org/10.3389/fcimb.2019.00375>
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., ... Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, *12*(9), 635–645. <https://doi.org/10.1038/nrmicro3330>
- Yu, G., Torres, J., Hu, N., Medrano-Guzman, R., Herrera-Goepfert, R., Humphrys, M. S., ... Goldstein, A. M. (2017). Molecular characterization of the human stomach microbiota in Gastric Cancer Patients. *Frontiers in Cellular and Infection Microbiology*, *7*(JUL), 1–11. <https://doi.org/10.3389/fcimb.2017.00302>
- Zakrzewski, M., Proietti, C., Ellis, J. J., Hasan, S., Brion, M. J., Berger, B., & Krause, L. (2017). Calypso: A user-friendly web-server for mining and visualizing microbiome-environment interactions. *Bioinformatics*, *33*(5), 782–783. <https://doi.org/10.1093/bioinformatics/btw725>
- Zhang, J., Chiodini, R., Badr, A., & Zhang, G. (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, *38*(3), 95–109. <https://doi.org/10.1016/j.jgg.2011.02.003>
- Zhu, X. Y., & Liu, F. (2017). Probiotics as an adjuvant treatment in *Helicobacter pylori* eradication therapy. *Journal of Digestive Diseases*, *18*(4), 195–202. <https://doi.org/10.1111/1751-2980.12466>

Anexo I

Anexo I

El siguiente anexo contiene los comandos para realizar los análisis de secuencias pareadas obtenidas por medio de la plataforma Illumina una vez “demultiplexadas” (separadas por muestras) y preferiblemente filtradas por calidad.

Para la utilización de los siguientes comandos será preciso disponer de una versión de QIIME2 de manera nativa o por medio de una máquina virtual. Las instrucciones para su instalación pueden encontrarse en su página web “<https://docs.qiime2.org/2020.8/install/>”. Para la realización del siguiente protocolo se recomienda su instalación por medio de una máquina virtual y optimizarla en función de las capacidades del ordenador (generalmente aumentando su disponibilidad de memoria RAM y procesadores dedicados en Configuración / Sistema).

Antes de empezar se recomienda preparar un archivo de datos “metadata/mapping” en el que se disponga de cada una de las muestras por fila y las variables por columna siendo la primera columna siempre el nombre de las muestras con el encabezado “**SampleID**” :

Ejemplo:

SampleID	Biopsia_Vs_Jugo	Pais	Hp
Muestra_1	Biopsia	Espana	HpPos
Muestra_2	Biopsia	Portugal	Neg
Muestra_3	Jugo_gastrico	Espana	Neg
Muestra_4	Biopsia	Espana	HpPos
Muestra_5	Biopsia	Portugal	HpPos
Muestra_6	Biopsia	Portugal	HpPos
Muestra_7	Jugo_gastrico	Portugal	HpPos
Muestra_8	Biopsia	Portugal	Neg
Muestra_9	Biopsia	Espana	Neg
Muestra_10	Jugo_gastrico	Espana	HpPos
Muestra_11	Biopsia	Espana	HpPos
Muestra_12	Biopsia	Espana	HpPos

Lo guardaremos como texto delimitado por tabulaciones para facilitar su implementación al análisis.

Anexo I

Una vez instalada siguiendo sus instrucciones procedemos de la siguiente forma:

- 1- Generamos una carpeta con el nombre del proyecto en el directorio de preferencia (generalmente el escritorio de linux).
- 2- Dentro de la carpeta del proyecto creamos otra carpeta con el nombre “seqs” y movemos las secuencias a dicha carpeta y volvemos a la carpeta del proyecto.
- 3- Abrimos la terminal en la carpeta del proyecto (botón derecho del ratón) y creamos las carpetas accesorias escribiendo:

```
mkdir artifact
mkdir visu
mkdir data
```

- 4- Cargamos las secuencias a qiime2*:

```
qiime tools import \
--type SampleData[PairedEndSequencesWithQuality] \
--input-path seqs/ \
--input-format CasavaOneEightSingleLanePerSampleDirFmt \
--output-path artifact/demux.qza
```

*Hay otros formatos que pueden aparecer al buscar proyectos en las plataformas web. En tal caso el formato podría no ser tipo Casava y podría ser necesario indicar a que muestra pertenece cada archivo y su dirección de lectura. En tal caso hay que crear un archivo de texto (.txt) con la dirección de cada muestra en cada línea al que llamaremos “manifiesto.txt”

Ejemplo de manifiesto.txt:

```
Muestra_1,/media/sf_qiime/Proyecto/seqs/Muestra_1_R1_all.fastq,forward  
Muestra_1,/media/sf_qiime/Proyecto/seqs/Muestra_1_R2_all.fastq,reverse  
Muestra_2...
```

Y cargamos las secuencias

```
qiime tools import \  
--type SampleData[PairedEndSequencesWithQuality] \  
--input-path manifiesto.txt \  
--output-path artifact/demux.qza \  
--input-format PairedEndFastqManifestPhred33
```

5- Generamos el archivo visual para comprobar el estado de las secuencias y determinar si hay que realizar algún corte para mejorar la calidad de las secuencias.

```
qiime demux summarize \  
--i-data artifact/demux.qza \  
--o-visualization visu/demux-summary-1.qzv
```

Anexo I

6- Generamos la tabla de secuencias ASVs mediante el algoritmo DADA2 y realizamos los cortes que viéramos que fueran necesarios (generalmente buscamos que la calidad de la mediana en una posición de los nucleótidos no baje mucho de 20) sustituyendo los "0" por el valor que corresponda :

```
qiime dada2 denoise-paired \  
--i-demultiplexed-seqs artifact/demux.qza \  
--o-representative-sequences artifact/rep-seqs.qza \  
--o-table artifact/table.qza \  
--p-trunc-len-f 0 \  
--p-trunc-len-r 0 \  
--p-trim-left-f 0 \  
--p-trim-left-r 0 \  
--p-max-ee 1 \  
--p-n-threads 0 \  
--o-denoising-stats data/stats.qza
```

QIIME2 tiene muchas herramientas que pueden aplicarse a partir de este punto y se recomienda profundizar en y navegar por sus foros. En nuestro caso procedemos a realizar la asignación taxonómica directamente. Para ello será necesario disponer de una base de datos de secuencias de referencia.

7- Obtención de la base de datos de referencia: hay diferentes protocolos pudiendo extraerse las tablas de las secuencias representativas y utilizarlas con la plataforma RDP o descargar la base de datos de referencia de Silva (se recomienda la más actual posible). Para la plataforma de QIIME2 hay diferentes posibilidades de descarga y el procedimiento general se encuentra en: https://github.com/mikerobeson/make_SILVA_db.

*Con el objetivo de facilitar el proceso de descarga y preparación, se recomienda descargar la versión de su foro: <https://forum.qiime2.org/t/silva-138-classifiers/13131>

El archivo es: SILVA-138-SSURef-full-length-classifier.qza

Según su desarrollador se recomienda preparar las secuencias en función de la región hipervariable que vaya a ser utilizada y agilizar los análisis siendo necesario realizarlo solo una vez por cada actualización de la base de datos. No es obligatorio y no se puede realizar si en tu estudio hay más de un proyecto con diferentes regiones hipervariables.

Las regiones más comunes son las regiones V3-V4* y para ellas se puede realizar el siguiente comando

```
#Extracción de las secuencias de referencia
qiime feature-classifier extract-reads \
--i-sequences SILVA-138-SSURef-Full-Seqs.qza \
--p-f-primer CCTACGGGNGGCWGCAG\
--p-r-primer GGATTAGATACCCBDGTAGTC\
--p-trunc-len 0 \
--p-min-length 150 \
--p-max-length 500 \
--o-reads ref-seqs-v138-V3V4.qza \

#Prepara el clasificador:

qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads ref-seqs.qza \
--i-reference-taxonomy Silva-v138-full-length-seq-taxonomy.qza \
--o-classifier silva138_v34_classifier.qza\
```

Las secuencias de los oligonucleótidos utilizados para realizar la extracción de las secuencias en este caso fueron: Primer_341F_CCTACGGGNGGCWGCAG y Primer_805R_revcomp_GACTACHVGGGTATCTAATCC

*En caso de realizar el estudio sobre otras regiones hipervariables será necesario realizar la extracción con las secuencias que corresponda.

Anexo I

8- Movemos el archivo generado “silva138_v34_classifier.qza” a la carpeta del proyecto y ejecutamos los siguientes comandos:

```
qiime feature-classifier classify-sklearn \  
--i-classifier silva132_v34_classifier.qza \  
--i-reads data/rep-seqs.qza \  
--p-n-jobs -2 \  
--o-classification data/taxonomy_dada.qza
```

9- Para finalizar con este protocolo necesitamos el archivo “metadata/mapping/datos” de las muestras en la carpeta del proyecto:

```
qiime taxa barplot \  
--i-table data/table.qza \  
--i-taxonomy data/taxonomy_dada.qza \  
--m-metadata-file mapping.txt \  
--o-visualization data/taxa-bar-plots.qzv
```

Buscaremos los siguientes archivos dentro de las carpetas que hemos generado al principio:

- La tabla con las asignaciones taxonómicas: data/taxa-barplots.qsv/data/level-7.csv
- La tabla con las secuencias agrupadas en ASVs: data/table.qza/data/feature-table.biom*

*En caso de no poder trabajar con la tabla biom lo ideal es extraer el archivo y ejecutar el comando en la carpeta donde se haya extraído:

```
biom convert -i feature-table.biom -o table.from_biom.txt --to-tsv
```

Con estos 2 archivos ya podemos ir Windows o continuar trabajando en Linux.

10- Damos formato a los archivos para el próximo script en EXCEL:

A) table.from_biom.txt

Abrimos el archive en EXCEL y eliminamos la primera fila.

Remplazamos “#OTU ID” por Subject

Guardamos con el nombre ASV.txt

Ejemplo:

Subject	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
01348dc4941f7b23ffecd89c9bfe6017	797	805	3499	3839	600
68f6ccd194c0d19027e167cafbd241ae	1773	1620	2454	3216	633
fb516f4f9e2cbe67ec17dc74d289770f	344	1056	3515	1699	630
0e93772fd31f97b68d540667b12c5c06	362	211	264	0	497
f75c563c8dc652f6bf5b363541cc01b2	799	485	484	1684	392
c5c02db5f958a29f4b17be41e9e88a85	1504	2109	1807	261	1059
7bb6ec696353e67a33c443f2af958b93	0	2827	1600	4637	0
d221a82b1aac065cc671149baeaad8ad	1089	492	49	624	190
fca18bb04fe7628d3675805a4e651c46	362	74	359	881	48

B) level-7.csv

El objetivo es extraer la información del número de secuencias por género y filos de este archivo. Por experiencia, no es óptimo utilizar exclusivamente los archivos de niveles taxonómicos superiores ya que solo te informan de las secuencias cuya asignación taxonómica se haya podido lograr obteniendo un número diferente de secuencias en el archivo de filos y el de géneros. Utilizar este archivo nos asegura tener todas las secuencias en un único archivo pero dificulta su procesado.

Abrimos el archivo en EXCEL

Comprobamos que cada asignación taxonómica ocupe un única celda y trasponemos, si fuese necesario, para tener las asignaciones en la primera columna y que cada columna sea ocupada por una única muestra.

Anexo I

Copiamos la primera columna en una hoja nueva y extraemos los géneros.

Ejemplo:

Es más fácil si se utiliza el método de reemplazar la parte común de la asignación a nivel de género “g__” por “|” o algún otro símbolo especial y separamos por dicho símbolo.

Esa columna de géneros tendrá celdas en blanco, correspondientes a aquellas identificaciones que no haya sido capaz de realizar el clasificador a nivel de género y clasificaciones como “*Candidatus*”, las reemplazamos por género_no_clasificado o “Género no clasificado”.

Con la nueva columna hay que realizar 2 acciones: Reemplazar la columna de la primera hoja y crear una nueva columna habiendo eliminado los duplicados con la herramienta de EXCEL pegándola justo debajo.

Ejemplo: remplazo de la columna original por la columna con los géneros extraídos:

Subject	Muestra_1	Muestra_2	Muestra_3	Muestra_4	Muestra_5
d_Bacteria;p_Campylobacteri	797	805	3499	3839	600
Unassigned_	1773	1620	2454	3216	633
d_Archaea;p_Euryarchaeota;c	344	1056	3515	1699	630
d_Bacteria;p_Actinobacteri	362	211	264	0	497
d_Bacteria;p_Actinobacteri	799	485	484	1684	392
d_Bacteria;p_Actinobacteri	1504	2109	1807	261	1059
d_Bacteria;p_Actinobacteri	0	2827	1600	4637	0
d_Bacteria;p_Actinobacteri	1089	492	49	624	190
d_Bacteria;p_Actinobacteri	362	74	359	881	48
d_Bacteria;p_Actinobacteri	314	351	1021	454	381

Subject	Muestra_1	Muestra_2	Muestra_3	Muestra_4	Muestra_5
Helicobacter	797	805	3499	3839	600
Unclassified_genera	1773	1620	2454	3216	633
Methanobrevibacter	344	1056	3515	1699	630
Methanosphaera	362	211	264	0	497
Actinomyces	799	485	484	1684	392
Actinomyces	1504	2109	1807	261	1059
Actinomyces	0	2827	1600	4637	0
Arcanobacterium	1089	492	49	624	190
Unclassified_genera	362	74	359	881	48
Mobiluncus	314	351	1021	454	381

Subject	Muestra_1	Muestra_2	Muestra_3	Muestra_4	Muestra_5
Helicobacter	797	805	3499	3839	600
Unclassified_genera	1773	1620	2454	3216	633
Methanobrevibacter	344	1056	3515	1699	630
Methanosphaera	362	211	264	0	497
Actinomyces	799	485	484	1684	392
Actinomyces	1504	2109	1807	261	1059
Actinomyces	0	2827	1600	4637	0
Arcanobacterium	1089	492	49	624	190
Unclassified_genera	362	74	359	881	48
Mobiluncus	314	351	1021	454	381
Subject	Muestra_1	Muestra_2	Muestra_3	Muestra_4	Muestra_5
Helicobacter	=SUMAR.SI(\$C\$4:\$C\$13;\$C17;D\$4:D\$13)				
Unclassified_genera	SUMAR.SI(rango; criterio; [rango_suma])				
Methanobrevibacter					
Methanosphaera					
Actinomyces					
Arcanobacterium					
Mobiluncus					

La expresión para sumar todas las celdas de cada columna en función del género es “SUMAR.SI”.

Copiamos la tabla nueva y pegamos valores en una hoja nueva desde la celda A1 y guardamos como texto delimitado por tabulaciones con el nombre de “genera.txt”

Repetimos el proceso a nivel de filo para obtener el archivo “phyla.txt”

Con las dos columnas que obtuvimos con los filios y con los géneros antes de quitarles los duplicados podemos obtener una lista de los géneros y los filios de nuestro estudio quedando relacionados y la llamamos "lista_filios_generos.txt".

Anexo I

Ejemplo:

Filos	Generos
Filo no clasificado	Género no clasificado
Euryarchaeota	Methanobrevibacter
Euryarchaeota	Methanosphaera
Actinobacteriota	Actinomyces
Actinobacteriota	Arcanobacterium
Actinobacteriota	Mobiluncus
Actinobacteriota	Varibaculum
Actinobacteriota	Alloscardovia
Actinobacteriota	Bifidobacterium
Actinobacteriota	Corynebacterium
Actinobacteriota	Brevibacterium
Actinobacteriota	Rothia
Actinobacteriota	Propionibacterium
Actinobacteriota	Atopobium
Actinobacteriota	Coriobacteriaceae
Actinobacteriota	Olsenella
Actinobacteriota	Collinsella
Actinobacteriota	Enorma

Esta tabla nos ayudará a poder realizar una serie de análisis que necesitan relacionar cada género con sus filos.

Finalmente tendremos 5 archivos diferentes que debemos situar en la misma carpeta:

- metadata.txt: Con la información de cada una de las muestras
- ASV.txt: Con la agrupación de las secuencias en formato ASVs
- genera.txt: Con las agrupaciones de las secuencias por géneros
- phyla.txt: Con las agrupaciones de las secuencias por filos
- lista_filos_generos.txt: que nos relaciona los géneros con cada uno de los filos.

Anexo II

Anexo II

El siguiente anexo contiene el script para realizar los análisis de secuencias una vez han sido formateadas con los comandos del Anexo1.

Antes de empezar es necesario disponer de los 5 archivos antes mencionados, disponer del programa Rstudio e instalar la primera vez las siguientes librerías:

```
library(ggbiplot), library(forcats), library(RColorBrewer), library(reshape2),
library(ggplot2), library(vegan), library(viridis), library(FactoMineR),
library(factoextra), library(ggpubr), library(gplots), library(magrittr), library(venn),
library(ggsignif), library(tidyverse), library(gtools), library(knitr), library(qgraph),
library(pairwiseAdonis).
```

Posteriormente solo hay que modificar el directorio en el script a la carpeta donde se localicen los otros 5 archivos

```
"directory<-"C:/Users/user/Desktop/qiime/proyecto/ ""
```

Para cargar el *scrip* en Rmarkdown solo es preciso copiar todo en un nuevo archivo Rmarkdown HTML borrando todo su contenido y pegando el script con sus comentarios que comienza a continuación:

```
<!-- Esto es una manera de representar los comentarios en Markdown. Se pueden
correr sin que repercuta en el programa -->
```

```
<!--Comienza el script. Módulo de carga e inicio -->
```

```
---
```

```
<!--Título y encabezado del proyecto con fecha y formato de salida HTML -->
```

```
title: "Proyecto Tesis Claudio"
```

```
author: "Claudio"
```

```
date: "`r format(Sys.time(), '%d %B, %Y')`"
```

```
output: html_document
```

```
---
```

```
---
```

```
####Packages:
```

```
<!-- Librerías de R utilizadas y su version. Hay que tenerlas instaladas previamente
mediante el comando install.packages("Nombre_de_la_libreria")-->
```

Anexo II

```
- library(ggbiplot) Version: `r packageVersion("ggbiplot")`  
- library(forcats) Version: `r packageVersion("forcats")`  
- library(RColorBrewer) Version: `r packageVersion("RColorBrewer")`  
- library(reshape2) Version: `r packageVersion("reshape2")`  
- library(ggplot2) Version: `r packageVersion("ggplot2")`  
- library(vegan) Version: `r packageVersion("vegan")`  
- library(viridis) Version: `r packageVersion("viridis")`  
- library(FactoMineR) Version: `r packageVersion("FactoMineR")`  
- library(factoextra) Version: `r packageVersion("factoextra")`  
- library(ggpubr) Version: `r packageVersion("ggpubr")`  
- library(gplots) Version: `r packageVersion("gplots")`  
- library(magrittr) Version: `r packageVersion("magrittr")`  
- library(venn) Version: `r packageVersion("venn")`  
- library(ggsignif) Version: `r packageVersion("ggsignif")`  
- library(tidyverse) Version: `r packageVersion("tidyverse")`  
- library(gtools) Version: `r packageVersion("gtools")`  
- library(knitr) Version: `r packageVersion("knitr")`  
- library(pairwiseAdonis) Version: `r packageVersion("pairwiseAdonis")`
```

```
``{r, warning=FALSE,message=FALSE, echo=FALSE}
```

```
<!-- Carga de las librerías en R-->
```

```
library(ggbiplot)
```

```
library(forcats)
```

```
library(RColorBrewer)
```

```
library(reshape2)
```

```
library(ggplot2)
```

```
library(vegan)
```

```
library(viridis)
```

```
library(FactoMineR)
```

```
library(factoextra)
```

```
library(ggpubr)
library(gplots)
library(magrittr)
library(venn)
library(ggsignif)
library(tidyverse)
library(gtools)
library(knitr)
library(qgraph)
library(pairwiseAdonis)
...
```

<!--Modulo 2: Definición del directorio de trabajo y el nombre de los archivos. Los archivos solicitados por defecto son 5. Tabla a nivel de ASV, a nivel de género, a nivel de filo lista con los géneros y su respectivo filo y la tabla de variables o "mapping file" -->

```
##ASV Samples resumen

``{r, echo=FALSE}

directory<-"C:/Users/Claudio/Desktop/qiime/proyecto/ "
setwd(directory)

#Nombre del archivo con las abundancias absolutas de las MATs, OTUs o ASVs
data<- read.table("ASV.txt", sep = "\t", header = TRUE, row.names = 1, check.names =
F)

md1<- read.table("metadata.txt", header=T, sep = "\t", row.names = 1, comment.char
= "*",check.names = F )

filo_genero<- read.table("lista_filos_generos.txt", sep = "\t", header = TRUE,
check.names = F)

#Variable a analizar
var<-"Biopsia_Vs_Jugo"
```

Anexo II

```
#Orden de los factores de var?
ordenados<- "No"

if (ordenados=="Si"){
md1[[var]] <- factor(md1[[var]], levels = c("Young", "Older"))
niveles <- c("Young", "Older")
}

#quitamos ASVs sin representación
ASV <- rowSums(data)
sin_0<-ASV!=0
data<-data[sin_0,]
...

<!--Resumen descriptivo -->
Number of samples: `r ncol(data)`

Number of ASVs: `r nrow(data)`

Number of reads: `r sum(apply(data, 2, sum))`

__Summary__
``{r, echo=FALSE}
data2<-data
data2<-as.matrix(t(data2))

data_filtrado<-data2
md1variable<- cbind(row.names(md1),as.character(md1[[var]]))
md1variable<-`colnames<-`(md1variable, c("muestra", "factor"))
md1variable<- as.data.frame(md1variable)
md1variable<-`row.names<-`(md1variable, md1variable$muestra)
md1variable[,1]<-NULL
```

```
a<-merge(data_filtrado,md1variable,by=0)
```

```
Factor<-c()
```

```
Samples<-c()
```

```
Seq_total<-c()
```

```
Min<-c()
```

```
Mediana<-c()
```

```
Q1<-c()
```

```
Q3<-c()
```

```
Max<-c()
```

```
for (i in unique (a$factor)){  
  var_i<-subset(a, factor == i)  
  var_i<-select(var_i, -factor, -Row.names)  
  sumar<-summary(apply(var_i, 1, sum))  
  Samples_temp<-nrow(var_i)  
  Seq_temp<-sum(var_i)  
  Factor_temp<-c(i)  
  Min_temp<-sumar[1]  
  Mediana_temp<-sumar[3]  
  Q1_temp<-sumar[2]  
  Q3_temp<-sumar[5]  
  Max_temp<-sumar[6]  
  Factor<-c(Factor,Factor_temp)  
  Samples<-c(Samples, Samples_temp)  
  Seq_total<-c(Seq_total, Seq_temp)  
  Min<-c(Min,Min_temp)  
  Mediana<-c(Mediana,Mediana_temp)  
  Q1<-c(Q1,Q1_temp)  
  Q3<-c(Q3,Q3_temp)
```

Anexo II

```
Max<-c(Max,Max_temp)
}

var_i<-select(a, -factor, -Row.names)
sumar<-summary(apply(var_i, 1, sum))
Samples_temp<-nrow(var_i)
Seq_temp<-sum(var_i)
Factor_temp<-c("Total")
Min_temp<-sumar[1]
Mediana_temp<-sumar[3]
Q1_temp<-sumar[2]
Q3_temp<-sumar[5]
Max_temp<-sumar[6]
Factor<-c(Factor,Factor_temp)
Samples<-c(Samples, Samples_temp)
Seq_total<-c(Seq_total, Seq_temp)
Min<-c(Min,Min_temp)
Mediana<-c(Mediana,Mediana_temp)
Q1<-c(Q1,Q1_temp)
Q3<-c(Q3,Q3_temp)
Max<-c(Max,Max_temp)

tablaguay<-cbind(Factor,Samples,Seq_total,Min,Mediana,Q1,Q3,Max)
kable(`row.names<-`(tablaguay,NULL))

...

<!-- Configuración actual del script-->
- Alfadiversity
  - Boxplots shannon
  -Boxplots Simpson
```

- 2 factors: Wilcoxon bonferroni adjusted
- 3 or more factors: Kruskal-Wallis bonferroni adjusted
- Pairwise comparisons using Wilcoxon rank sum test adjusted with Bonferroni.

- Betadiversity
 - Bray-Curtis distance matrix with Adonis test
 - Dispersion to the centroid study
 - Binnary-Jaccard distance matrix with Adonis test
 - Dispersion to the centroid study

-Diagramas de venn y heatmaps

##Results:

```
``{r, echo=FALSE}
```

```
data2<-data
```

```
data2<-as.matrix(data2)
```

```
ft<-prop.table(data2, 2)*100
```

```
tft<-t(ft)
```

```
orden<-order(apply(tft, 2, sum)/100, decreasing = T)
```

```
datos<-tft[,orden]
```

```
...
```

<!--Modulo 3: Inicio de los análisis de la diversidad alfa y beta usando los archivos: 1 tabla ASVs, OTUs o MATs y el "mapping file" -->

```
###Alfiversity
```

Anexo II

```
####Shannon and Simpson
```

```
`{r, echo=FALSE}
```

```
Simpson_Index<-diversity(datos, index="simpson")
```

```
Shannon_Index<-diversity(datos, index="shannon")
```

```
indices<-data.frame(Shannon_Index, Simpson_Index)
```

```
md1variable<- cbind(row.names(md1),as.character(md1[[var]]))
```

```
md1variable<-`colnames<-`(md1variable, c("muestra", "factor"))
```

```
md1variable<- as.data.frame(md1variable)
```

```
md1variable<-`row.names<-`(md1variable, md1variable$muestra)
```

```
md1variable[,1]<-NULL
```

```
a<-merge(indices,md1variable,by=0)
```

```
if (ordenados=="Si"){
```

```
  a$factor<- factor(a$factor, levels = niveles)
```

```
}
```

```
a$Row.names<-NULL
```

```
p1<-ggplot(a, aes(x=factor, y=Shannon_Index, fill= factor)) + geom_boxplot() +  
scale_fill_viridis(discrete = TRUE) + stat_compare_means(aes(group = factor),label.y =  
max(Shannon_Index)+0.1, size= 5.3) + ylim (NA,max(Shannon_Index)+0.2) +  
theme(axis.text=element_text(size=20), axis.title=element_text(size=20)) + ggtitle("")
```

```
p2<-ggplot(a, aes(x=factor, y=Simpson_Index, fill= factor)) + geom_boxplot() +
scale_fill_viridis(discrete = TRUE) + stat_compare_means(aes(group = factor),label.y =
max(Simpson_Index)+0.04, size= 5.3)+ ylim (NA,max(Simpson_Index)+0.05) +
theme(axis.text=element_text(size=20), axis.title=element_text(size=20)) + ggtitle("")
```

```
#p2<-ggplot(a, aes(x=factor, y=simpson, fill= factor)) + geom_boxplot() +
scale_fill_viridis(discrete = TRUE)
```

```
...
```

```
```{r, echo=FALSE, fig.height = 5, fig.width = 15, fig.align = "center"}
ggarrange(p1, p2 , labels = c(" Shannon", " Simpson"),ncol = 2, nrow = 1)
```

```
...
```

```
#####Estadisticos
```

```
```{r, echo=FALSE, warning=FALSE}
```

```
Factor<-c()
```

```
Div.Index<-c()
```

```
Samples<-c()
```

```
Min<-c()
```

```
Mediana<-c()
```

```
Q1<-c()
```

```
Q3<-c()
```

```
Max<-c()
```

```
for (i in unique (a$factor)){
```

```
  var_i<-subset(a, factor == i)
```

Anexo II

```
ncol(var_i)
for (l in 1: (ncol(var_i)-1)) {
  Factor_temp<-c(i)
  Filo_temp<-colnames(var_i)[l]
  Frec.detec_temp<-paste0(sum(var_i[,l]>0))
  Min_temp<-round(summary(var_i[,l])[1],2)
  Mediana_temp<-round(summary(var_i[,l])[3],2)
  Q1_temp<-round(summary(var_i[,l])[2],2)
  Q3_temp<-round(summary(var_i[,l])[5],2)
  Max_temp<-round(summary(var_i[,l])[6],2)
  Factor<-c(Factor,Factor_temp)
  Div.Index<-c(Div.Index,Filo_temp)
  Samples<-c(Samples,Frec.detec_temp)
  Min<-c(Min,Min_temp)
  Mediana<-c(Mediana,Mediana_temp)
  Q1<-c(Q1,Q1_temp)
  Q3<-c(Q3,Q3_temp)
  Max<-c(Max,Max_temp)
}

}

tablaguay<-cbind(Factor,Div.Index,Samples,Min,Mediana,Q1,Q3,Max)
tablaguay<-as.data.frame(tablaguay)
levels(tablaguay$Min) <- c(levels(tablaguay$Min), "<0.01")
levels(tablaguay$Mediana)<-c(levels(tablaguay$Mediana),"<0.01")
levels(tablaguay$Q1)<-c(levels(tablaguay$Q1),"<0.01")
levels(tablaguay$Q3)<-c(levels(tablaguay$Q3),"<0.01")
levels(tablaguay$Max)<-c(levels(tablaguay$Max),"<0.01")

tablaguay[tablaguay == "0"] <- "<0.01"
kable(`row.names<-`(tablaguay,NULL))
```

```

...

```{r, echo=FALSE}
b<-colnames(a[,-3])
resultados<-lapply(a[,b], function(x) pairwise.wilcox.test(x , g=a$factor,
p.adjust.method = "bonf"))
resultados
...

####Betadiversity

```{r, echo=FALSE}
#res.pca <- prcomp(datos)
#
#p1<-fviz(res.pca, geom="point", element="ind",axes = c(1, 2) ,label="none",
habillage=a$factor, addEllipses=TRUE, ellipse.level=0.8, invisible="quali")
#p2<-fviz(res.pca, geom="point", element="ind",axes = c(1, 3) ,label="none",
habillage=a$factor, addEllipses=TRUE, ellipse.level=0.8, invisible="quali")
#p3<-fviz(res.pca, geom="point", element="ind",axes = c(2, 3) ,label="none",
habillage=a$factor, addEllipses=TRUE, ellipse.level=0.8, invisible="quali")
...
```{r, echo=FALSE, fig.height = 7, fig.width = 15, fig.align = "center"}
#ggarrange(p1, p2, p3 + rremove("x.text"), labels = c("A", "B", "C"),ncol = 3, nrow = 1)

...

####PCoA

#####Bray-Courtis

```{r, echo=FALSE,warning=FALSE}
a$factor<-droplevels(a$factor)
datos.dist<-vegdist(datos, method='bray')

```

Anexo II

```
datos.div<-adonis2(datos~a$factor, permutations = 1000, method="bray", by= NULL)

dispersion<-betadisper(datos.dist, group=a$factor)

...

``{r, echo=FALSE, fig.height = 7, fig.width = 10, fig.align = "center", warning=FALSE}
labs<- paste("Dimension", 1:4, "(", round(100*dispersion$eig / sum(dispersion$eig), 2),
"%)")

par(mfrow=c(2,3), xpd = F, mar = par()$mar + c(0,0,0,7))

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(1,2), xlab=labs[1], ylab=labs[2],
conf=0.50) ##sd ellipse

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(1,3), xlab=labs[1], ylab=labs[3],
conf=0.50) ##sd ellipse

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(2,3), xlab=labs[2], ylab=labs[3],
conf=0.50) ##sd ellipse

# Si molesta la etiqueta del centroide usar el de abajo

#par(mfrow=c(2,3), xpd = T, mar = par()$mar + c(0,0,0,7))

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(1,2), xlab=labs[1], ylab=labs[2],
conf=0.50, label= FALSE) ##sd ellipse

legend("bottomright",inset=c(-0.6,0),pch=c(1:length(levels(a$factor))),bty = "n",col =
c(1:length(levels(a$factor))),xpd=TRUE, legend=levels(a$factor))

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(1,3), xlab=labs[1], ylab=labs[3],
conf=0.50, label= FALSE) ##sd ellipse

legend("bottomright",inset=c(-0.6,0),pch=c(1:length(levels(a$factor))),bty = "n",col =
c(1:length(levels(a$factor))),xpd=TRUE, legend=levels(a$factor))

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(2,3), xlab=labs[2], ylab=labs[3],
conf=0.50, label= FALSE) ##sd ellipse

legend("bottomright",inset=c(-0.6,0),pch=c(1:length(levels(a$factor))),bty = "n",col =
c(1:length(levels(a$factor))),xpd=TRUE, legend=levels(a$factor))

...

```

```

``{r, echo=FALSE}
print(datos.div)
...

``{r, echo=FALSE}
print(pairwise.adonis(datos.dist,a$factor, perm= 1000,p.adjust.m = "holm" ))
...

#####Bray-Courtis (dispersion study)

``{r, echo=FALSE, fig.height = 3, fig.width = 3, fig.align = "center",warning=FALSE}
boxplot(dispersion, axes = c(1,3))

...

``{r, echo=FALSE}
permutest(dispersion, pairwise = T , permutations = 1000)
...

-----

#####Binnary-Jaccard

``{r, echo=FALSE,warning=FALSE}
a$factor<-droplevels(a$factor)
datos.dist<-vegdist(datos, method="jaccard", binary = T)
datos.div<-adonis2(datos.dist~a$factor, permutations = 1000, method="jaccard",
strata="PLOT")
dispersion<-betadisper(datos.dist, group=a$factor)
...

``{r, echo=FALSE, fig.height = 7, fig.width = 10, fig.align = "center",warning=FALSE}
labs<- paste("Dimension", 1:4, "(", round(100*dispersion$eig / sum(dispersion$eig), 2),
"%)")

```

Anexo II

```
par(mfrow=c(2,3), xpd = T, mar = par()$mar + c(0,0,0,7))

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(1,2), xlab=labs[1], ylab=labs[2],
conf=0.50) ##sd ellipse

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(1,3), xlab=labs[1], ylab=labs[3],
conf=0.50) ##sd ellipse

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(2,3), xlab=labs[2], ylab=labs[3],
conf=0.50) ##sd ellipse

# Si molesta la etiqueta del centroide el de abajo

#par(mfrow=c(2,3), xpd = T, mar = par()$mar + c(0,0,0,7))

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(1,2), xlab=labs[1], ylab=labs[2],
conf=0.50, label= FALSE) ##sd ellipse

legend("bottomright",inset=c(-0.6,0),pch=c(1:length(levels(a$factor))),bty = "n",col =
c(1:length(levels(a$factor))),xpd=TRUE, legend=levels(a$factor))

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(1,3), xlab=labs[1], ylab=labs[3],
conf=0.50, label= FALSE) ##sd ellipse

legend("bottomright",inset=c(-0.6,0),pch=c(1:length(levels(a$factor))),bty = "n",col =
c(1:length(levels(a$factor))),xpd=TRUE, legend=levels(a$factor))

plot(dispersion, hull=FALSE, ellipse=TRUE, axes = c(2,3), xlab=labs[2], ylab=labs[3],
conf=0.50, label= FALSE) ##sd ellipse

legend("bottomright",inset=c(-0.6,0),pch=c(1:length(levels(a$factor))),bty = "n",col =
c(1:length(levels(a$factor))),xpd=TRUE, legend=levels(a$factor))

...

``{r, echo=FALSE}

print(datos.div)

...

``{r, echo=FALSE}

print(pairwise.adonis(datos.dist,a$factor, perm= 10000, p.adjust.m = "holm"))

...
```

```
#####Binnary-Jaccard (dispersion study)
```

```
``{r, echo=FALSE, fig.height = 3, fig.width = 3, fig.align = "center",warning=FALSE}
```

```
boxplot(dispersion, axes = c(1,3), las=1)
```

```
...
```

```
``{r, echo=FALSE}
```

```
permutest(dispersion, permutations = 10000, pairwise = T)
```

```
...
```

<!-- Módulo 4: Estudio de los filios más abundantes. Está configurado para seleccionar los 5 mayoritarios y juntar el resto en filios minoritarios. Los archivos necesarios son la tabla de filios y el "mapping file"-->

```
###Most abundant phyla study
```

```
``{r, echo=FALSE}
```

```
setwd(directory)
```

```
data<- read.table("phyla.txt", sep = "\t", header = TRUE, row.names = 1, check.names  
= F)
```

```
...
```

```
Numero de muestras: `r ncol(data)`
```

```
Numero de filios identificados: `r nrow(data)`
```

Anexo II

```
``{r, echo=FALSE}
data2<-data
data2<-as.matrix(data2)
ft<-prop.table(data2, 2)*100
tft<-t(ft)
orden<-order(apply(tft, 2, sum)/100, decreasing = T)
datos<-tft[,orden]
data_filtrado2<-datos[,1:5]
Minor_phyla<-100-apply(data_filtrado2, 1, sum)
data_filtrado1<-data.frame(data_filtrado2,Minor_phyla)
```

```
data_filtrado<-data_filtrado1
data_filtrado[data_filtrado<0] <- 0
```

```
...
```

####Boxplots

<!-- Importante: El script precisa una abundancia minima en cada comparación por pares para realizar el estadístico (al menos una secuencia por par comparado). En caso de no haber esa abundancia mínima saltará error y habrá que modificar el script por lo que te diga el programa-->

```
``{r, echo=FALSE , fig.width=15, fig.height=10, warning=FALSE}
md1variable<- cbind(row.names(md1),as.character(md1[[var]]))
md1variable<-`colnames<-`(md1variable, c("muestra", "factor"))
md1variable<- as.data.frame(md1variable)
md1variable<-`row.names<-`(md1variable, md1variable$muestra)
md1variable[,1]<-NULL
```

```
a<-merge(data_filtrado,md1variable,by=0)

a$Row.names<-NULL
if (ordenados=="Si"){
a$factor<- factor(a$factor, levels = niveles)
}
df.m <- melt(a, id.var = "factor")

b<-colnames(a[,1:6])

ajustando<-lapply(a[,b], function(x) pairwise.wilcox.test(x , g=a$factor,
p.adjust.method = "bonf")$p.value)

ajuste<-na.omit(unlist(ajustando, use.names=FALSE))

df.m$factor<-factor(df.m$factor)

anno_df2 = compare_means(value ~ factor, group.by = "variable", data = df.m, method
= "wilcox.test", p.adjust.method= "bonferroni") %>% mutate(y_pos = rep(seq(130, 100, by = -
(20/nrow(combinations(length(levels(df.m$factor)),2))))[1:nrow(combinations(length(levels(df
.m$factor)),2)],length(levels(df.m$variable))))

#Posible error si ninguna deteccion cambiar el 1:19 al numero que corresponda

#anno_df2 = compare_means(value ~ factor, group.by = "variable", data = df.m,
method = "wilcox.test", p.adjust.method= "bonferroni") %>% mutate(y_pos = rep(seq(130, 50,
by = -(80/nrow(combinations(length(levels(df.m$factor)),2))))[1:84]))

anno_df2$p.adj<-round(ajuste, digits = 4)

gr <- anno_df2$p.adj <= 0.051
```

Anexo II

```
anno_df2$p.adj2 <- cut(anno_df2$p.adj, breaks=c(-1, 0.001, 0.01, 0.05, 1),
labels=c("***", "**", "*", "ns"))

p <- ggplot(data = df.m, aes(x=factor, y=value))
p <- p + geom_boxplot(aes(fill = factor), size=0.5 )
p <- p + geom_point(aes(y=value, group=factor), position =
position_dodge(width=0.75))
p <- p + facet_wrap( ~ variable, scales="free", nrow = 1)
p <- p + xlab("") + ylab("") + ggtitle("")
p <- p + guides(fill=guide_legend(title="Legend"))
p <- p + scale_fill_viridis(discrete = TRUE)
p <- p + ylim (0,135)
#p <- p + theme(axis.title.y = element_text(size = rel(1.8), angle = 90))
#p <- p + theme(axis.title.x = element_text(size = rel(1.8), angle = 00))
p <- p + theme (axis.text = element_text(size = 00, angle = 45, hjust = 1), axis.text.y =
element_text(size = 15,angle = 00), title = element_text(face = "italic"), plot.margin =
unit(c(0,0,0,0),"cm"))
p <- p + stat_compare_means(aes(group = factor),label.y = 135)

if (any(gr==TRUE)){
  p <- p + geom_signif(data=anno_df2[gr,],aes(xmin=group1,hide.ns= TRUE
,xmax=group2, annotations=p.adj2, y_position=y_pos), manual=TRUE)
}

p

...

``{r, echo=FALSE, warning=FALSE}
Factor<-c()
Filo<-c()
Frec.detec<-c()
```

```
Min<-c()
Mediana<-c()
Q1<-c()
Q3<-c()
Max<-c()

for (i in unique (a$factor)){
  var_i<-subset(a, factor == i)
  ncol(var_i)
  for (l in 1: (ncol(var_i)-1)) {
    Factor_temp<-c(i)
    Filo_temp<-colnames(var_i)[l]
    Frec.detec_temp<-paste0(sum(var_i[,l]>0) ,"
(",round(100*(sum(var_i[,l]>0))/nrow(var_i), 2),"|%)", "")
    Min_temp<-round(summary(var_i[,l])[1],2)
    Mediana_temp<-round(summary(var_i[,l])[3],2)
    Q1_temp<-round(summary(var_i[,l])[2],2)
    Q3_temp<-round(summary(var_i[,l])[5],2)
    Max_temp<-round(summary(var_i[,l])[6],2)
    Factor<-c(Factor,Factor_temp)
    Filo<-c(Filo,Filo_temp)
    Frec.detec<-c(Frec.detec,Frec.detec_temp)
    Min<-c(Min,Min_temp)
    Mediana<-c(Mediana,Mediana_temp)
    Q1<-c(Q1,Q1_temp)
    Q3<-c(Q3,Q3_temp)
    Max<-c(Max,Max_temp)
  }
}
```

Anexo II

```
}  
tablaguay<-cbind(Factor,Filo,Frec.detec,Min,Mediana,Q1,Q3,Max)  
tablaguay<-as.data.frame(tablaguay)  
levels(tablaguay$Min) <- c(levels(tablaguay$Min), "<0.01")  
levels(tablaguay$Mediana)<-c(levels(tablaguay$Mediana),"<0.01")  
levels(tablaguay$Q1)<-c(levels(tablaguay$Q1),"<0.01")  
levels(tablaguay$Q3)<-c(levels(tablaguay$Q3),"<0.01")  
levels(tablaguay$Max)<-c(levels(tablaguay$Max),"<0.01")  
  
tablaguay[tablaguay == "0"] <- "<0.01"  
  
tablaguay$medianIQR<-paste(tablaguay$Mediana, " (", tablaguay$Q1, "-",  
tablaguay$Q3,")", sep="")  
  
kable(`row.names<-`(tablaguay,NULL))  
  
  
tabla2<-as.character(unique(tablaguay$Filo))  
nombre<- c("Filo")  
for (i in levels(tablaguay$Factor)) {  
  tabla_temp<-tablaguay$Factor== i  
  tabla_temp2<-tablaguay[tabla_temp,]  
  Frec.detec_temp<-as.character(tabla_temp2$Frec.detec)  
  mediana_tempo<-tabla_temp2$medianIQR  
  nom1<-paste("frec.detec_", i, sep = "")  
  nom2<-paste("medianIQR_", i, sep = "")  
  nombre<-c(nombre,nom1,nom2)  
  tabla2<-cbind(tabla2,Frec.detec_temp,mediana_tempo )  
}  
tabla_filos<-`colnames<-`(tabla2, nombre)  
  
...
```

```
##### Test estadísticos

##### 1er Kruskal-Wallis

```{r, echo=FALSE, warning=FALSE}
b<-colnames(a[,1:6])

resultados<-lapply(a[,b], function(x) kruskal.test(x , g=a$factor, p.adjust.method =
"bonf"))

resultados2<-lapply(a[,b], function(x) pairwise.wilcox.test(x , g=a$factor,
p.adjust.method = "bonf"))

#resultados
...

2er Wilcoxon rank sum test

```{r, echo=FALSE, warning=FALSE}
#resultados2
...

###Heatmap

```{r, echo=FALSE , fig.width=15, fig.height=10, warning=FALSE}
a$factor<-droplevels(a$factor)

cols <- c("pink1", "brown", "palegreen4", "violet", "mediumpurple1", "slateblue1",
"purple", "purple3", "turquoise2", "skyblue", "steelblue", "blue2", "navyblue", "orange",
"tomato", "coral2", "palevioletred", "violetred", "red2", "springgreen2", "yellowgreen",
"wheat2", "tan", "tan2", "tan3", "grey70", "grey50", "grey30")

myvar <- a$factor
legend.col <- cols[1:length(levels(myvar))]
names(legend.col) <- levels(myvar)
```

## Anexo II

---

```
myRowSideColors <- cols[as.numeric(myvar)]

agrupamos muestras a parte
hr <- hclust(dist(data_filtrado), method="complete")

pintamos heatmap

p<-heatmap.2(as.matrix(data_filtrado), Rowv=as.dendrogram(hr), cexCol=1.2,
cexRow=0.7, col="terrain.colors", trace="none", dendrogram="row", Colv=FALSE,
RowSideColors=myRowSideColors, lhei=c(3, 8), margins=c(10, 20), xlab="Phyla",
ylab="Samples", key.xlab="Relative abundance (%)")

legend("topright", legend=names(legend.col), fill=legend.col)
...

###Venn diagram (Genus presented in 80% of the samples, in at least one cohort, with
3 minimum different relative abundance limits).

<pre>
 0.01% 0.1% 1%

</pre>

``{r, echo=FALSE, fig.height = 7, fig.width = 15, fig.align = "center",warning=FALSE}
par(mfrow=c(1,3))

a<-merge(datos,md1variable,by=0)
a$Row.names<-NULL

a$factor<-droplevels(a$factor)
if (length(levels(a$factor)) < 8){

unidas001<-list()
```

```

unidas01<-list()
unidas1<-list()
for (i in unique (a$factor)){
 var_i<-subset(a, factor == i)
 var_i <- subset(var_i, select = -factor)
 unidas001[[i]]<-apply(var_i, MARGIN=2, FUN=function(x)
if((100*length(which(x>0.01))/nrow(var_i))>=80) return(TRUE) else return(FALSE))
 unidas01[[i]]<-apply(var_i, MARGIN=2, FUN=function(x)
if((100*length(which(x>0.1))/nrow(var_i))>=80) return(TRUE) else return(FALSE))
 unidas1[[i]]<-apply(var_i, MARGIN=2, FUN=function(x)
if((100*length(which(x>1))/nrow(var_i))>=80) return(TRUE) else return(FALSE))

}
unidas001 <- data.frame(unidas001)
unidas01 <- data.frame(unidas01)
unidas1 <- data.frame(unidas1)

venn(unidas001, zcolor = "style", cexil = 3, cexsn = 3)
venn(unidas01, zcolor = "style", cexil = 3, cexsn = 3)
venn(unidas1, zcolor = "style", cexil = 3, cexsn = 3)

vector001 <- apply(unidas001, 1, any)
vector01 <- apply(unidas01, 1, any)
vector1 <- apply(unidas1, 1, any)

print(unidas001 <- unidas001[vector001,])
print(unidas01 <- unidas01[vector01,])
print(unidas1 <- unidas1[vector1,])

} else {
 print ("Variable no apta para el venn (¿Muchos factores?)")

```

## Anexo II

---

```
}
```

```
...
```

<!-- Módulo 4: Estudio de los géneros más abundantes. Está configurado para seleccionar los 19 mayoritarios y juntar el resto en géneros minoritarios. Precisa de la tabla de géneros y el "mapping file"-->

```
###Most abundant genera study
```

```
``{r, echo=FALSE}
setwd(directory)
data<- read.table("genera.txt", sep = "\t", header = TRUE, row.names = 1, check.names
= F)
```

```
...
```

```
Number of genera detected: `r nrow(data)`
```

```
``{r, echo=FALSE}
data2<-data
data2<-as.matrix(data2)
ft<-prop.table(data2, 2)*100
tft<-t(ft)
orden<-order(apply(tft, 2, sum)/100, decreasing = T)
datos<-tft[,orden]
```

```
data_filtrado2<-datos[,1:19]
Minor_genera<-100-apply(data_filtrado2, 1, sum)
data_filtrado1<-data.frame(data_filtrado2,Minor_genera)
```

```
data_filtrado<-data_filtrado1
data_filtrado[data_filtrado<0] <- 0
...

```

#### ####Boxplots

<!-- Importante: El script precisa una abundancia mínima en cada comparación por pares para realizar el estadístico (al menos una secuencia por par comparado). En caso de no haber esa abundancia mínima saltará error y habrá que modificar el script por lo que te diga el programa-->

```
``{r, echo=FALSE , fig.width=25, fig.height=10, warning=FALSE}
md1variable<- cbind(row.names(md1),as.character(md1[[var]]))
md1variable<-`colnames<-`(md1variable, c("muestra", "factor"))
md1variable<- as.data.frame(md1variable)
md1variable<-`row.names<-`(md1variable, md1variable$muestra)
md1variable[,1]<-NULL
```

```
a<-merge(data_filtrado,md1variable,by=0)
```

```
a$Row.names<-NULL
if (ordenados=="Si"){
a$factor<- factor(a$factor, levels = niveles)
}
df.m <- melt(a, id.var = "factor")
```

## Anexo II

---

```
b<-colnames(a[,1:20])

ajustando<-lapply(a[,b], function(x) pairwise.wilcox.test(x , g=a$factor,
p.adjust.method = "bonf")$p.value)

ajuste<-na.omit(unlist(ajustando, use.names=FALSE))

df.m$factor<-factor(df.m$factor)

anno_df2 = compare_means(value ~ factor, group.by = "variable", data = df.m, method
= "wilcox.test", p.adjust.method= "bonferroni") %>% mutate(y_pos = rep(seq(130, 50, by = -
(80/nrow(combinations(length(levels(df.m$factor)),2))))[1:nrow(combinations(length(levels(df
.m$factor)),2)]),length(levels(df.m$variable))))

#Posible error si ninguna deteccion cambiar el 1: 287 al numero que corresponda

#anno_df2 = compare_means(value ~ factor, group.by = "variable", data = df.m,
method = "wilcox.test", p.adjust.method= "bonferroni") %>% mutate(y_pos = rep(seq(130, 50,
by = -(80/nrow(combinations(length(levels(df.m$factor)),2))))[1:287]))

anno_df2$p.adj<-round(ajuste, digits = 4)

gr <- anno_df2$p.adj <= 0.051

anno_df2$p.adj[anno_df2$p.adj<0.0001]<-0.0001

anno_df2$p.adj2 <- cut(anno_df2$p.adj, breaks=c(-1, 0.001, 0.01, 0.05, 1),
labels=c("****", "***", "**", "ns"))

p <- ggplot(data = df.m, aes(x=factor, y=value))
p <- p + geom_boxplot(aes(fill = factor), size=0.5)
p <- p + geom_point(aes(y=value, group=factor), position =
position_dodge(width=0.75))
p <- p + facet_wrap(~ variable, scales="free", nrow = 4)
p <- p + xlab("") + ylab("") + ggtitle("")
```

```

p <- p + guides(fill=guide_legend(title="Legend"))

p <- p + scale_fill_viridis(discrete = TRUE)

p <- p + ylim (0,150)

#p <- p + theme(axis.title.y = element_text(size = rel(1.8), angle = 90))

#p <- p + theme(axis.title.x = element_text(size = rel(1.8), angle = 00))

p <- p + theme (axis.text = element_text(size = 00, angle = 45, hjust = 1), axis.text.y =
element_text(size = 15,angle = 00), title = element_text(face = "italic"), plot.margin =
unit(c(0,0,0,0),"cm"))

p <- p + stat_compare_means(aes(group = factor),label.y = 150)

if (any(gr==TRUE)){

 p <- p + geom_signif(data=anno_df2[gr,],aes(xmin=group1,hide.ns= TRUE
,xmax=group2, annotations=p.adj2, y_position=y_pos), manual=TRUE)
}

p

...

``{r, echo=FALSE, warning=FALSE}

Factor<-c()

Genero<-c()

Frec.detec<-c()

Min<-c()

Mediana<-c()

Q1<-c()

Q3<-c()

Max<-c()

```

```
for (i in unique (a$factor)){
 var_i<-subset(a, factor == i)
 ncol(var_i)
 for (l in 1: (ncol(var_i)-1)) {
 Factor_temp<-c(i)
 Genero_temp<-colnames(var_i)[l]
 Frec.detec_temp<-paste0(sum(var_i[,l]>0) ,"
(",round(100*(sum(var_i[,l]>0))/nrow(var_i), 2),"|%" , ""))
 Min_temp<-round(summary(var_i[,l])[1],2)
 Mediana_temp<-round(summary(var_i[,l])[3],2)
 Q1_temp<-round(summary(var_i[,l])[2],2)
 Q3_temp<-round(summary(var_i[,l])[5],2)
 Max_temp<-round(summary(var_i[,l])[6],2)
 Factor<-c(Factor,Factor_temp)
 Genero<-c(Genero,Genero_temp)
 Frec.detec<-c(Frec.detec,Frec.detec_temp)
 Min<-c(Min,Min_temp)
 Mediana<-c(Mediana,Mediana_temp)
 Q1<-c(Q1,Q1_temp)
 Q3<-c(Q3,Q3_temp)
 Max<-c(Max,Max_temp)
 }
}

tablaguay<-cbind(Factor,Genero,Frec.detec,Min,Mediana,Q1,Q3,Max)
tablaguay<-as.data.frame(tablaguay)
levels(tablaguay$Min) <- c(levels(tablaguay$Min), "<0.01")
levels(tablaguay$Mediana)<-c(levels(tablaguay$Mediana),"<0.01")
levels(tablaguay$Q1)<-c(levels(tablaguay$Q1),"<0.01")
levels(tablaguay$Q3)<-c(levels(tablaguay$Q3),"<0.01")
levels(tablaguay$Max)<-c(levels(tablaguay$Max),"<0.01")
```

```

tablaguay[tablaguay == "0"] <- "<0.01"

tablaguay$medianIQR<-paste(tablaguay$Mediana," (", tablaguay$Q1, "-",
tablaguay$Q3,")", sep="")

kable(`row.names<-`(tablaguay,NULL))

tabla2<-as.character(unique(tablaguay$Genero))
nombre<- c("generos")
for (i in levels(tablaguay$Factor)) {
tabla_temp<-tablaguay$Factor== i
tabla_temp2<-tablaguay[tabla_temp,]
Frec.detec_temp<-as.character(tabla_temp2$Frec.detec)
mediana_tempo<-tabla_temp2$medianIQR
nom1<-paste("frec.detec_", i, sep = "")
nom2<-paste("medianIQR_", i, sep = "")
nombre<-c(nombre,nom1,nom2)
tabla2<-cbind(tabla2,Frec.detec_temp,mediana_tempo)
}
tabla_generos<-`colnames<-`(tabla2, nombre)

...

Test estadísticos

1er Kruskal-Wallis

```{r, echo=FALSE, warning=FALSE}
b<-colnames(a[,1:20])

```

Anexo II

```
resultados<-lapply(a[,b], function(x) kruskal.test(x , g=a$factor, p.adjust.method =  
"bonf"))
```

```
resultados2<-lapply(a[,b], function(x) pairwise.wilcox.test(x , g=a$factor,  
p.adjust.method = "bonf"))
```

```
#resultados
```

```
...
```

```
##### 2er Wilcoxon rank sum test
```

```
``{r, echo=FALSE, warning=FALSE}
```

```
#resultados2
```

```
...
```

```
###Heatmap
```

```
``{r, echo=FALSE , fig.width=15, fig.height=10, warning=FALSE}
```

```
a$factor<-droplevels(a$factor)
```

```
cols <- c("pink1", "brown", "palegreen4", "violet", "mediumpurple1", "slateblue1",  
"purple", "purple3", "turquoise2", "skyblue", "steelblue", "blue2", "navyblue", "orange",  
"tomato", "coral2", "palevioletred", "violetred", "red2", "springgreen2", "yellowgreen",  
"wheat2", "tan", "tan2", "tan3", "grey70", "grey50", "grey30")
```

```
myvar <- as.factor(a$factor)
```

```
legend.col <- cols[1:length(levels(myvar))]
```

```
names(legend.col) <- levels(myvar)
```

```
myRowSideColors <- cols[as.numeric(myvar)]
```

```
## agrupamos muestras a parte
```

```
hr <- hclust(dist(data_filtrado), method="complete")
```

```
## pintamos heatmap
```

```
p<-heatmap.2(as.matrix(data_filtrado), Rowv=as.dendrogram(hr), cexCol=1,
cexRow=0.7, col="terrain.colors", trace="none", dendrogram="row", Colv=FALSE,
RowSideColors=myRowSideColors, lhei=c(3, 8), margins=c(10, 20), xlab="Genera",
ylab="Samples", key.xlab="Relative abundance (%)")
```

```
legend("topright", legend=names(legend.col), fill=legend.col)
```

```
...
```

```
``{r, echo=FALSE, fig.height = 7, fig.width = 15, fig.align = "center",warning=FALSE,
error=TRUE}
```

```
Min<-c()
```

```
a$factor<-droplevels(a$factor)
```

```
for (i in unique (a$factor)){
```

```
  var_i<-subset(a, factor == i)
```

```
  var_i$factor<-NULL
```

```
  Min_temp<-nrow(var_i)
```

```
  Min<-c(Min,Min_temp)
```

```
}
```

```
Min2<-min(Min)-1
```

```
if (Min2>20){
```

```
  Min2<-19
```

```
}
```

```
generos2<-colnames(data_filtrado[,1:Min2])
```

```
colnames(filo_genero)<-c("Filos", "generos")
```

```
filo_genero$Match <- as.integer(filo_genero$generos %in% generos2)
```

```
bb<-with(filo_genero, which(Match ==TRUE))
```

```
filo_genero2<-filo_genero[bb,]
```

Anexo II

```
grupo_filo<-filo_genero2[match(generos2, filo_genero2$generos),]
grupo_filo$Filos<-droplevels(grupo_filo$Filos)

for (i in unique (a$factor)){
  print(i)
  var_i<-subset(a, factor == i)
  var_i$factor<-NULL
  correlacion <- var_i[,1:Min2] # Select only variables
  corMat <- cor_auto(correlacion)
  qqgraph(corMat, graph = "glasso", layout = "spring", edge.labels = F, tuning = 0.25,
sampleSize = nrow(correlacion), nodeNames = colnames(correlacion), groups =
grupo_filo$Filos, legend.cex = 0.3, palette = "pastel")
}

...

###Diagrama de Venn (Abundancia minima requerida en el 80% de las muestras)


```

 0.01% 0.1% 1%


```


```

```
``{r, echo=FALSE, fig.height = 7, fig.width = 15, fig.align = "center",warning=FALSE}
par(mfrow=c(1,3))

a<-merge(datos,md1variable,by=0)
a$Row.names<-NULL
a$factor<-droplevels(a$factor)
if (length(levels(a$factor)) < 8){
  unidas001<-list()
  unidas01<-list()
  unidas1<-list()
  for (i in unique (a$factor)){
```

```
var_i<-subset(a, factor == i)

var_i <- subset( var_i, select = -factor )

unidas001[[i]]<-apply(var_i, MARGIN=2, FUN=function(x)
if((100*length(which(x>0.01))/nrow(var_i))>=80) return(TRUE) else return(FALSE))

unidas01[[i]]<-apply(var_i, MARGIN=2, FUN=function(x)
if((100*length(which(x>0.1))/nrow(var_i))>=80) return(TRUE) else return(FALSE))

unidas1[[i]]<-apply(var_i, MARGIN=2, FUN=function(x)
if((100*length(which(x>1))/nrow(var_i))>=80) return(TRUE) else return(FALSE))

}

unidas001 <- data.frame(unidas001)
unidas01 <- data.frame(unidas01)
unidas1 <- data.frame(unidas1)

venn(unidas001, zcolor = "style", cexil = 3, cexsn = 3)
venn(unidas01, zcolor = "style", cexil = 3, cexsn = 3)
venn(unidas1, zcolor = "style", cexil = 3, cexsn = 3)

vector001 <- apply(unidas001, 1, any)
vector01 <- apply(unidas01, 1, any)
vector1 <- apply(unidas1, 1, any)

print(unidas001 <- unidas001[vector001,])
print(unidas01 <- unidas01[vector01,])
print(unidas1 <- unidas1[vector1,])

} else {
  print ("Variable no apta para el venn (¿Muchos factores?)")
}
```

Anexo II

...

<!-- Módulo 6: Utilizando la información adquirida en los módulos 4 y 5 procesa la información y prepara una tabla resumen de los resultado obtenidos. Precisa los módulos anteriores y la tabla de los géneros con los fillos-->

```
``{r, echo=FALSE, warning=FALSE}
```

```
generos2<-colnames(data_filtrado)
colnames(filo_genero)<-c("Filos", "generos")
filo_genero$Match <- as.integer(filo_genero$generos %in% generos2)
bb<-with(filo_genero, which(Match ==TRUE))
filo_genero2<-filo_genero[bb,]
grupo_filo<-filo_genero2[match(generos2, filo_genero2$generos),]
grupo_filo$Filos<-droplevels(grupo_filo$Filos)
grupo_filo<-grupo_filo[,1:2]
b<-data.frame(levels(grupo_filo$Filos), levels(grupo_filo$Filos))
b<-`colnames<-`(b,colnames(grupo_filo))
grupo_filo<-rbind(grupo_filo,b)[,1:2]

#####

tablafinal<-rbind(tabla_filos, tabla_generos)
unidos3<-merge(grupo_filo,tablafinal,by.x="generos",by.y="Filo", all =T)
kable(`row.names<-`(unidos3,NULL))
```

...

<!-- ----->

Anexo III

Anexo III

Al terminar de ejecutar el programa se puede ejecutar este comando para preparar los diagramas de barras apiladas a nivel de género. Es posible que sea necesario incorporar géneros nuevos a la lista que hemos incluido, en tal caso se procede en la lista colores con el color correspondiente en formato HEX entre comillas y separados por una coma y el género en la lista bichos de igual manera.

#Módulo extra: Este módulo se ejecuta en R. Son complementos que ayudan a la visualización de los resultados. Entre otras funciones tenemos los diagramas de barras apilados por factor y gráficos de correlaciones

#Barplot Generos

#Lista de bichos de filios y generos más abundantes, lo ideal es ir ampliando y/o actualizando colores

```
mycolors<-
c("#FBB4AE", "#B3CDE3", "#CCEBC5", "#DECBE4", "#FED9A6", "#FFFFCC", "#E5D8BD", "#FDDAEC",
"#F2F2F2", "#E41A1C", "#377EB8", "#4DAF4A", "#984EA3", "#FF7F00", "#91cc78", "#248f49", "#5e
74a6", "#FFFF33", "#A65628", "#F781BF", "#e8db92", "#999999", "#A6CEE3", "#1F78B4", "#999095
", "#8fbfb7", "#FB9A99", "#CC79A7", "#FDBF6F", "#FF7F00", "#CAB2D6", "#32a8a6", "#32a8a6", "#3
240a8", "#6349a6", "#8c80ab", "#b09a68", "#853072", "#8a3f2f", "#2c8238", "#389447", "#b844b0
", "#64a3a2", "#2f4037", "#daebe2", "#fae6e3", "#d9aa77", "#ced16f", "#a2cc68", "#5eb55c", "#5cb
5b5", "#5377a6", "#41437a", "#851e63", "#54511d", "#5b8549", "#d1d8f0", "#edecbe", "#f0efdd",
"#70533f", "#7d3d11", "#320f59", "#6b4794", "#9c3892", "#49857d", "#5b8781", "#302d5c", "#28
1ebd", "#b81d2d", "#d4aeb1", "#c91c2a", "#f57d87", "#f5f783", "#737509", "#db620b")
```

```
bichos<-
c("Helicobacter", "Escherichia", "Streptococcus", "Lactobacillus", "Fusobacterium", "Prevotella", "L
actococcus", "Pseudomonas", "Clostridium", "Veillonella", "Methylobacterium", "Alloprevotella",
"Leptotrichia", "Chryseobacterium", "Staphylococcus", "Citrobacter", "Haemophilus", "Selenomo
nas", "Campylobacter", "Bacteroides", "Treponema", "Sphingomonas", "Minor_genera", "Género
no clasificado", "Filo no
clasificado", "Minor_phyla", "Bacteroidetes", "Proteobacteria", "Firmicutes", "Epsilonbacteraeota
", "Actinobacteria", "Dubosiella", "Pedobacter", "Akkermansia", "Faecalibacterium", "Sphingobiu
m", "Rothia", "Corynebacterium", "Herbaspirillum", "Acinetobacter", "Atopostipes", "Sediminibact
erium", "Megasphaera", "Virgibacillus", "Novosphingobium", "Rhodanobacter", "Porphyromonas
", "Alkalibacterium", "Granulicatella", "Enterococcus", "Neisseria", "Gemella", "Actinomyces", "Mic
rococcus", "Tannerella", "Brachyobacterium", "Sarcina", "Actinobacillus", "Brevundimonas", "Serrat
ia", "Thermus", "Cutibacterium", "Gordonia", "Tsukamurella", "Enhydrobacter", "Burkholderia", "B
acillus", "Pannonibacter", "Haematomicrobium", "Ochrobactrum")
```

```
#Relacionamos los colores con los microorganismos
colores <-mycolors[1:70]
lista_colores_bichos<-cbind(bichos, colores)
lista_colores_bichos<-as.data.frame(lista_colores_bichos)
```

Anexo III

```
#Lo ideal es que los Unclassified aparezcan al final. Los buscamos y marcamos posiciónposicio
en levels(df.m$variable)[-2]
levels(df.m$variable)
recolocar<-levels(df.m$variable)[-2]
recolocar2<-c(recolocar,"Género no clasificado")
df.m$variable<-factor(df.m$variable, levels = recolocar2)
```

```
#Definimos titulos de columnas y orden de aparición en el boxplot
df.m<-`colnames<-`(df.m, c("factor", "Generos", "Abundancia_relativa"))
df.m$Generos<-forcats::fct_rev(df.m$Generos)
levels(df.m$Generos)
```

```
#Ordenamos la lista de géneros y filas con la lista de objetivos
lista_colores_bichos<-lista_colores_bichos[order(match(lista_colores_bichos$bichos,
levels(df.m$Generos))),]
```

```
#Pintamos el diagrama de barras apilado
ggplot(df.m, aes(fill=Generos, y=Abundancia_relativa, x=factor)) + geom_bar(stat =
"identity",position = position_fill(reverse=F)) +scale_fill_manual(values=
as.vector(lista_colores_bichos$colores[c(1:21)]))
```

#Gráfico de correlaciones de los géneros

```
#Necesita la librería "corrplot" instalada
library("corrplot")
```

```
#Preparamos la función para los p-valores
```

```
cor.mtest <- function(mat, ...) {
  mat <- as.matrix(mat)
  n <- ncol(mat)
  p.mat<- matrix(NA, n, n)
  diag(p.mat) <- 0
  for (i in 1:(n - 1)) {
    for (j in (i + 1):n) {
      tmp <- cor.test(mat[, i], mat[, j], ...)
      p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
    }
  }
  colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
  p.mat
}
```

```
#Gráfico correlaciones
para_cor<- cor(a[,1:20],y=NULL,use="pairwise.complete.obs" ,"spearman")
corrplot.mixed(para_cor, upper = "ellipse",tl.pos ="lt",tl.cex=1,number.cex=0.8,
order="hclust",addrect = 3)
#Gráfico correlaciones y p-valores
```

```
p.mat <- cor.mtest(a[,1:20])  
cors<-cor(a[,1:20])  
corrplot.mixed(cors, tl.pos = "lt", order="hclust", diag = "f", upper = "ellipse", number.cex =  
0.6, p.mat = p.mat, sig.level = 0.05, insig = "blank")
```