UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

Duodenal microbiome characterization in patients with and without *Helicobacter pylori* infection

Andrés Eduardo Suárez Jaramillo

Paúl Cárdenas MD, PhD
Director del Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito
para la obtención del título de Magister en Microbiología.

Quito, 26 de noviembre de 2019

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

Duodenal microbiome characterization in patients with and without *Helicobacter pylori* infection

Andrés Eduardo Suárez Jaramillo

Firmas

Paúl Cárdenas MD, PhD.
Director del Trabajo de Titulación

Gabriel Trueba, PhD.
Director del Programa de Posgrados del Instituto de Microbiología.
Miembro del comité de tesis.

Antonio Machado, PhD.
Miembro del comité de tesis.

Hugo Burgos, PhD.
Decano del Colegio de Posgrados.

Quito, 26 de noviembre de 2019.

© DERECHOS DE AUTOR

Firma del estudiante:	
Nombre:	Andrés Eduardo Suárez Jaramillo.
Código de estudiante:	00140972.
C. I.:	1719249235.
Lugar, Fecha	Quito, 26 de noviembre de 2019.

DEDICATORIA

A todas las personas que durante mi paso por USFQ me inspiraron, me dieron ideas y me recordaron porque amo la ciencia y el conocimiento. Y lo más importante, me hicieron dar cuenta porque es trascendente la búsqueda de la felicidad. Este trabajo está dedicado a: Carlos Freile, Jaime Costales, Valeria Espinosa, Pablo de la Cerda, Daniela Vásquez, Farah Cuadros, Ana Ulloa, Ana Beatriz Dueñas, Mónica Arévalo, Andrea Velasco, María Olivia Franco, Elena Çatelan, Doménica Costa, Doménica Vargas, “Stefanie Scott”, Marthina Aguirre y Daniela Moreira.

AGRADECIMIENTOS

A mi familia por su constante apoyo.

A la Universidad San Francisco de Quito por haberme dado la oportunidad de formarme como médico y como microbiólogo.

A mi director de tesis, Paúl Cárdenas, por sus enseñanzas.

A la University of North Carolina at Chapel Hill, específicamente, al equipo del High-Throughput Sequencing Facility por su ayuda en la elaboración de las bibliotecas genómicas y el secuenciamiento.

A Daniela Garcés por el tiempo que invirtió en enseñarme a hacer extracciones de ADN.

A Belén Prado por su apoyo y enseñanzas en el análisis bioinformático.

A Sully Márquez por su ayuda en múltiples ocasiones, aunque no era su obligación hacerlo.

A Viviana Albán quien me ayudó a hacer MIC para otro proyecto mientras trataba de terminar mi tesis.

RESUMEN

La infección por *H. pylori* afecta aproximadamente al 44.3% de la población mundial y se asocia con patologías gastrointestinales y extra gastrointestinales. Sin embargo, no todos los pacientes infectados desarrollan enfermedad. Las razones para este fenómeno no están completamente entendidas, pero es posible que dependa de muchos factores como las interacciones con el sistema inmune, factores de virulencia y cambios en el microbioma. No se conocen todos los detalles de cómo *H. pylori* modifica el microbioma, especialmente el duodenal. Por lo tanto, este estudio tuvo como objetivo estudiar el microbioma de biopsias duodenales en pacientes ecuatorianos con y sin infección por *H. pylori.* Para esto se hizo secuenciamiento del gen 16s rRNA con Illumina Miseq. Las secuencias obtenidas fueron analizadas con Qiime2 para obtener índices de diversidad, asignación taxonómica y abundancia diferencial de taxones en los grupos con infección y sin infección. Con MEGA X se realizaron árboles filogenéticos de los OTUs (del inglés Operational Taxonomic Unit) asignados como *Helicobacter*. También se realizó BLAST de los OTUs de *Helicobacter* para determinar si son *H. pylori.* Además, se utilizó PICRUSt2 y DESeq2 para realizar una predicción del perfil funcional. Como resultados se obtuvo que la infección por *H.* pylori modifica la alfa diversidad (Índice de Shannon y diversidad filogenética de Faith) y beta diversidad (Jaccard, Bray-Curtis, unweighted UniFrac y weighted UniFrac). Se encontró que el género *Ralstonia* fue más abundante en las biopsias de los pacientes sin infección. Por otro lado, los géneros/especies *Haemophilus, Neisseria, Prevotella pallens, Prevotella 7, Streptococcus* y una bacteria (phylum no definido) fueron más abundantes en los infectados. Los géneros diferencialmente abundantes pueden ser asociados con diferentes patologías humanas. Se observó un OTU de *Helicobacter* diferente en cada muestra, excepto en un caso en el que se describieron varios (muestra HDRP0079). En el BLAST, todos los OTUs de *Helicobacter* fueron identificados como *H. pylori.* Sin embargo al analizar los porcentajes de identidad, dos OTUs (OTU18 y OTU19) tuvieron menos de 97% lo que sugiere que pueden ser otra especie. Adicionalmente, en los árboles filogenéticos no todos los OTUs de *Helicobacter* se agruparon en un mismo clado y los OTUs con porcentaje de identidad menos de 97% se separan del resto de OTUs. Con PICRUSt2 y DESeq2 se identificó un total de 65 genes ortólogos de la Enciclopedia de Kyoto de genes y genomas, y 155 genes de enzimas que se expresaron diferencialmente entre los grupos. Entre los genes encontrados los más importantes fueron: *comB4, comB8, comB9*, enzima de restricción tipo II, ADN adenina metilasa, *yafQ*, *hicB*, hidroxietiltiazol, tiamina difosfoquinasa, GMP-reductasa, fosfolipasas (A1 y A2), asparaginasa y UDP-4-amino-4,6-dideoxi-L-N-acetil-beta-L-altrosamine transaminasa. Todos estos genes están asociados con virulencia o adaptación bacteriana. Finalmente, se realizó un análisis de ITS (del inglés Internal Transcribed Spacer) para caracterizar el micobioma. Las taxa que se identificaron en más muestras fueron Fungi (phylum no definido), Recurvomyces, Ascomycota and *Xenoacremonium falcatum*.

Palabras clave: *H. pylori*, microbioma duodenal, alfa diversidad, beta diversidad, filogenia, Gneiss, PICRUSt2, DESeq2, micobioma duodenal.

ABSTRACT

H. pylori infection affects approximately 44.3% of the world's population and is associated with gastrointestinal and extra gastrointestinal pathologies. However, not all infected patients develop disease. The reasons for this phenomenon are not completely understood, but it is possible that it depends on many factors such as interactions with the immune system, virulence factors and changes in the microbiome. Not all the details of how *H. pylori* modifies the microbiome, especially the duodenal, are known. Therefore, this study's aim is to study the microbiome of duodenal biopsies in Ecuadorian patients with and without *H. pylori infection*. For this, sequencing of the 16s rRNA gene with Illumina Miseq was done, and sequences obtained were analyzed with Qiime2 to obtain indices of diversity, taxonomic assignment and differential abundance of taxa in groups with infection and without infection. With MEGA X, phylogenetic trees of the OTUs (Operational Taxonomic Unit) assigned as *Helicobacter* were made. BLAST of Helicobacter OTUs was also performed to determine if the OTUs are *H. pylori*. In addition, PICRUSt2 and DESeq2 were used to make a prediction of the functional profile. We found that *H. pylori* infection modifies alpha diversity (Shannon Index and Faith phylogenetic diversity) and beta diversity (Jaccard, Bray-Curtis, unweighted UniFrac and weighted UniFrac). The genus *Ralstonia* was more abundant in the biopsies of patients without infection. On the other hand, genera/species *Haemophilus, Neisseria, Prevotella pallens, Prevotella 7, Streptococcus* and a bacterium (undefined phylum) were more abundant in the infected. Differentially genus abundance can be associated with different human pathologies. A different *Helicobacter* OTU was observed in each sample, except in one case in which several were described (sample HDRP0079). In BLAST, all Helicobacter OTUs were identified as H. pylori. However, when analyzing the percentages of identity, two OTUs (OTU18 and OTU19) had less than 97% suggesting that they may be another species. Additionally, in the phylogenetic trees not all *Helicobacter* OTUs were grouped in the same clade and OTUs with an identity percentage less than 97% were separated from the rest of OTUs. With PICRUSt2 and DESeq2 were identified a total of 65 orthologous genes of the Kyoto Encyclopedia of Genes and Genomes, and 155 enzyme genes that expressed differentially between the groups. Among the genes found the most important were: *comB4, comB8, comB9,* type II restriction enzyme, DNA adenine methylase, *yafQ, hicB*, hydroxyethylthiazole, thiamine diphosphokinase, GMP-reductase, phospholipases (A1 and A2), aspisaraginase and UDP-4-amino-4,6-dideoxy-LN-acetyl-beta-L-altrosamine transaminase. All these genes are associated with virulence or bacterial adaptation. Finally, an ITS (Internal Transcribed Spacer) analysis was performed to characterize the mycobiome. Taxa that were identified in more samples were Fungi (undefined phylum), Recurvomyces, Ascomycota and *Xenoacremonium falcatum*.

Key words: *H. pylori*, duodenal microbiome, alpha diversity, beta diversity, phylogeny, Gneiss, PICRUSt2, DESeq2, duodenal mycobiome.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Helicobacter pylori is a small, microaerophilic, non-spore-forming, gram negative curved-rod. It is encountered in the stomach mucosa of humans (and other animals) which can persist lifelong (Jorgensen et al., 2015; Portal-Celhay & Perez-Perez, 2006). Approximately 44.3% of world's population is infected (Ansari & Yamaoka, 2018; Zamani et al., 2018). However, the prevalence changes according to geographic region, income level, age and gender (Zamani et al., 2018). In a meta-analysis with data from 73 countries in six continents, the mean prevalence in developing countries was 50.8% (CI 95%: 46.8-54.7). On the other hand, the prevalence in developed countries was 34.7% (CI 95%: 39-46.5) (Zamani et al., 2018). Nevertheless, there is significant variation between countries. For example, Nigeria has the highest prevalence (89.7%); while, Yemen has the lowest (8.9%) (Zamani et al., 2018). Infection frequency also depends on gender. Men were more affected (46.3%, CI 95%: 42.1-50.5) than women (42.7%, CI 95%: 39-46.5) (Zamani et al., 2018). Age is also a determinant of infection. Adult patients (18 years old or more) had a prevalence of 48.6% (CI 95%: 43.8-53.3) in comparison with children that had 32.6% (CI 95%: 28.4-36.8) (Zamani et al., 2018). There are few epidemiologic studies in Ecuador, in which the prevalence of infection in adult patients varies from 40.2% (CI 95%: 37.4-43) to 72.2% (Buitrón, 2013; Sasaki et al., 2009; Zamani et al., 2018). In children, Egorov and colleagues reported a percentage of 61% (Egorov et al., 2010).

H. pylori infection is associated with dyspepsia, gastrointestinal cancer (gastric adenocarcinoma, MALT lymphoma, colon cancer), peptic ulcers (duodenal and gastric), hematological disorders, metabolic alterations, growth retardation, ischemic heart disease, dermatological diseases, and even neurological diseases (Ansari & Yamaoka, 2018;

Benavides-Ward et al., 2018; Gravina et al., 2018; Teimoorian, Ranaei, Hajian Tilaki, Shokri Shirvani, & Vosough, 2018). Not all infected patients develop symptoms or complications. The reasons are not completely understood; nevertheless, it is possible that depends on many factors. For example, interactions with the immune system, virulence factors and microbiome changes (Kao, Sheu, & Wu, 2016; Sgouras, Thi, Trang, & Yamaoka, 2015).

There are studies that try to determine the impact of *H. pylori* in gastrointestinal microbiome and how the interactions can cause diseases. Many observations suggest that the infection produces changes in different bacterial phyla (Actinobacteria, Bacteroidetes and Proteobacteria, Firmicutes, Spirochaetes, and Acidobacteria) and genera (*Veillonella*, *Neisseria, Haemophilus*, *Prevotella* and *Streptococcus*). This information has been obtained mainly from analyses of gastric or fecal microbiome on several studies (Bühling, Radun, Müller, & Malfertheiner, 2001; Gao et al., 2018; Klymiuk et al., 2017; Maldonado-Contreras et al., 2011; Schulz et al., 2018). However, these studies had limitations; for example, different methodologies have been applied and many studies have been made with small number of samples. Additionally, there are not studies that analyze the impact of *H. pylori* infection in gastrointestinal mycobiome. Consequently, the details of how *H. pylori* modifies the gut microbiome are still unknow (Luyi Chen et al., 2018). Moreover, information about duodenal microbiome in patients with and without *H. pylori* is limited (Schulz et al., 2018; Shiotani et al., 2018).

It has been hypothesized that alterations in duodenal microbiome can produce dysfunction of this organ which is related with dyspepsia and maybe other disease manifestations (H. Han et al., 2019). Schulz and colleagues made the only study that directly analyzed the relation between *H. pylori* infection and duodenal microbiome. The study was

made in Germany applying V1-V2 16s rRNA sequencing. The subjects were 24 adults (mean age 52 years old) with and without *H. pylori* infection that were underwent to upper endoscopy. Aspirates and biopsies form duodenum were taken (Schulz et al., 2018). Proteobacteria were more abundant in the aspirate of *H. pylori* positive subjects. The phylotype Phy103 (*Staphylococcus aureus*) had more abundance in the aspirate of *H. pylori* negatives (Schulz et al., 2018). In duodenal biopsies, patients with *H. pylori* infection had less abundance of Firmicutes and more abundance of Bacteroidetes compared with the no infected group. Nevertheless, these observed differences were not statistically significative (Schulz et al., 2018). Also, in biopsies from patients with *H. pylori* infection the abundance of phylotypes Phy78 (*Rhodobacteriaceae*) and Phy269 (*Campylobacter gracilis*) was significantly lower. Moreover, only in duodenal biopsies of not infected patients was possible to observe the following phylotypes: Phy140 (*Streptococcus infantis*), Phy168 (*Actinomyces*), Phy81 (*Enterococcus*) and Phy478 (*Lachnospiraceae*)(Schulz et al., 2018).In Ecuador, there are not studies of *H. pylori* and its relationship with gastrointestinal microbiome (including duodenum).

Hence, the main goal of this study was to increase our knowledge of *H. pylori*-microbiome interactions in the duodenal mucosa; in particular, in the Ecuadorian population. The findings obtained in this work attempt to be useful as starting point for future investigations and to propose possible therapeutic interventions that targets microbiome.

Objectives:

General objective: Characterize duodenal microbiome (bacterial and fungal) in patients with and without *Helicobacter pylori* infection.

Specific objectives:

- Describe the general characteristics of positive and negative biopsies for *H. pylori* in the current study participants.
- Compare alpha and beta diversity indexes between positive and negative biopsies for *H. pylori*.
- Determine if taxa are differentially abundant between patients with and without *H. pylori* infection.
- Analyze the phylogenetic relationships based on 16S rRNA gene of the *Helicobacter* found in infected patients.
- Establish a functional prediction with PICRUSt2 to determine bacterial genes or metabolic pathways differentially expressed between patients with and without *H. pylori* infection.

METHODOLOGY AND STUDY DESIGN.

Study design and population:

This is a cross-sectional study that was conducted at Axxis Hospital (Quito, Ecuador), Biodimed (Quito, Ecuador) and Instituto ecuatoriano de seguridad social Hospital (Sangolquí, Ecuador) from 2016 to 2017. The inclusion criteria were ambulatory patient, age of 18 years old or more, not antibiotics use in the last two weeks, medical indication for upper endoscopy and agree to participate in the study (demonstrated by the signing of informed consent). The exclusion criteria were inpatient, age less than 18 years old, antibiotics use in the last two weeks, pre- or probiotic use, chronic use of corticosteroids, use of medications that can alter nutrient absorption, contraindications for upper endoscopy or biopsy, inflammatory bowel disease, malabsorption syndromes and not accept to participate in the study. Duodenal and gastric biopsies of every participant were obtained by upper endoscopy which was made by a gastroenterologist. Gastric samples were analyzed by a pathologist to determine the presence of *H. pylori* infection. Duodenal tissue was conserved in phosphate-buffered saline (PBS) at -80°C until DNA extraction. In order to respect all ethical principles, the study protocol was reviewed and approved by a human research ethics committee of the Universidad de las Americas (UDLA).

DNA extraction:

FastDNA SPIN Kit for Soil (MP Biomedicals) and DNeasy blood and tissue kit (Qiagen) were tested to determine the best for the sample's extraction. Consequently, the DNA of two duodenal tissues were extracted with every kit. In both cases the procedure was done

following the manufacturer's instructions. Later, using NanoVue spectrophotometer DNA concentrations and purity (A260/280 and A260/230) were measured. Reference values of the purity indexes are: A260/280: 1.8; A260/230: 2.0-2.2. A260/280 detects contamination by proteins and phenols. A260/230 detects contamination by carbohydrates and phenols (Matlock, 2015). Next, 16S rRNA PCR was made in order to verify the correct amplification of bacterial DNA. The PCR protocol used at the Institute of Microbiology of the Universidad San Francisco de Quito was applied (Lane, 1991; Turner, Pryer, Miao, & Palmer, 1999). For one reaction, the master mix was composed of 12.88 µL of PCR water, 2.2 µL of $MgCl_2$ (final concentration 2.5mM), 2.5 µL of buffer Go Taq (Promega, Madison, WI, US) (final concentration 1X), 1.5 µL of dNTPs (final concentration 0,25mM), 0.4 µL of primer 27F: AGAGTTTGATCCTGGCTCAG (final concentration 0.2 uM), 0.4 µL of primer 1429R: GGTTACCTTGTTACGACTT (final concentration 0.2 uM), 0.12 µL of Taq polymerase (Promega, Madison, WI, US) (final concentration 2.5 U) and 5 µL DNA. A termocycler Bio-Rad T100™ was used with the following program: 1) 94°C for 4 minutes; 2) 30 cycles of 94°C for 1 minute, 57°C for 0.3 minutes and 72°C for 2 minutes; 3) 72°C for 8 minutes. Results were visualized in an agarose gel 1.5% after an electrophoresis time of 30 minutes (amplicon size 1500 bp). Finally, the results were analyzed (see results section). FastDNA SPIN Kit for Soil (MP Biomedicals) was chosen because the DNA extracted with this procedure had the best concentration and amplification in PCR.

With the FastDNA SPIN Kit for Soil (MP Biomedicals) the DNA of 93 more samples were extracted. Also, the DNA concentrations and purities were made as mentioned above to determine quality. The samples were stored at -20°C. An aliquot of 30 µL of every DNA

was lyophilized in order to be send to UNC-High-Throughput Sequencing Facility (University of North Carolina at Chapel Hill).

Analysis of the bacterial microbiome:

Genomic library preparation and sequencing:

These procedures were made at UNC-High-Throughput Sequencing Facility. Libraries of the V4 region of 16s rRNA were prepared. Paired-end sequencing was made with illumina MiSeq system. Read length was 2 x 300 and PhiX 20% was used as sequencing control.

Bioinformatic analysis:

Raw sequences were analyzed with the bioinformatic package Qiime2 (Quantitative Insights Into Microbial Ecology) version 2018.11 (https://qiime2.org/) (Boylen et al., 2018). Below are the steps made during the process.

Metadata construction and validation:

Of every patient was taken age, gender and *H. pylori* status according to gastric biopsy (positive and negative). Additionally, a unique code was assigned to each sample of duodenal tissue. With this information the metadata was built in Google Sheets. Keemei was used to validate the information (Rideout et al., 2016). Finally, the metadata document was exported in .tsv format for subsequent analyzes.

Importing data and denoising:

Demultiplexed sequences were imported to Qiime2 using qiime tools import method. Once the information was imported, a summary with the sequences obtained per sample and the distribution of sequence qualities at each position was obtained using qiime demux summarize commands (Boylen et al., 2018). From this information were determined the trim-left (position 6) and trunc-len (position 290) parameters that are necessary for

denoising. Afterwards, denoising was made using DADA2. This tool allows to detect and correct Illumina sequencing errors, and filter chimeric sequences (Callahan et al., 2016). Forward and reverse sequences were analyzed with DADA2. The results of denoising step were visualized with feature-table summarize and feature-table tabulate-seqs commands. With the first one was obtained a summary of the number of sequences associated with every sample and OTU (operational taxonomic unit). With the second one was generated a table of representative sequences which shows OTUs and their sequences associated with links to make BLAST.

Alpha and beta diversity analysis:

Groups analyzed were *H. pylori* positive vs negative and men vs women. For all diversity analysis was applied the q2-diversity plugin. A phylogenetic tree was created using the representative sequences of the previous step and the align-to-tree-mafft-fasttree pipeline (Katoh, Misawa, Kuma, & Miyata, 2002; Price, Dehal, & Arkin, 2010). It was performed a multiple sequence alignment, where highly variable sequences were removed, and an unrooted tree was generated. Later a root was placed at a midpoint between the longest tip-to-tip distance (Katoh et al., 2002; Price et al., 2010). The tree was used in the following steps to support some diversity metrics such as Faith, unweighted UniFrac and weighted Unifrac.

Next, core-metrics-phylogenetic method was applied to obtain alpha and beta diversity metrics including Principal Coordinates Analysis (PCoA). In this step the sampling depth was established in 2755 (the lowest number of sequences in a sample) for a single rarefaction of the samples. The alpha diversity metrics calculated were Faith's Phylogenetic Diversity, Shannon's diversity and evenness (Faith, 1992; Morgan & Huttenhower, 2012). On

every metric there were obtained boxplots and statistical significances with Kruskal-Wallis test (qiime diversity alpha-group-significance commands). Alpha rarefaction curves of Shannon and Faith indexes were visualized with qiime diversity alpha-rarefaction. Maximum depth was established in 6000. In the case of beta diversity were determined Jaccard distance, Bray-Curtis distance, weighted UniFrac and unweighted UniFrac (Goodrich et al., 2014; Kuczynski et al., 2012; Lozupone, Hamady, Kelley, & Knight, 2007; Morgan & Huttenhower, 2012). Boxplots and statistical significances with PERMANOVA were made for all beta diversity metrics (qiime diversity beta-group-significance commands). PCoAs, of the four metrics, were visualized and analyzed using the Emperor tool (Vázquez-Baeza, Pirrung, Gonzalez, & Knight, 2013).

Taxonomic analysis:

Taxonomy was assigned to the OTUs using the database Silva version 132 (https://www.arb-silva.de/documentation/release-132/) (used in April 2018) (Quast et al., 2013; Yilmaz et al., 2014) and the q2-feature-classifier plugin (Bokulich et al., 2018). Inside the plugin, with the command feature-classifier classify-sklearn a Naive Bayes Classifier (NBC) was ran. This classifier compare the pattern of sub-sequences (k-mers) of every study sequence with the k-mers of the reference sequences of the database. Based on this comparison, it establishes a probability based on Bayes' theorem of which reference sequence corresponds most likely with the every study sequence. NBC also makes a bootstrap to calculate a percentage of confidence. In Qiime2 the default cut-off point of the confidence percentage is 70%, this means that if a taxonomic assignment is made with a confidence percentage less than this cut point, the system gives the result “unassigned” (Bokulich et al., 2018; Wang, Garrity, Tiedje, & Cole, 2007). As a result of the taxonomic

assignment, a table was obtained which contains the OTUs with their respective assigned taxa (silva-taxonomy table). Then, taxa barplots were generated with qiime taxa barplot commands (Boylen et al., 2018). Groups compared in plots were *H. pylori* positive vs negative and men vs women. In all cases, men and female groups contain samples positive and negative for *H. pylori* infection.

It was not possible to assign taxonomy in an appropriate way to all the OTUs. Some were classified only as bacteria (undefined phylum) and others as unassigned (see results and discussion). To try to determine what these OTUs correspond to, BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschup, Gish, Pennsylvania, & Park, 1990) was performed on each of the sequences that were not correctly identified. Silva-taxonomy table and representative sequences table were used to identify the sequences that needed to be reanalyzed. The best match of every sequence was reported.

Helicobacter phylogeny and OTUs distribution:

OTUs and their associated sequences that were assigned as *Helicobacter,* were identified using Silva-taxonomy and representative sequences tables. BLAST was made to confirm that the sequences belong to *H. pylori.* The distribution of *Helicobacter* OTUs in every sample together with its relative abundance was represented in a heatmap made with qiime feature-table heatmap commands (Hunter, 2007). With MEGA X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018), two phylogenetic trees were generated. One with all *Helicobacter* OTUs and other with *Helicobacter* OTUs present in sample HDRP0079. A separated three was made with sample HDRP0079 because it was the only one that had several OTUs. Sequence alignment was made with Clustal W 2.1 (using MEGA X default conditions) and the phylogenetic relationships were inferred with maximum likelihood and Hasegawa-Kishino-

Yano model (Hasegawa, Kishino, & Yano, 1985; Thompson, Higgins, & Gibson, 1994). Also, it was applied a bootstrap of 100 replicates (Felsenstein, 1985). In both phylogenetic trees, it was incorporated four *H.pylori* reference sequences of the GeneBank: *Helicobacter pylori* NY40 (Accession AP014523.1), *Helicobacter pylori* strain G-Mx-2003-250 (Accession CP032048.1), *Helicobacter pylori* strain ATCC 43504 (Accession MN326691.1) and *Helicobacter pylori* strain B147 (Accession CP024946.1).

Differential abundance analysis:

In order to identify taxa that are more abundant between groups, two methods were applied: Analysis of composition of microbiomes (ANCOM) (Mandal et al., 2015) and gneiss (Morton et al., 2017). To use ANCOM the q2-composition plugin was applied (Mandal et al., 2015). In order to compare the groups at different taxonomic levels, first qiime taxa collapse commands were utilized. ANCOM does not admit a frequency of 0 in any OTU; hence, the command add-pseudocount was used next. With this the minimum frequency was established on one (Mandal et al., 2015). Subsequently, qiime composition ancom commands were utilized to visualize volcano plots, W value (number of times the null hypothesis is rejected for a particular taxa, see supplementary material of Mandal and colleagues) and percentiles (Mandal et al., 2015).

The plugin q2-gneiss was also applied to do differential abundance analysis (Morton et al., 2017). The first step was to create a tree based on correlation-clustering (qiime gneiss correlation-clustering commands). In other words, groups of microorganisms that are correlated with each other were clustered together. Then in the internal nodes of the tree, balances were calculated applying log ratios between groups (qiime gneiss ilr-hierarchical commands) (Morton et al., 2017). Lastly, proportion plots were generated (qiime gneiss

balance-taxonomy) to show differential abundance between groups. The taxonomic levels used were family, genus and specie.

Linear models with Gneiss:

Using the balances generated previously, linear models were generated to determine which variables contribute more to the variation observed in the community. The commands qiime gneiss ols-regression were applied and the examined variables were age, gender and *H. pylori* status (Boylen et al., 2018; Morton et al., 2017). This analysis was made with 70 samples. Samples of the patients with no age information were excluded. Also was excluded sample HDRP0081.

Prediction of the functional profile:

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) was used to infer the functional profile based on 16s rRNA sequences (Douglas et al., 2019). For this, the plugin q2-picrust2 was applied in Qiime2 version 2019.4. As a result, were obtained abundances of orthologs of the Kyoto Encyclopedia of Genes and Genomes (https://www.genome.jp/kegg/ko.html) (KEGG) and of the enzymes classified according to the enzyme commission number (EC). With enzyme information, MetaCyc pathway abundances were also predicted. All abundance information was exported in .txt format and analyzed in the R based web tool MicrobiomeAnalyst (https://www.microbiomeanalyst.ca/) (Dhariwal et al., 2017). Data filtering with default parameters and data transformation with centered log ratio method was made. Genes and pathways differential expression was analyzed with DESeq2 method (Love, Huber, & Anders, 2014).

Analysis of the mycobiome:

Library preparation of ITS and paired-end sequencing with Illumina MiSeq was done at UNC-High-Throughput Sequencing Facility. Denoising was made with DADA2 using trim-left of 6 and trunc-len of 180. When forward and reverse sequences were denoised only 608 of the initial 17364987 were obtained. For this reason denoising was made only with forward sequences obtaining 6681809 sequences and 1192 OTUs. Taxonomic assignment was done with the database UNITE 8.0 for Fungi (https://unite.ut.ee/) (used in August 2019) (UNITE Community, 2019), but the majority (1072/1192, 89.93%) of OTUs were defined as Fungi (undefined phylum). Consequently, diversity metrics and differential abundance analyzes were not performed. In results section are summarized the taxonomic findings.

Study population statistics:

For the analysis of the variables age, gender, *H. pylori* status and fungal taxa, descriptive statistics were made using mean, maximum, minimum, percentages and standard deviation. The statistical package used was SPSS version 25 (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.).

RESULTS

DNA quality:

The results of the analysis of the two methods of DNA extraction (MPbio and Qiagen) are shown in table 1. On average, a higher concentration was obtained with MPbio (88 ng/µL mean concentration of HDRP0054 and HDRP0052 samples), while with Qiagen it was obtained 59.75 ng/µL (mean concentration of HDRP0034 and HDRP0004).

Table 1. Summary of concentrations and purities obtained in DNA extraction with MPbio and Qiagen kits.

Sample ID	Kit	Concentration (ng/µL)	A260/280	A260/230
HDRP0054	MPbio	40.5	1.588	0.163
HDRP0052	MPbio	135.5	1.683	0.313
HDRP0034	Qiagen	70	1.728	2.258
HDRP0004	Qiagen	49.5	1.768	4.583

In 16S rRNA PCR, the bands of the two samples extracted with the MPbio kit were clear. Conversely, only the bands of one of the samples extracted with Qiagen kit were clear. Considering the results of mean concentration and PCR amplification, MPBio kit was selected to do the DNA extraction of all study samples.

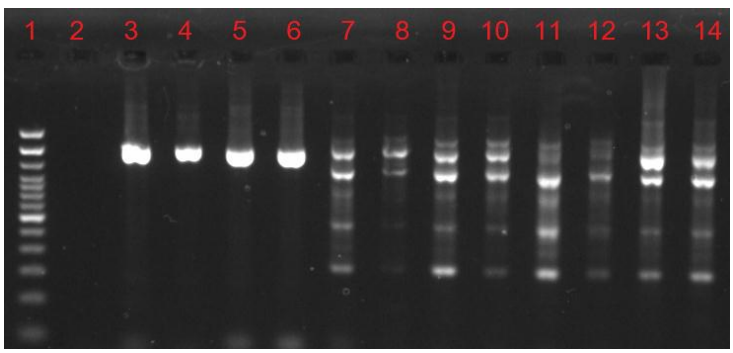


Figure 1. PCR of 16s rRNA made with the DNA extracted using MPbio and Qiagen kits.

1) Ladder, 2) Negative control, 3) Positive control *Listeria* without dilution, 4) Positive control *Listeria* with dilution 1/10, 5) Positive control *Salmonella* without dilution, 6) Positive control *Salmonella* with dilution 1/10, 7) Sample HDRP0054 (MPbio, dilution 1/10), 8) Sample HDRP0054 (MPbio, dilution 1/100), 9) Sample HDRP0052 (MPbio, dilution 1/10), 10) Sample HDRP0052 (MPbio, dilution 1/100), 11) Sample HDRP0034 (Qiagen, dilution 1/10), 12) Sample HDRP0034 (Qiagen, dilution 1/100), 13) Sample HDRP0004 (Qiagen, dilution 1/10), 14) Sample HDRP0004 (Qiagen, dilution 1/100).

In all study samples that DNA was extracted, the mean concentration was 148.77 ng/μL (range: 21.5-467 ng/μL), A260/280 was 1.77 (range: 1.644-1.957) and A260/230 was 0.170 (range: 0.529-0.028).

Population general data:

Duodenal biopsies of 97 patients were obtained; however, 22 samples were discarded. One for having a low DNA concentration (10 ng/μL) and 21 for lacking pathology information; 51 (68%) were women and 24 (32%) were men. The percentage of gastric biopsies positive for *H. pylori* were 49.33% (n=37) and the negatives were 50.67% (n=38). Table 2 shows the frequency and percentage of positive and negative biopsies according to gender. Most of the negative biopsies were in the group of women (31 of 38). On the other hand, the number of positive biopsies is similar between the two genders (20 and 17 in women and men, respectively).

Table 2. Results of gastric biopsies according to the gender.

	Women	Men	**Total**
Negative biopsy	31 (41.33%)	7 (9.33%)	38 (50.67%)
Positive biopsy	20 (26.67%)	17 (22.67%)	37 (49.33%)
Total	51 (68%)	24 (32%)	75 (100%)

The mean age was 36.94 years (minimum 21, maximum 69, standard deviation 10.12). It should be considered that the age of four participants could not be obtained. Therefore, the statistics referring to this variable were calculated from the information of 71 patients.

Analysis of the bacterial microbiome:

Sequences quality control:

After denoising with DADA2 and the exclusion of the samples that were eliminated from the study, it was obtained 1,219,867 sequences and 3,350 OTUs.

Diversity:

Alpha diversity:

H. pylori infection contributed significantly to richness measured with Faith's phylogenetic diversity (FD) and Shannon index (SI). Boxplots showed a small difference when the presence and absence of infection is compared (figure 2 and 3). In these figures, FD and SI were higher in the positive group. With Kruskal-Wallis statistics the difference was confirmed. The results for FD were: H= 4.054054 (*P*= 0.044065). In the case of SI, the values were: H= 5.336977 (*P* = 0.020878).

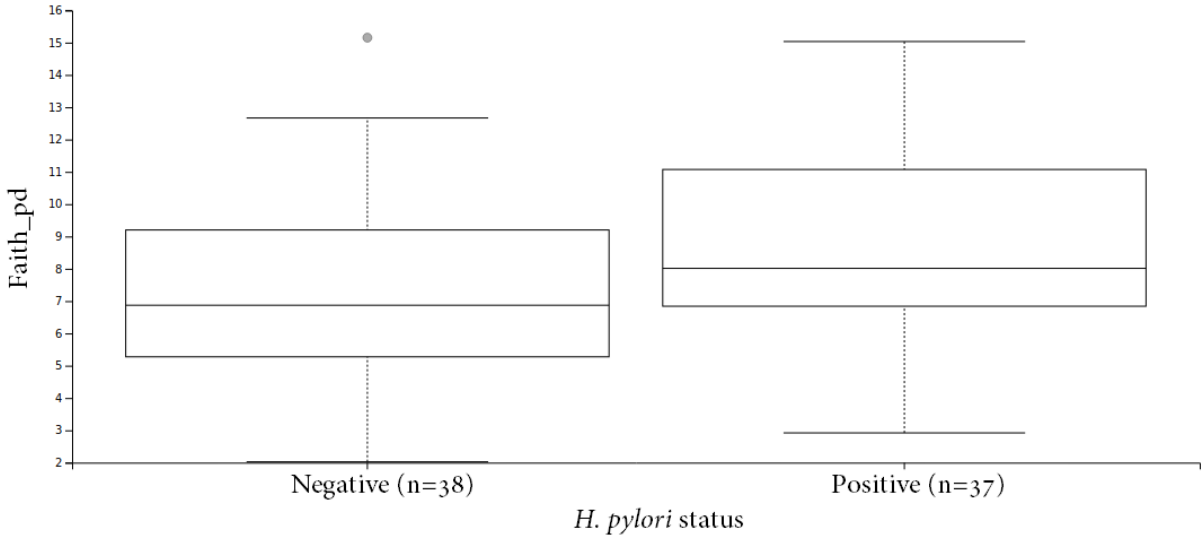


Figure 2. Boxplots of Faith's phylogenetic diversity that compare *H. pylori* negative and positive groups. The left boxplot represents *H. pylori* negative group and the right boxplot represents *H. pylori* positive group. In the y-axis is the numerical value of the Faith's phylogenetic diversity index. Kruskal-Wallis statistics: H= 4.054054 (P= 0.044065).

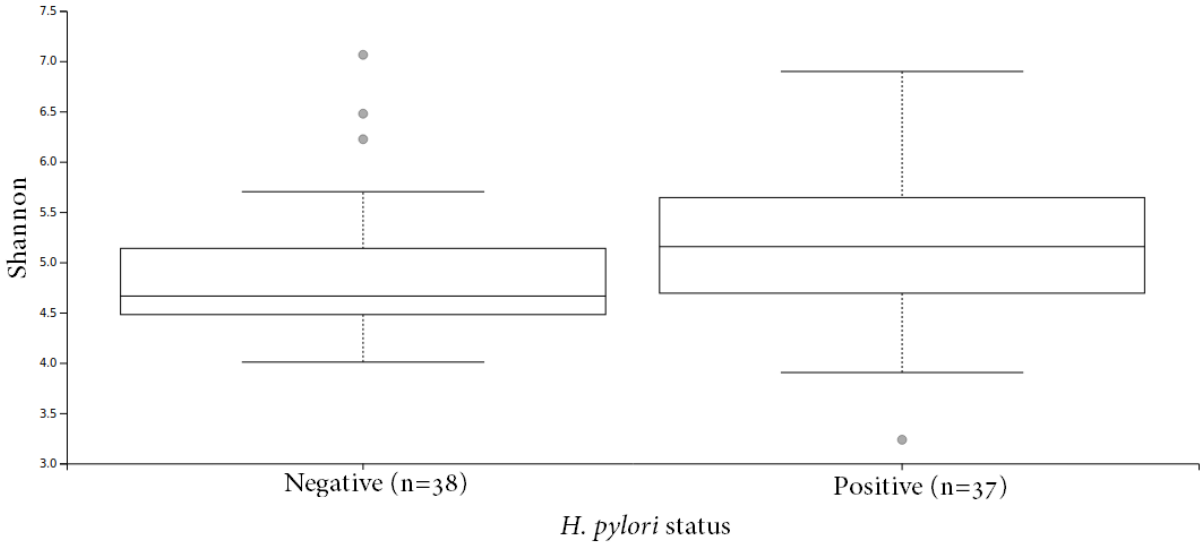


Figure 3. Boxplots of Shannon index that compare *H. pylori* negative and positive groups. The left boxplot represents *H. pylori* negative group and the right boxplot represents *H. pylori*

positive group. In the y-axis is the numerical value of Shannon index. Kruskal-Wallis statistics: H= 5.336977 (P = 0.020878).

The presence or absence of *Helicobacter pylori* did not produce differences in evenness (H= 1.214644, *P*= 0.270414). In boxplots there was not a clear difference (figure 4).

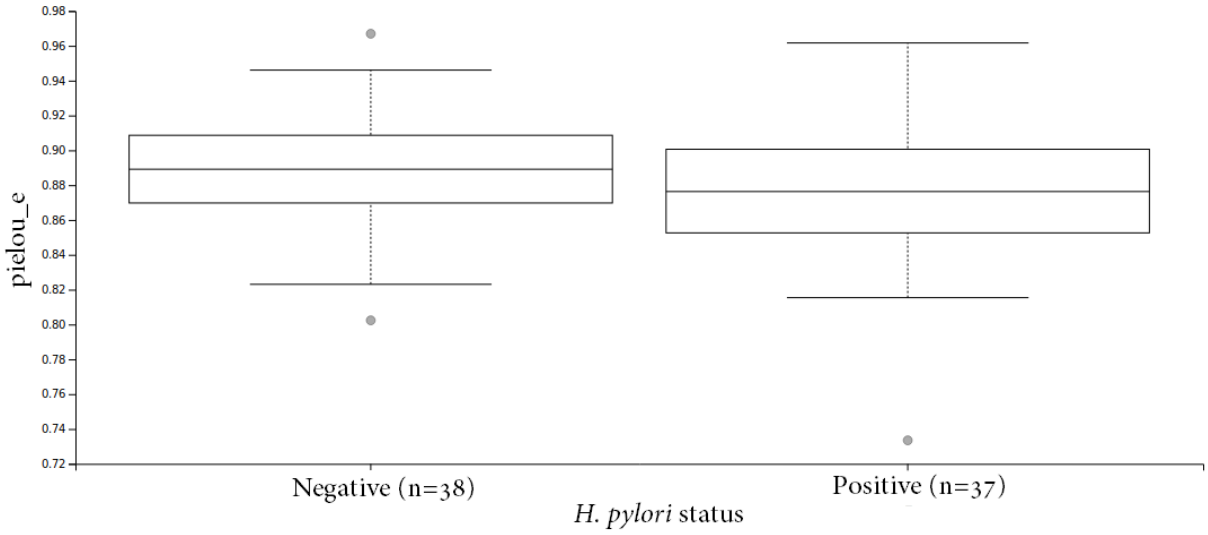
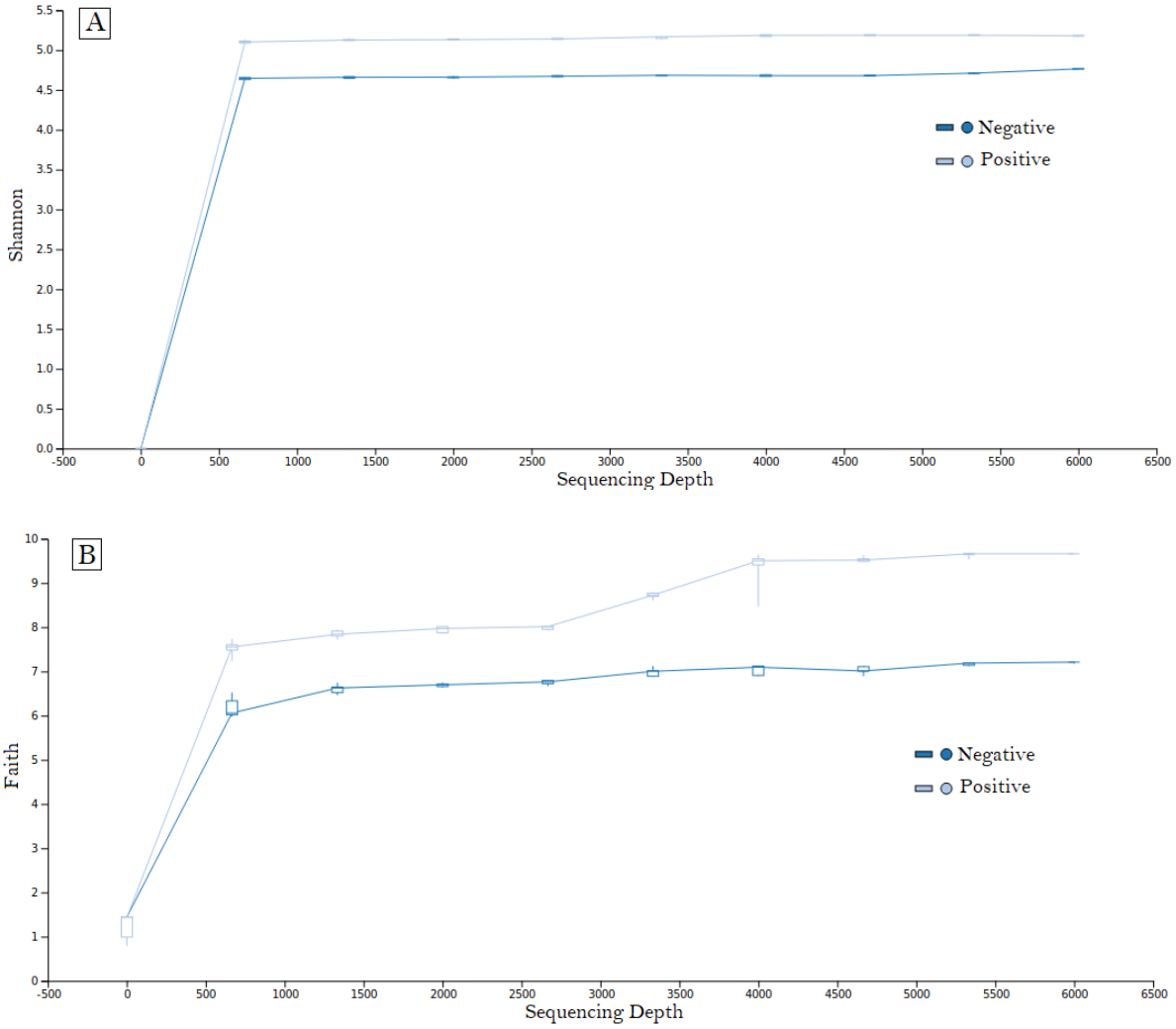


Figure 4. Boxplots of evenness that compare *H. pylori* negative and positive groups. The left boxplot represents *H. pylori* negative group and the right boxplot represents *H. pylori* positive group. In the y-axis is the numerical value of the Pielou's evenness index. Kruskal-Wallis statistics: H= 1.214644 (P= 0.270414).

There were not differences in alpha diversity between men and women: evenness (H= 1.826754, *P* = 0.176511), FD (H= 1.422214, *P* = 0.233039), SI (H= 1.264319, *P* = 0.260835).

Alpha rarefaction curves comparing the groups positive and negative for *H. pylori* infection were made. The metric used were Shannon index and Faith's phylogenetic diversity (figure 5). With both diversity indices, the lines reached a slope of 0 at a point close to 500 of depth of sequencing. But when Faith was used, the curve of the positives presented an

increase during the interval that goes from approximately 2700 to 4000 of depth of sequencing. After this the curve is leveled again (panel B of figure 5). The number of samples analyzed at sequencing depths higher than 2,700, are less than the actual number of samples (38 for positive and 37 for negative) (panel C of figure 5).



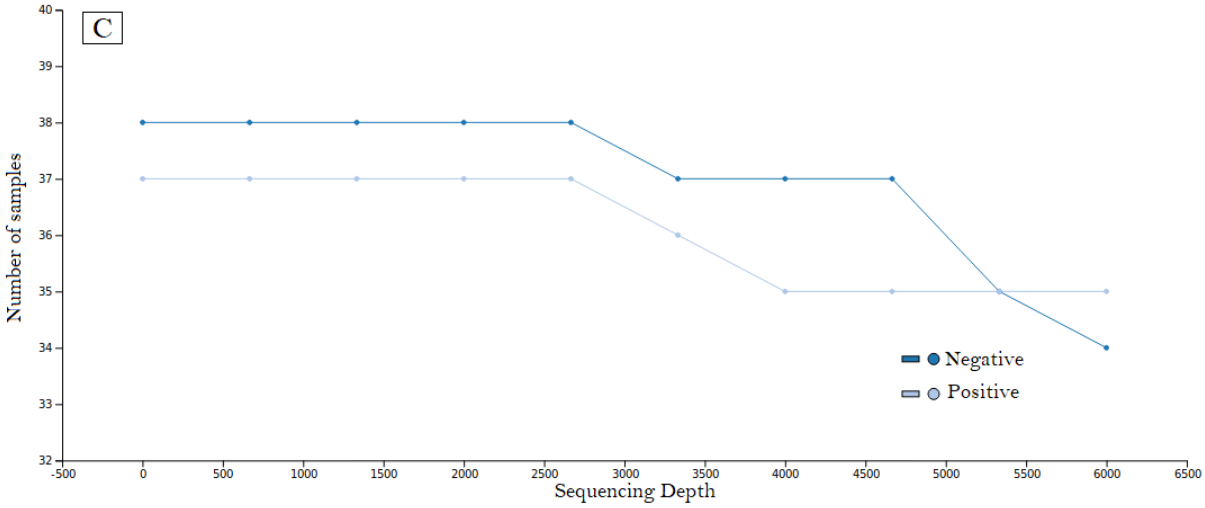


Figure 5. Alpha rarefaction curves comparing *H. pylori* positive and negative groups.

A) Comparison using Shannon index. B) Comparison using Faith's phylogenetic diversity. C) Number of samples in function of sequencing depth.

Beta diversity:

All beta diversity metric used were statistically significant (with PERMANOVA) when comparing *H. pylori* positive and negative groups. When PCoA was applied, there was not a clear difference between groups with all metrics. Similarly, there are not clear differences in boxplots. The details of the findings obtained with each of the beta diversity metrics are shown below.

Jaccard had a pseudo-F value of 1.515798 (*P*= 0.004) (boxplot in figure 6). Therefore, there was a difference in terms of presence-absence of OTUs between groups. The PCoA showed all samples (positive and negative) dispersed.

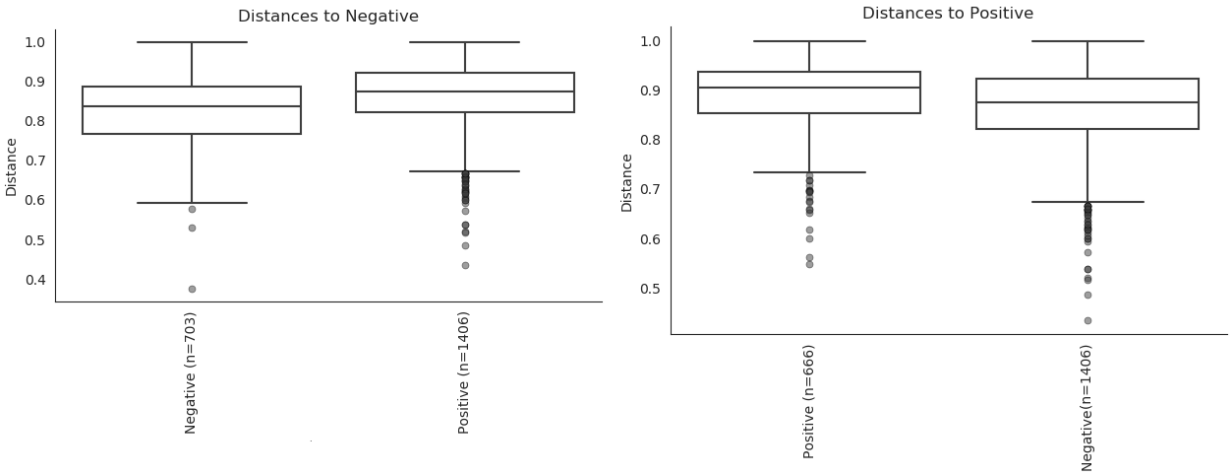


Figure 6. Boxplots of Jaccard comparing groups positive and negative for *H. pylori* infection. Boxplots from left to right show: distance between negative samples, distance from positive to negative samples, distance between positive samples and distance from negative to positive samples. Numbers between parenthesis that are next to positive and negative names are the number of measured distances.

In the PCoA of Bray-Curtis, there were positive and negative samples clustered together (green circle). Most negative samples are in the cluster and only six are outside. But there were positive and negative samples that moved away from the group (figure 7).

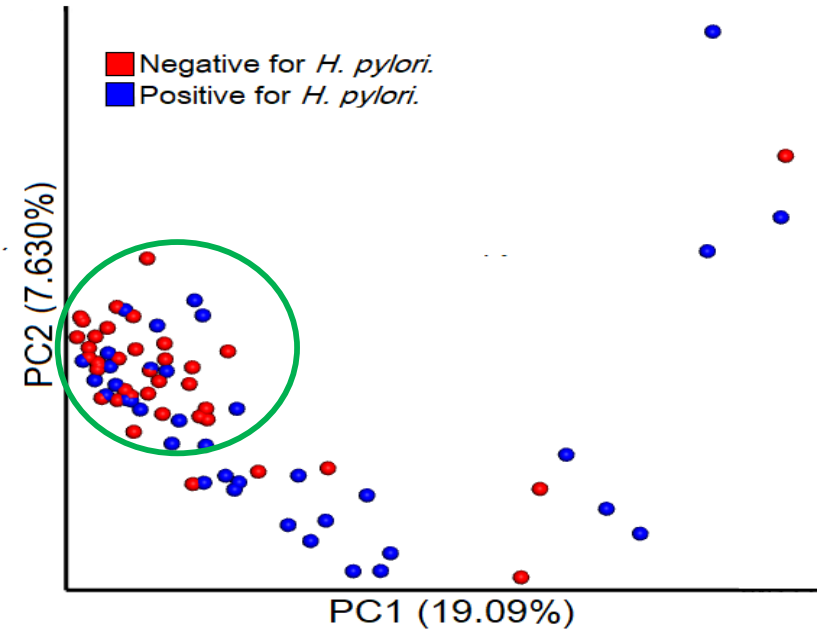


Figure 7. PCoA of Bray-Curtis comparing samples positive and negative for *H. pylori*. Green circle marks the cluster. Red and blue dots represent negative and positive samples respectively. PC means principal coordinate and represent the explained percentage of variation.

With PERMANOVA of Bray-Curtis, there was a significant difference between patients with and without *H. pylori* infection (pseudo-F= 2.1389, *P*= 0.005) (boxplots in figure 8). Hence, the groups are different in terms of abundance of OTUs.

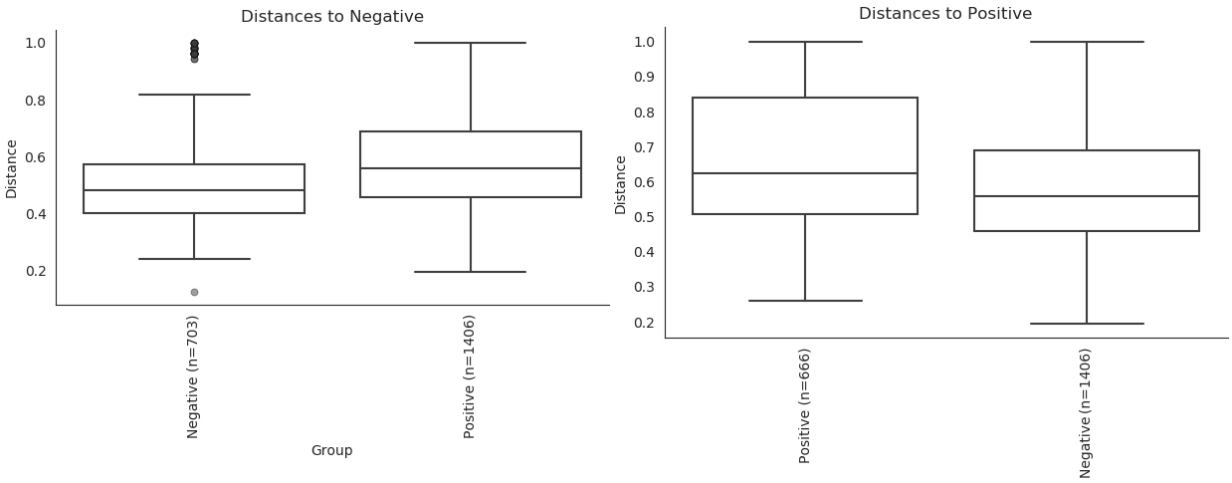


Figure 8. Boxplots of Bray-Curtis comparing groups positive and negative for *H. pylori* infection. Boxplots from left to right show: distance between negative samples, distance from positive to negative samples, distance between positive samples and distance from negative to positive samples. Numbers between parenthesis that are next to positive and negative names are the number of measured distances.

Pseudo-F value of Unweighted UniFrac was 1.997311 (*P*= 0.031) (see figure 9 for boxplots), which indicates that positive and negative samples were different in presence-absence and phylogenetic relationships of OTUs.

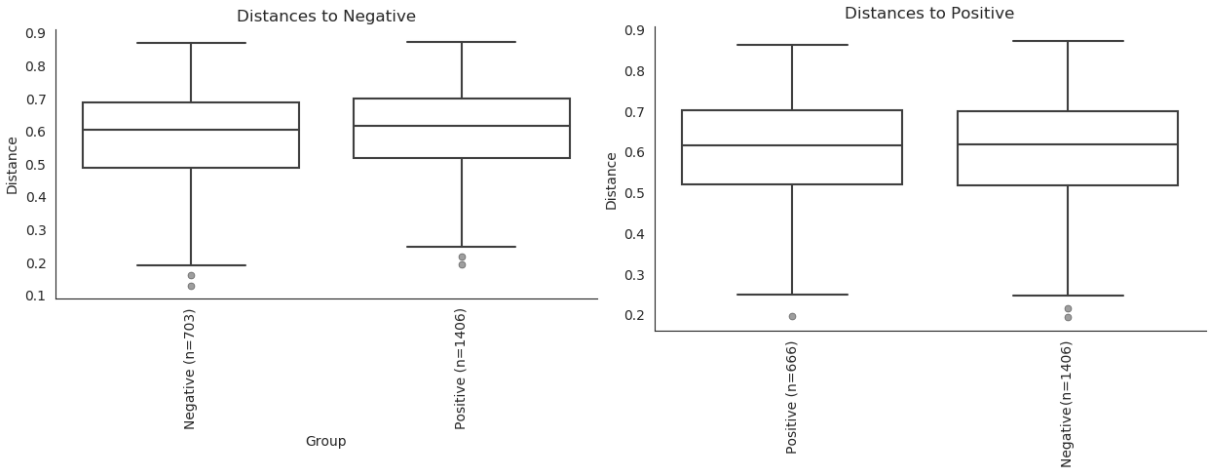


Figure 9. Boxplots of Unweighted UniFrac comparing groups positive and negative for *H. pylori* infection. Boxplots from left to right show: distance between negative samples, distance from positive to negative samples, distance between positive samples and distance from negative to positive samples. Numbers between parenthesis that are next to positive and negative names are the number of measured distances.

The PCoA showed three separate clusters (enclosed in green circles) composed of positive and negative samples (figure 10).

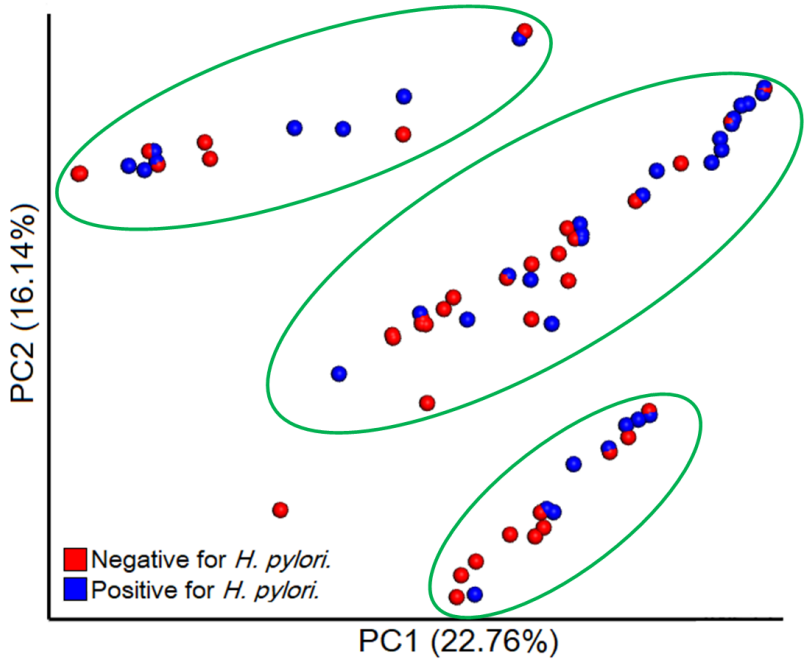


Figure 10. PCoA of Unweighted UniFrac comparing samples positive and negative for *H. pylori.* Green circles mark clusters. Red and blue dots represent negative and positive samples respectively. PC means principal coordinate and represent the explained percentage of variation.

Figure 11 shows the PCoA of Weighted UniFrac. It was observed that many samples (positive and negative) cluster together (green circle). While other samples dispersed and moved away from the cluster.

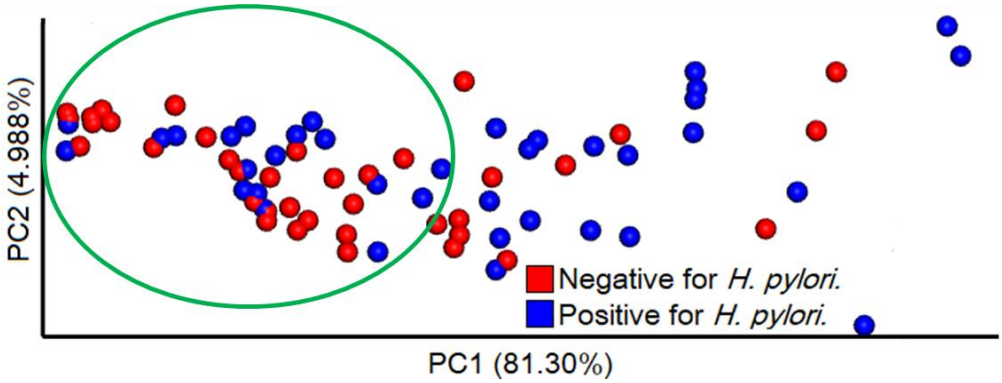


Figure 11. PCoA of Weighted UniFrac comparing samples positive and negative for *H. pylori.* Green circle marks the cluster. Red and blue dots represent negative and positive

samples respectively. PC means principal coordinate and represent the explained percentage of variation.

On the other hand, when PERMANOVA was applied a significant difference between groups was seen (pseudo-F= 5.308483, *P*= 0.017). Boxplots are shown in figure 12. This mean that *H. pylori* infected and non-infected are different in presence-absence, abundance, and phylogenetic relationships of OTUs.

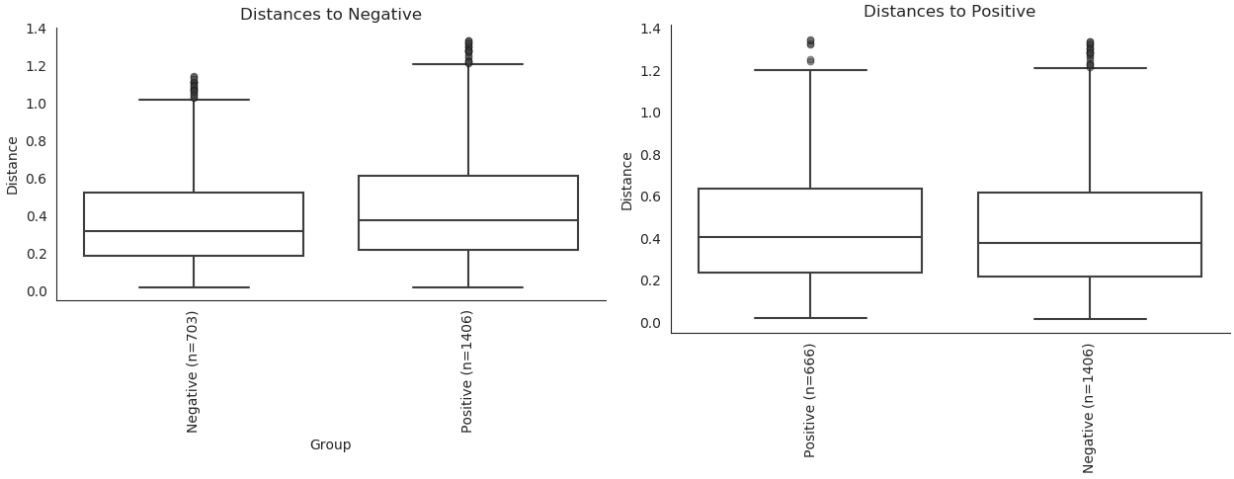


Figure 12. Boxplots of Weighted UniFrac comparing groups positive and negative for H. pylori infection. Boxplots from left to right show: distance between negative samples, distance from positive to negative samples, distance between positive samples and distance from negative to positive samples. Numbers between parenthesis that are next to positive and negative names are the number of measured distances.

Beta diversity indices were also applied to determine differences between men and women. Weighted Unifrac was the only metric statistically significant (pseudo-F= 4.214712, *P*= 0.038) (figure 13 for boxplots). PCoA, that is in figure 14, was like that performed using the variable *H. pylori* infection (figure 11). Gender, therefore, also has an influence on diversity of the studied population.

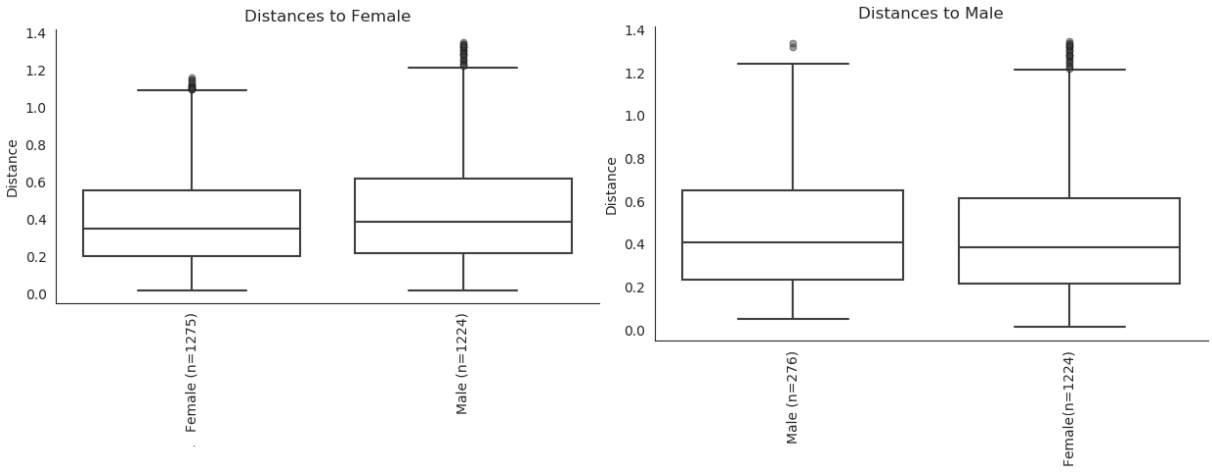


Figure 13. Boxplots of Weighted UniFrac comparing men and women. Boxplots from left to right show: distance between negative samples, distance from positive to negative samples, distance between positive samples and distance from negative to positive samples. Numbers between parenthesis that are next to positive and negative names are the number of measured distances.

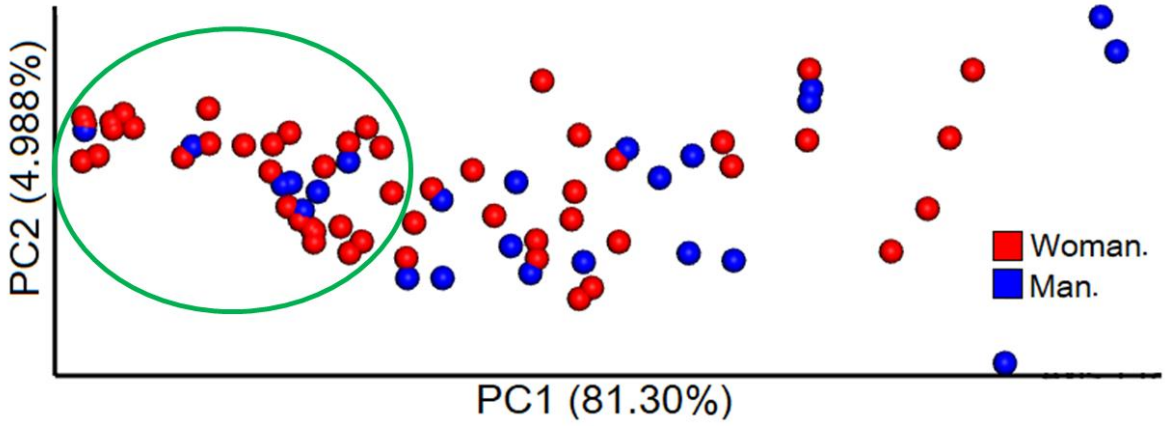
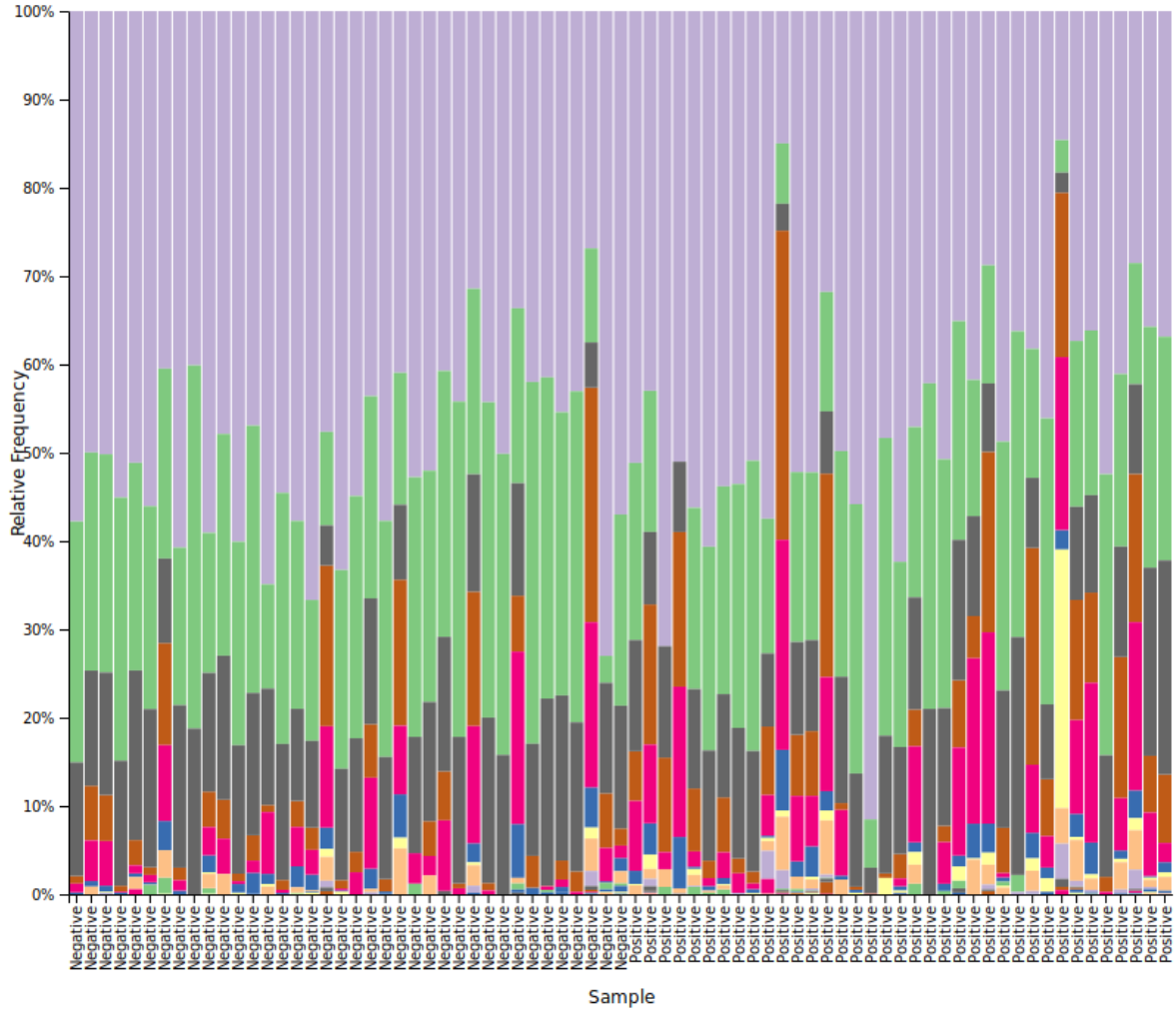


Figure 14. Figure 14. PCoA of Weighted UniFrac comparing samples of women and men. Green circle marks the cluster. Red and blue dots represent woman and man samples respectively. PC means principal coordinate and represent the explained percentage of variation.

Relative abundance:

In figure 15 are taxa barplots comparing *H. pylori* positive and negative samples at phylum level. The more predominant phyla in both groups were (in descending order): Proteobacteria (47.485% average in positive versus 48.159% average in negatives), Bacteroidetes (6.690% average in positive versus 6.516% average in negatives), Firmicutes (5.782% average in positive versus 5.607% average in negatives), Actinobacteria (1.383% average in positive versus 1.334% average in negatives) and Fusobacteria (1.162% average in positive versus 1.114% average in negatives).



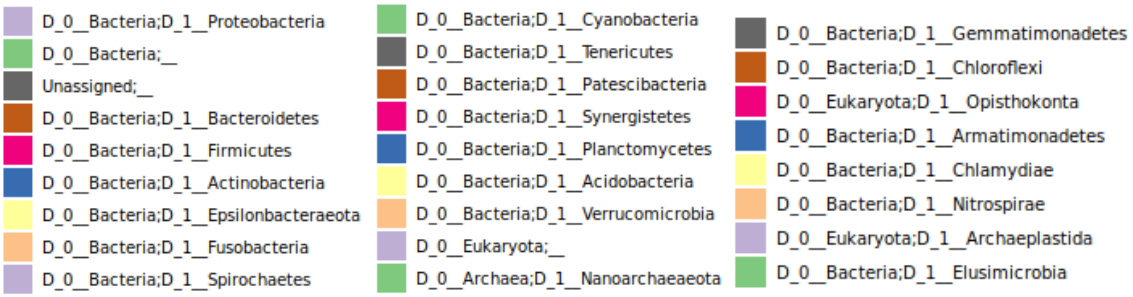


Figure 15. Taxa barplots comparing *H. pylori* positive and negative samples at phylum level.

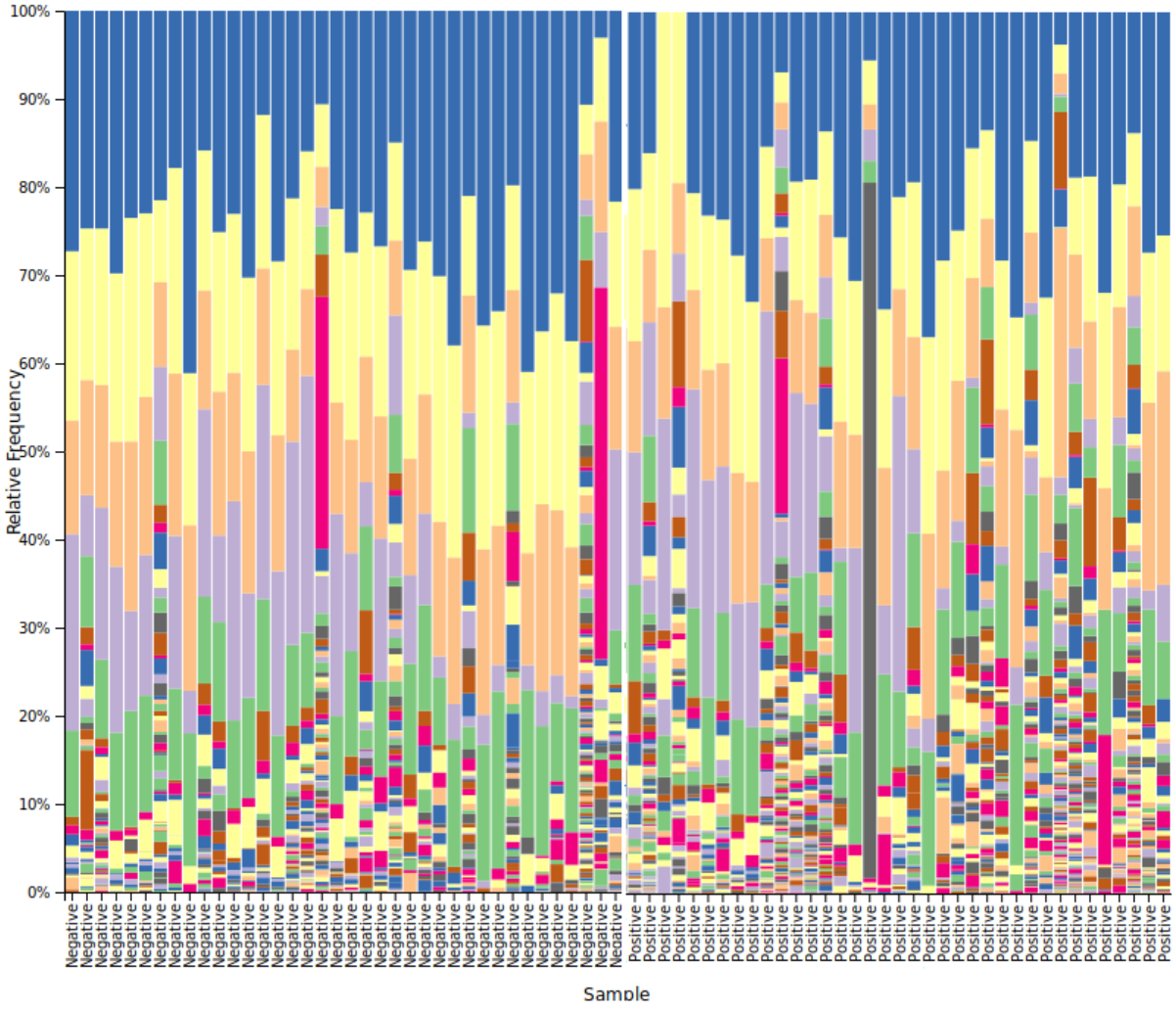
Also, fig 15 showed bars identified as bacteria (undefined phylum) and unassigned. From the inspection of the sequences was possible to identify 39 OTUs defined as bacteria (undefined phylum) and 99 OTUs defined as unassigned. After doing BLAST of the related sequences was possible to obtain the results presented in table 3.

In table 3 is possible to see that the majority of OTUs initially assigned as bacteria (undefined phylum) were later defined as uncultured bacterium isolate DGGE gel band Eub_1 16S ribosomal RNA gene, partial sequence (71.79%). Other types of uncultured bacteria were also found, and, as a group, this kind of bacteria constitute 94.87% of the OTUs. BLAST was also made for unassigned OTUs which were identified as 16S rRNA of Uncultured bacteria.

Table 3. Summary of the identification by BLAST of the OTUs defined as Bacteria (undefined phylum).

Bacteria	Accession	Number of OTUs	Percentage of OTUs
Uncultured bacterium isolate DGGE gel band Eub_1 16S ribosomal RNA gene, partial sequence.	GQ411109.1	28	71.79%
Uncultured bacterium clone 3-3 16S ribosomal RNA gene, partial sequence.	AY333095.1	9	23.08%
Cardinium endosymbiont of *Nitocra spinipes* partial 16S rRNA gene, clone 27.	FR878030.1	1	2.56%
Uncultured bacterium clone 6731 16S ribosomal RNA gene, partial sequence.	MF562002.1	1	2.56%

Ralstonia (10.163% average in positive versus 10.463% average in negatives), *Herbaspirillum* (1.230% average in positive versus 1.148% average in negatives), *Streptococcus* (2.248% average in positive versus 2.156% average in negatives), *Pseudomonas* (1.549% average in positive versus 2.039% average in negatives), *Haemophilus* (1.509% average in positive versus 1.501% average in negatives), *Burkholderia-Caballeronia-Paraburkholderia* (1.920% average in positive versus 1.946% average in negatives), *Neisseria* (1.149% average in positive versus 1.099% average in negatives) and *Veillonella* (0.623% average in positive versus 0.607% average in negatives) were the more abundant genera in positive and negative samples (figure 16).



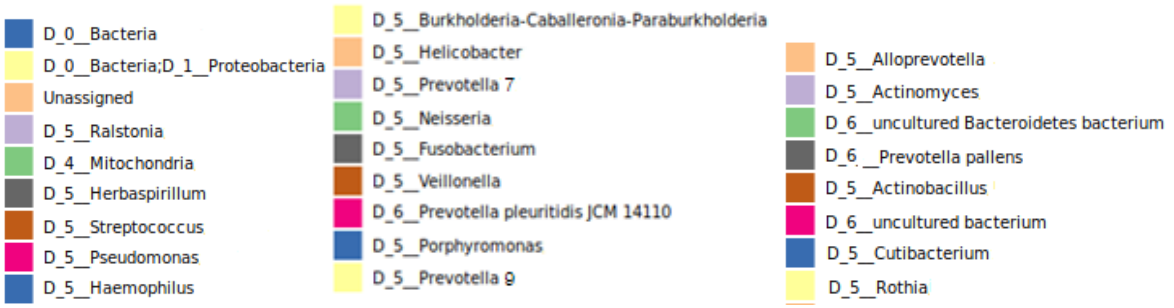


Figure 16. Taxa barplots comparing *H. pylori* positive and negative samples at genus/species level. It is presented the color key of the first 26 taxa.

Figure 17 shows in which positive samples (by gastric biopsy) was also possible to detect *Helicobacter* in duodenum. In 17 (45.95%) of the 37 positive samples there was presence of *Helicobacter*. Mean relative abundance was 2.376%, with a minimum of 0.072% (sample HDRP0006) and a maximum of 28.237% (HDRP0079). Furthermore, 8 samples correspond to women and 9 to men.

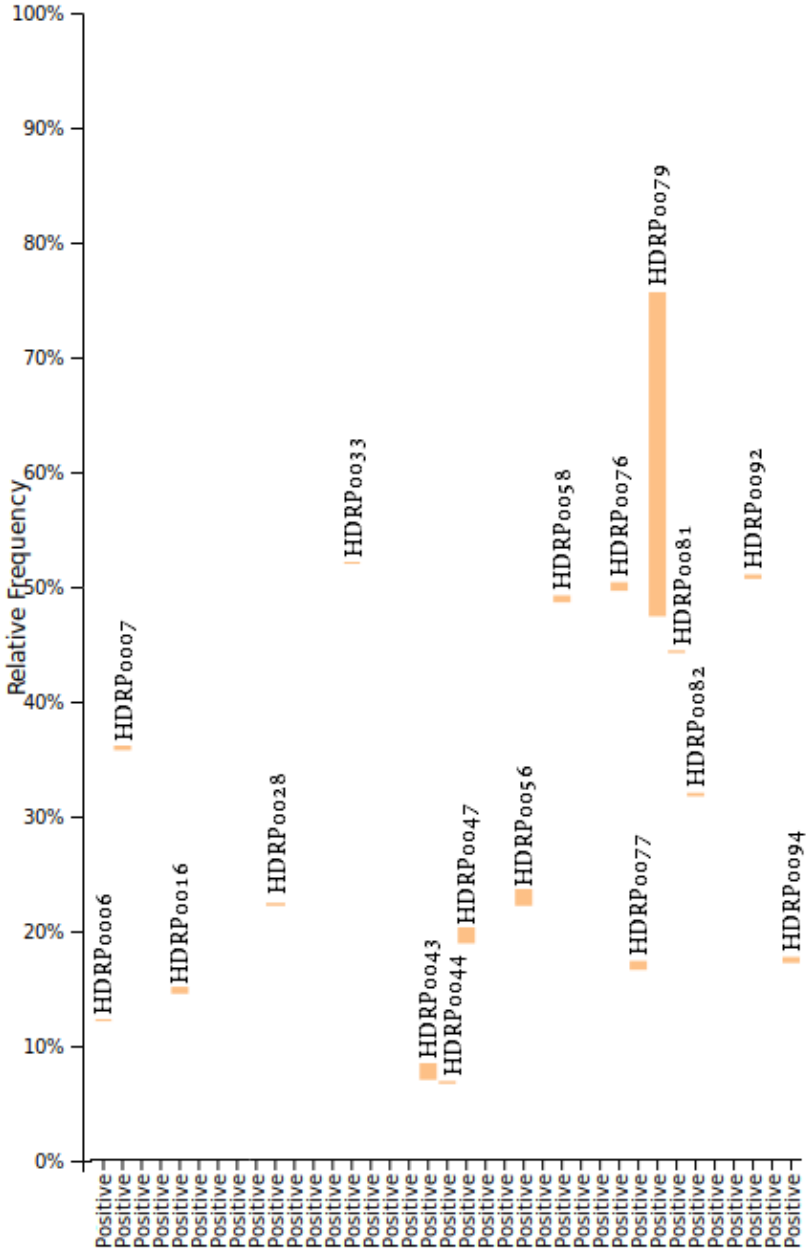


Figure 17. Taxa barplots showing positive samples (by gastric biopsy) in which *Helicobacter* could be detected in duodenum through 16S rRNA sequencing. Orange bars represent presence of Helicobacter. Above the bars are sample codes.

Helicobacter phylogeny and OTUs distribution:

Twenty-one OTUs were taxonomically assigned as *Helicobacter.* Each duodenal sample in which it was possible to find a *Helicobacter*, there was a different OTU of this

bacterium. The exception is sample HDRP0079, which has six different OTUs (figure 18). It even shared an OTU in common with the HDRP0007 sample. Moreover, four of the OTUs preset in HDRP0079 had a high frequency. In BLAST, all *Helicobacter* OTUs were identified as *H. pylori*. Identity percentages were greater than 98%, except in OTUs 18 (95.82%) and 19 (91.71%).

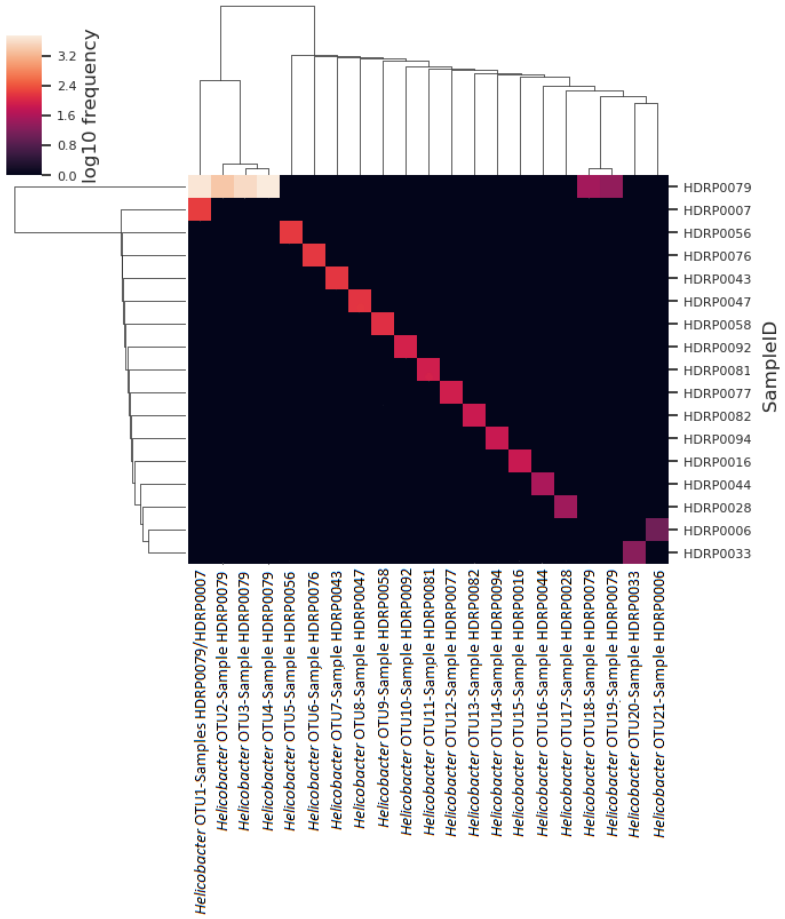


Figure 18. Heatmap showing the distribution and frequency of OTUs of Helicobacter in duodenal samples. At the bottom are *Helicobacter* OTUs and on the right the code of samples. Frequency is represented by colors: black means absence and the clearer the square, the greater the frequency.

With *Helicobacter* OTUs encountered in the study and reference sequences obtained from the GeneBank (see methodology), a phylogenetic tree was made using maximum likelihood and Hasegawa-Kishino-Yano model (figure 19). Notice that *Helicobacter* OTUs 18 and 19 were separated of the other OTUs.

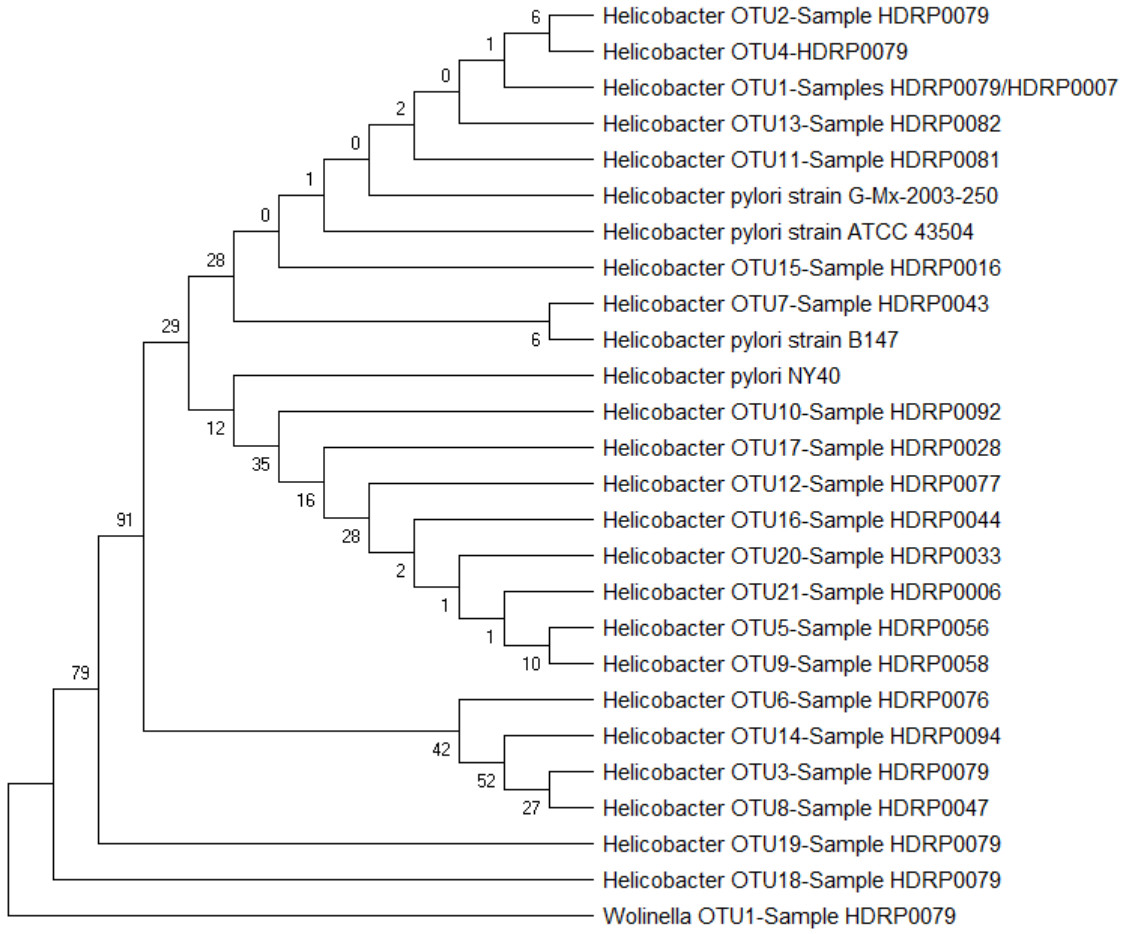


Figure 19. Phylogenetic tree of OTUs assigned as *Helicobacter.* In the tree are all *Helicobacter* OTUs encountered in the study. Also are incorporated four reference sequences. In the tip of the branches are the codes for the OTUs (Including the sample/s from which they come) or the manes of reference sequences. An OTU of Wolinella, obtained

in one sample clinical sample of this study, was used as outgroup. Numbers in the branches represent the bootstrap percentage. Highest log likelihood (-985.96)

Finally, a phylogenetic tree was made with the same OTUs (figure 20). Like the previous tree, OTUs did not form a unique clade and, therefore, they are generally different. Also OTUs 18 and 19 are separated from the rest.

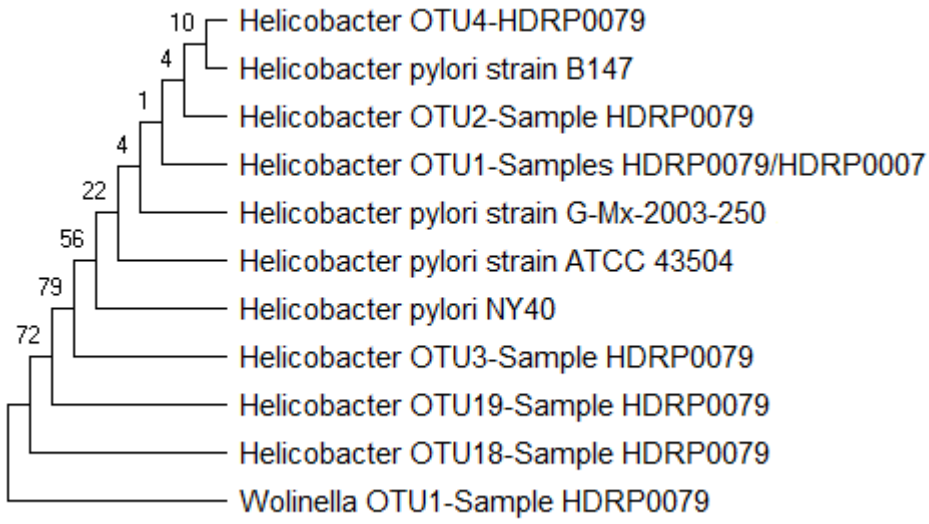


Figure 20. Phylogenetic tree of OTUs assigned as Helicobacter in sample HDRP0079.
In the tree are *Helicobacter* OTUs encountered in sample HDRP0079. Also are incorporated four reference sequences. In the tip of the branches are the codes for the OTUs (Including the sample/s from which they come) or the manes of reference sequences. An OTU of Wolinella, obtained in one sample clinical sample of this study, was used as outgroup. Numbers in the branches represent the bootstrap percentage. Highest log likelihood (941,56).

Differential abundance:

When comparing *H. pylori* positive and negative samples with ANCOM, the genus *Helicobacter* (W=383, clr=1.94) was the only taxa more abundant in the positives (figure 21).

This finding is expected because only in positive duodenal samples was found *Helicobacter.* However, this information did not contribute in the understanding of the microbiome differences between groups.

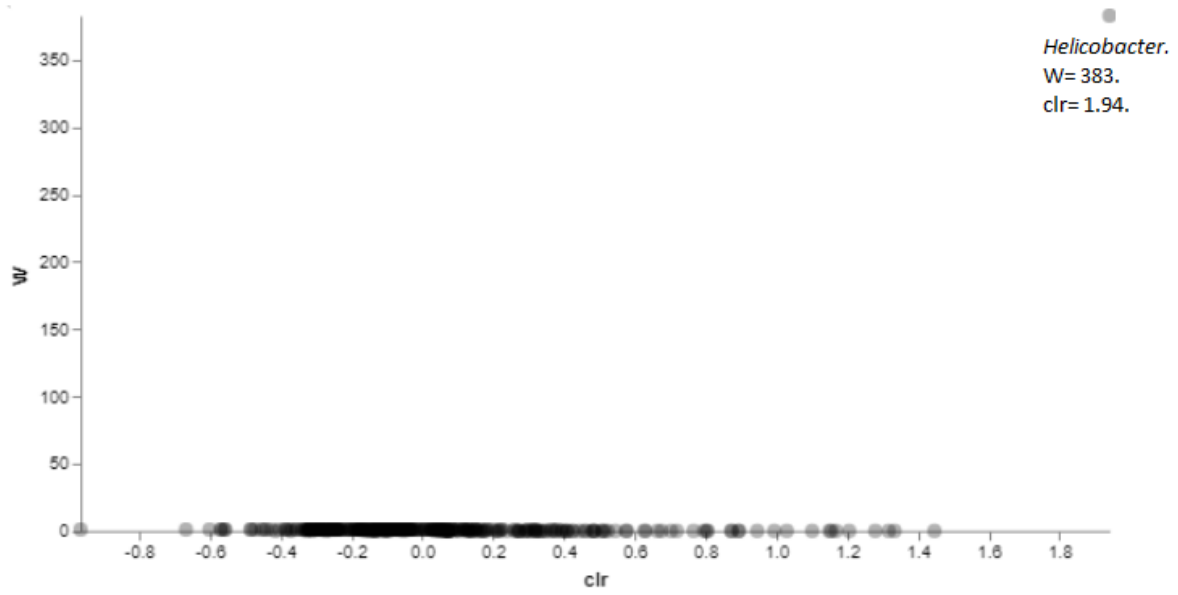


Figure 21. Volcano plot showing differentially abundant taxa between *H. pylori* positive and negative samples (genus level).

In the case of gender, there was not differentially abundant taxa between male and female when using ANCOM. This was seen in all taxonomic levels.

In figure 22 are the differentially abundant taxa between *H. pylori* positive and negative samples using Gneiss at order level. Positive samples were more abundant in: Pasteurellales, Betaproteobacteriales, Bacteroidales, Lactobacillales and Bacteria (undefined phylum). Negative samples were more abundant in other type of Betaproteobacteriales.

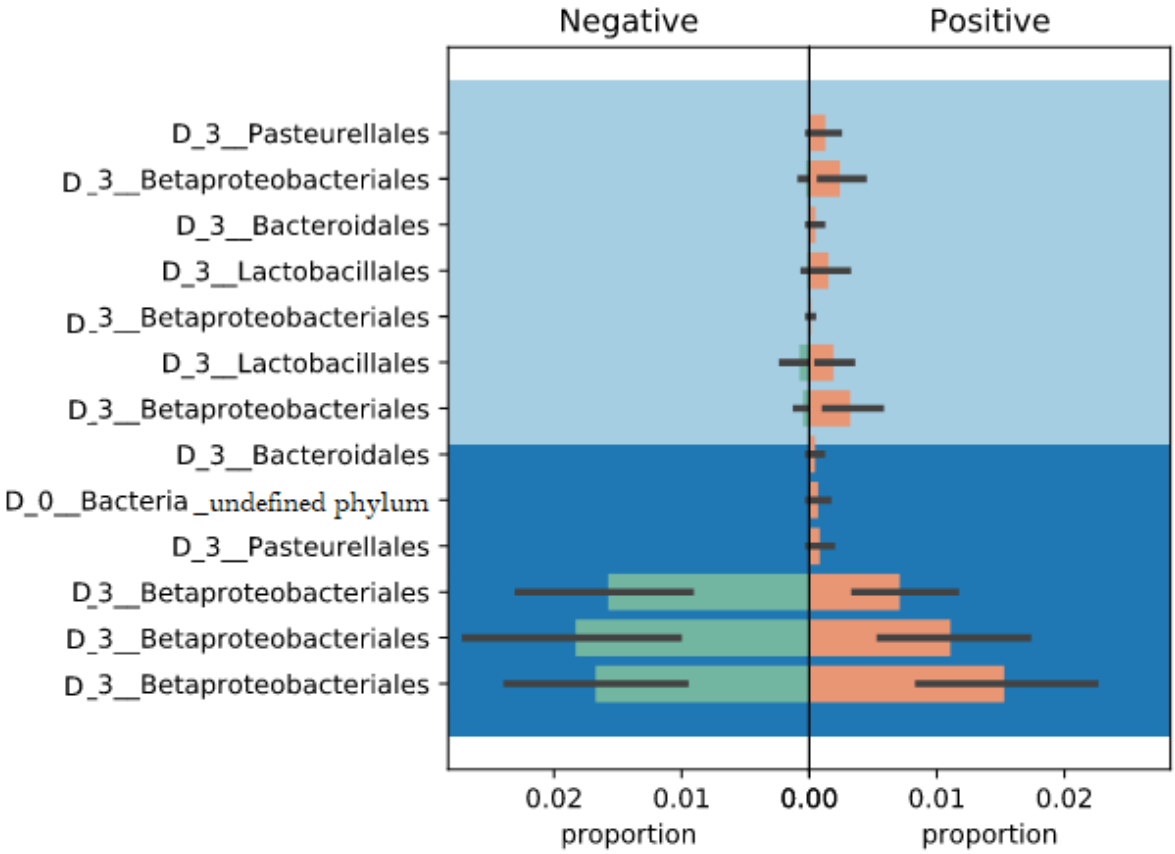


Figure 22. Gneiss proportion plot of differential abundance between H. pylori positive and negative samples (order level). Black bar represents error bars. Orange bars represent an increase in the relative abundance of the taxa in the positive group. Green bars represent an increase in the relative abundance of the taxa in the negative group.

Pasteurellaceae, Neisseriaceae, Prevotellaceae, *Streptococcaceae* and Bacteria (undefined phylum) were more abundant in positive samples when Gneiss analysis was done at family level (figure 23). On the other hand, in negative samples the more abundant taxon was *Burkholderiaceae.*

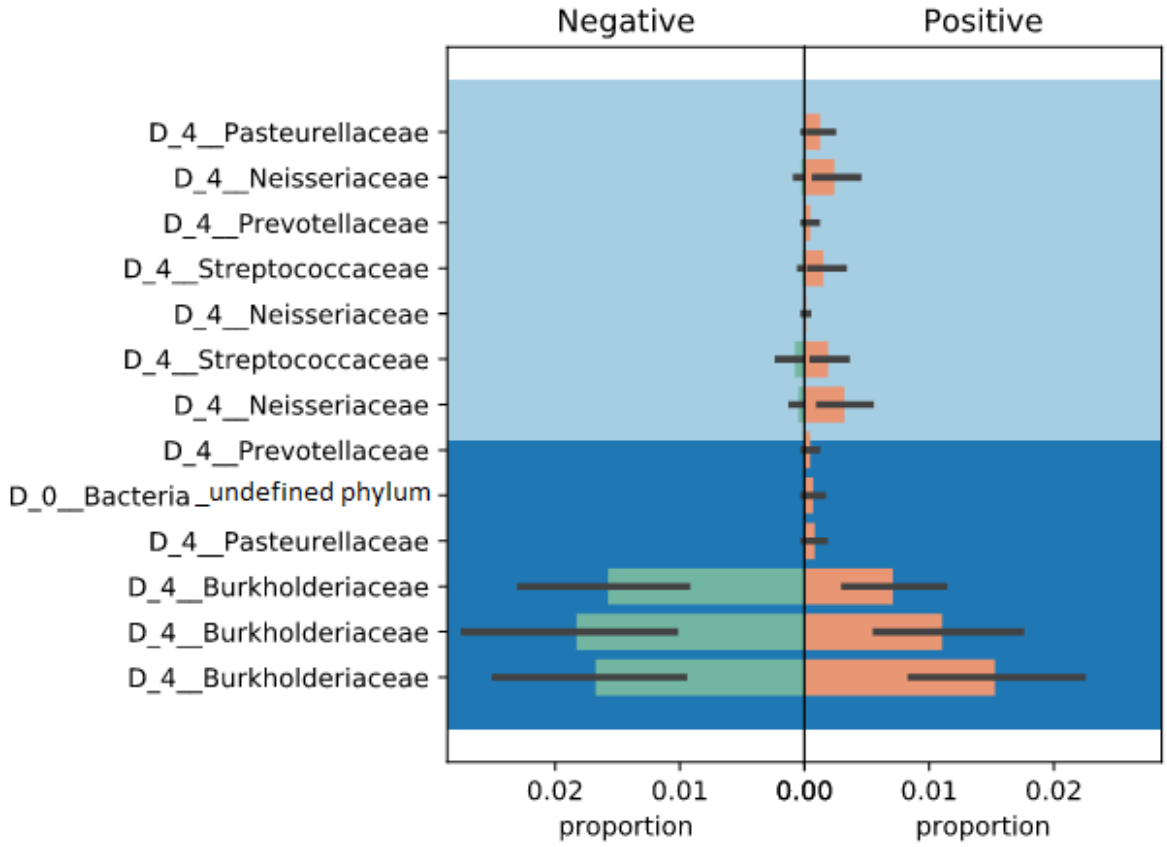


Figure 23. Gneiss proportion plot of differential abundance between H. pylori positive and negative samples (family level). Black bar represents error bars. Orange bars represent an increase in the relative abundance of the taxa in the positive group. Green bars represent an increase in the relative abundance of the taxa in the negative group.

When Gneiss was applied at genus/species, the taxa *Haemophilus, Neisseria, Prevotella pallens, Prevotella 7, Streptococcus* and Bacteria (undefined phylum) were the most abundant in positive samples. *Ralstonia*, on the other hand, was more abundant in the negative samples (figure 24).

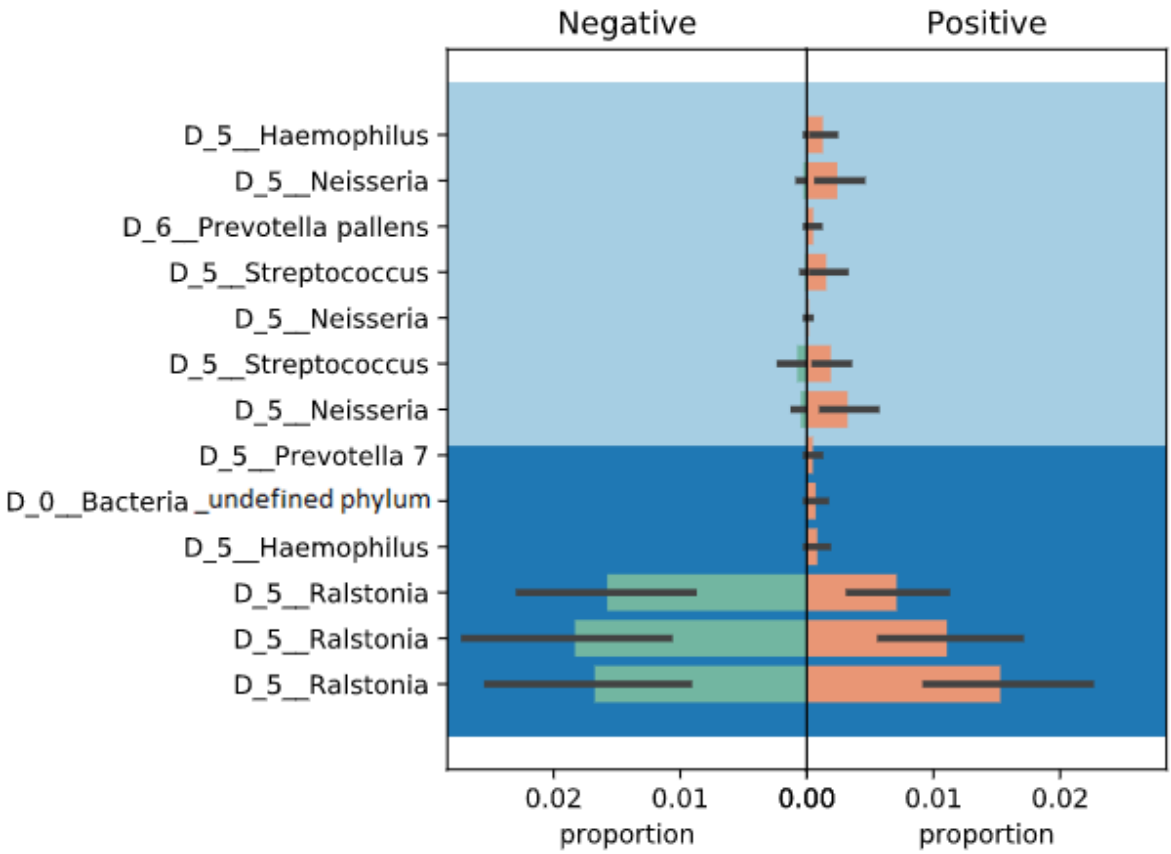


Figure 24. Gneiss proportion plot of differential abundance between *H. pylori* positive and negative samples (genus/species level). Black bar represents error bars. Orange bars represent an increase in the relative abundance of the taxa in the positive group. Green bars represent an increase in the relative abundance of the taxa in the negative group.

Gender was also analyzed with Gneiss. In figure 25 are shown the differentially abundant taxa at genus level. *Ralstonia* was more abundant in men while in women were *Streptococcus* and *Pseudomonas*. Notice that in women also a *Ralstonia* was more abundant compared with men. It was not possible to differentiate between these two bacteria because the taxonomic assignment only could reach genus.

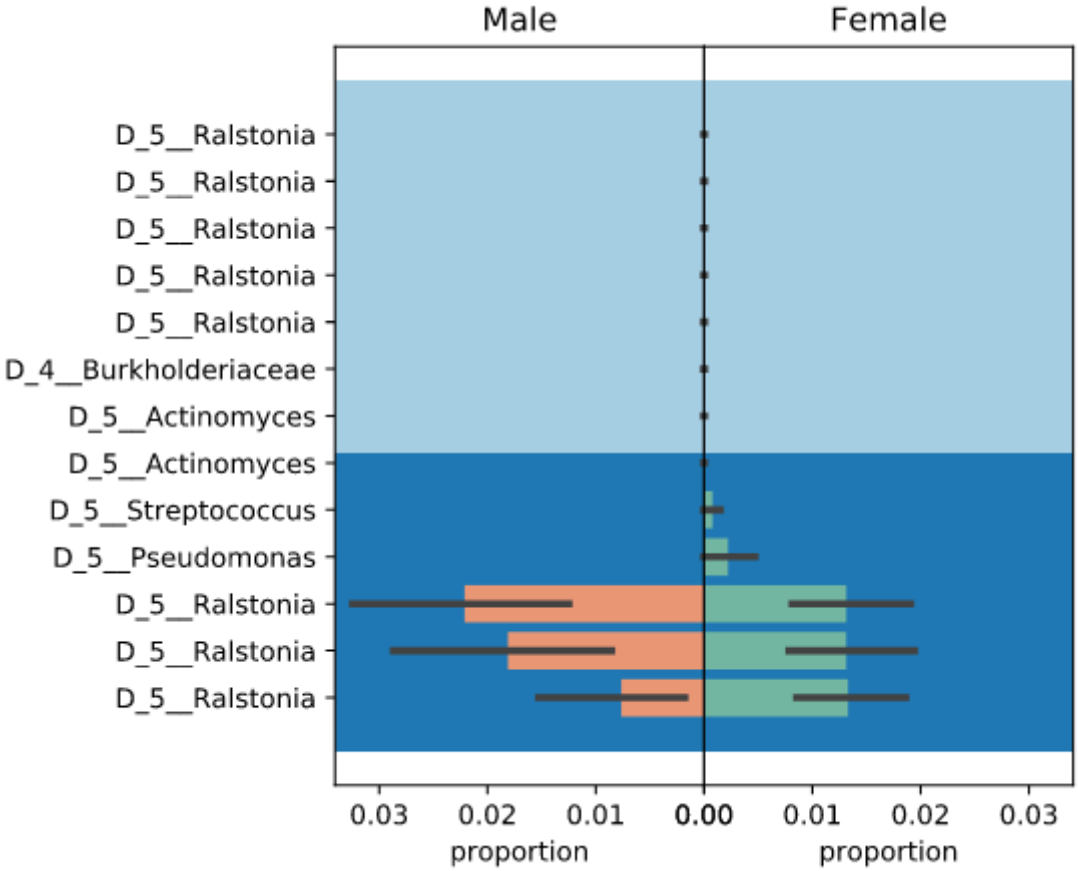


Figure 25. Gneiss proportion plot of differential abundance between women and men (genus level). Black bar represents error bars. Orange bars represent an increase in the relative abundance of the taxa in men. Green bars represent an increase in the relative abundance of the taxa in women.

Linear models with Gneiss:

Using the balances obtained in Gneiss, a linear regression model was generated. The model can explain approximately 4% (R2 = 0.0427) of the variance of the population. The variable that contributes most was gender (1.44%), followed by age (1.37%) and positive *H. pylori* (1.30%). From the validation process it can be concluded that there is no over-fitting. Because the accuracy of the prediction (pred_mse) is less than the model error (model_mse) in all partitions of data (table 4). The interpretion of the validation process and how to

conclude that there is not over-fitting is based on Morton and colleagues paper and Qiime2 tutorials (Morton et al., 2017; Qiime2, 2019).

Table 4. Validation process of the linear model made with balances in Gneiss.

	Model_mse	R^2	Pred_mse
Fold 0	21.9166	0.0570	3.5241
Fold 1	22.5087	0.0549	2.7213
Fold 2	23.0401	0.0582	2.1836
Fold 3	21.0928	0.0575	4.1789
Fold 4	20.2808	0.0582	5.0668
Fold 5	22.5904	0.0575	2.5835
Fold 6	23.6022	0.0543	1.4938
Fold 7	22.1094	0.0580	3.1041
Fold 8	21.0238	0.0557	4.2612
Fold 9	25.0022	0.0501	0.0000

Model_mse: model error; Pred_mse: Accuracy of the prediction.

Prediction of the functional profile realized with PICRUSt2:

Using as reference the orthologs of Kyoto Encyclopedia of Genes and Genomes (KEGG) and the statistical method DESeq2, it was identified 65 genes which are differentially expressed between the positive and negative groups for *H. pylori*. In table 5 are the results of the first 10 genes, ordered according to the value of the adjusted P-value (FDR), from lowest to highest. Notice that the first three genes code proteins that are part of the ComB system, which allows the transformation of *H. pylori*. The other genes are related with enzymatic restriction, methylation (DNA adenine methylase), toxin-antitoxin systems (mRNA interferase yafQ and Antitoxin hicB.), thiamine metabolism (hydroxyethylthiazole kinase), purine metabolism (GMP reductase) and carbohydrate metabolism (UDP-4-amino-4,6-dideoxy-L-N-acetyl-beta-L-altrosamine transaminase).

Table 5. Orthologs of the KEGG which were differentially expressed between *H. pylori* positive and negative groups.

KEEG cod	Description	log2FC	lfcSE	*P*-values	FDR
K12053	ComB4 competence protein.	6.0405	0.85304	1.43E-12	8.28E-09
K12049	ComB9 competence protein.	5.503	0.84368	6.91E-11	1.57E-07
K12050	ComB8 competence protein.	5.5005	0.84644	8.12E-11	1.57E-07
K01155	Type II restriction enzyme.	1.4392	0.30028	1.64E-06	0.0023813
K06223	DNA adenine methylase.	1.1793	0.25583	4.03E-06	0.0046733
K19157	mRNA interferase YafQ.	1.3521	0.32432	3.06E-05	0.025532
K00878	Hydroxyethylthiazole kinase.	1.548	0.37148	3.09E-05	0.025532
K00364	GMP reductase.	1.6157	0.3916	3.69E-05	0.026737
K15895	UDP-4-amino-4,6-dideoxy-L-N-acetyl-beta-L-altrosamine transaminase.	3.5494	0.87414	4.90E-05	0.027544
K18843	Antitoxin HicB.	1.4212	0.35092	5.13E-05	0.027544

Log2FC: log2 fold change, expresses how much the gene expression change due to the presence of *H. pylori.* lfcSE: standard error of the log2FC. FDR: False discovery rate, the p value adjusted with the Benjamini-Hochberg method.

Differential expression was also evaluated using as reference the Enzyme Commission number (EC). In this case was identified 155 enzymes with statistically significant differential expression between positives and negatives. As previously, in table 6 are the first 10 enzymes according to the FDR. Type II restriction enzyme, GMP reductase, UDP-4-amino-4,6-dideoxy-L-N-acetyl-beta-L-altrosamine transaminase, Hydroxyethylthiazole kinase and DNA adenine methylase were also detected when the orthologs of the KEGG was used. The exceptions were Thiamine diphosphokinase (thiamine metabolism), Sorbitol-6-phosphate 2-dehydrogenase (carbohydrate metabolism), Phospholipase A(1/2) (lipid metabolism) and asparaginase (aminoacids metabolism).

Table 6. Enzymes which were differentially expressed between *H. pylori* positive and negative groups.

EC number.	Description	log2FC	lfcSE	*P*-values	FDR
3.1.21.4	Type II restriction enzyme.	1.454	0.30705	2.18E-06	0.004066
1.7.1.7	GMP reductase.	1.6037	0.39813	5.62E-05	0.031463
2.6.1.92	UDP-4-amino-4,6-dideoxy-L-N-acetyl-beta-L-altrosamine transaminase.	3.4971	0.87807	6.81E-05	0.031463
2.7.6.2	Thiamine diphosphokinase.	0.77091	0.19886	0.00010591	0.031463
2.7.1.50	Hydroxyethylthiazole kinase.	1.4236	0.36747	0.00010708	0.031463
1.1.1.140	Sorbitol-6-phosphate 2-dehydrogenase.	0.58876	0.15483	0.00014321	0.031463
3.1.1.4	Phospholipase A(2).	1.5786	0.4207	0.00017522	0.031463
3.1.1.32	Phospholipase A(1).	1.5777	0.42086	0.0001777	0.031463
3.5.1.1	Asparaginase.	0.44564	0.11992	0.00020214	0.031463
2.1.1.72	DNA adenine methylase.	0.35438	0.095523	0.00020734	0.031463

Log2FC: log2 fold change, expresses how much the gene expression change due to the presence of *H. pylori.* lfcSE: standard error of the log2FC. FDR: False discovery rate, the p value adjusted with the Benjamini-Hochberg method.

When entire metabolic pathways were analyzed (using EC and MetaCyc Metabolic Pathway Database), none had a statistically significant differential expression. The best result was the superpathway of demethylmenaquinol-6 biosynthesis II (log2FC: 2.8877, lfcSE: 0.8823, p-value: 0.0010643 and FDR: 0.082874). This pathway is involved in the synthesis of menaquinone form chorismite, using as intermediary futalosine (Hiratsuka et al., 2008).

Analysis of the mycobiome:

After the taxonomic assignment using the database UNITE 8.0, it was observed that the majority of samples had OTUs that were defined as Fungi (undefined phylum): 36 samples *H. pylori* positive and 38 samples *H. pylori* negative. Only one positive samples did

not have unassigned OTUs. The three fungal taxa encountered in more samples were Recurvomyces (36 samples; 16 positive and 20 negative), Ascomycota (22 samples; 11 positive and 11 negative) and *Xenoacremonium falcatum* (13 samples; 5 positive and 8 negative). There was a taxon that was found only in negative samples (Agaricomycetes). All taxa encountered are summarized in table 7.

Table 7. Taxa identified in samples positive and negative for *H. pylori* infection during the analyses of ITS sequences.

Taxa	*H. pylori* positive	%	*H. pylori* negative	%	Total
Fungi (undefined phylum)	36	48.65	38	51.35	74
Ascomycota	11	50	11	50	22
Recurvomyces	16	44.44	20	55.56	36
Xenoacremonium falcatum	5	38.46	8	61.54	13
Basidiomycota	5	41.67	7	58.33	12
Agaricomycetes	0	0	3	100	3
Chytridiomycota	5	71.43	2	28.57	7

In *H. pylori* positive and negative columns are the number of samples in which was identified every taxa.

DISCUSSION

This is the first study in Ecuador describing the duodenal microbiome in patients with *H. pylori* infection and one the few describing the duodenal microbiome worldwide. Our results showed that *H. pylori* infection produces changes in diversity (alpha and beta diversity) and increases the abundance of some potentially pathogenic taxa (*Haemophilus, Neisseria, Prevotella pallens, Prevotella 7* and *Streptococcus*).

General characteristics of the population

The *H. pylori* infection rate was 50.67% which is similar to previous reports in Ecuador (40.2% to 72.2%). As mentioned above, the average age was calculated based on information from only 71 patients (94.67% of the total) constituting a limitation of the study. Additionally, the group of negatives is composed mostly of women and their microbiome could have be influenced by this factor.

16s rRNA V4 region

There is not an absolute standard of what 16s rRNA region must be used in microbiome studies. V4 region is recommended in the Earth Microbiome Project, has been applied in many intestinal microbiome studies, permits the identification of bacteria and archaea (specially *Methanobrevibacter* and *Methanosphaera* which are important in the gut), have good genus level resolution and it is useful in environment where there may be new bacterial lineages (duodenum) (Allaband et al., 2019; Yang, Wang, & Qian, 2016). For these reasons, it is a good option for this study.

Alpha and beta diversity

Alpha diversity means diversity inside the sample and beta diversity means diversity between samples (Cox, Cookson, & Moffatt, 2013). Alpha diversity indexes used in this study have been applied in many microbiome studies and are reliable measure of richness and evenness of the community. We used two indices of richness (Faith's Phylogenetic Index and Shannonn Index) because it gives complementary information. In both cases they are quantitative measures of richness but Faith's Phylogenetic Index also takes into account the phylogenetic relationships between OTUs (Cox et al., 2013; Faith, 1992). *H. pylori* infection modified alpha diversity (Faith's phylogenetic diversity and Shannon index but not evenness) in the present study. Shannon Index differs and evenness coincide with previous studies on fecal microbiome. Gao et al. reported that there is not difference in SI between the groups without current *H. pylori* infection and with current infection (Gao et al., 2018). Similarly, Chen et al. showed that there is not significant change in SI and evenness between positive and negative groups (Chen et al., 2018).

Schulz et al. studied the gastric and duodenal microbiome. Phylotype richness, Shannon index and evenness were significantly different when comparing gastric biopsies of patients with and without *H. pylori* infection. These results were not seen in duodenal biopsies where the differences, for the three indices, were not significative (Schulz et al., 2018). As a result, at the duodenum only evenness findings obtained by Schulz et al. are consistent with the current study. In a more recent study by Han et al, Chao1 and SI were used. Both indices were significantly higher when *H. pylori* was present, a finding consistent with the current work (H. Han et al., 2019).

The curves of alpha rarefaction showed, lines that reached a slope of zero when the depth of sequencing was approximately 500, indicating that at this sequencing depth we appreciate a real picture of the microbiome richness. The curve of the positive samples generated using Faith´s phylogenetic index showed an increase at 4000 of depth of sequencing. This could suggest that at 4,000 of depth of sequencing (and not 500) all the richness is seen. However, above 2700 of depth of sequencing, there were samples excluded from the analysis and; therefore, from that value the information is not reliable.

The presence of *H. pylori* infection also influenced beta diversity. Although PCoA and boxplots did not show a clear difference between groups in Jaccard, Bray-Curtis, Unweighted UniFrac and Weighted UniFrac, the statistic test PERMANOVA showed significant differences. In the PcoA of Bray-Curtis, Unweighted UniFrac and Weighted UniFrac there is a group of positive and negative samples that form a cluster. On the other hand, there is other group of positive and negative samples that are not clustered. This suggests the presence of small differences between the patients with and without infection.

In relation with beta diversity, previous studies like Gao et al. showed no differences between groups. They compared the fecal microbiome using Bray-Curtis (groups analyzed: current infection vs negative) (Gao et al., 2018). In other work that studied gastric biopsies, authors found statistically significant differences using Bray-Curtis. In the same study, duodenal biopsies of *H. pylori* negative and positive patients were not distinct also using Bray-Curtis (Schulz et al., 2018).

Changes in diversity due to the presence of *H. pylori* infection has not always been observed, and can be explained because different studies have used different methodologies, sample sizes and populations. Likewise, the same anatomical site has not

always been analyzed. There is little information on diversity at the level of the duodenum to be able to make a better comparison with the results of the present study.

Relative abundance of taxa

It was determined that Proteobacteria, Bacteroidetes, Firmicutes, Actinobacteria and Fusobacteria were the more abundant phyla in duodenal tissue. These phyla had been described in previous studies of the duodenal microbiome made in aspirates and biopsies in patients positive and negative for *H. pylori* (Li et al., 2015; Schulz et al., 2018).

In a previous study they found at genus level, *Streptococcus, Prevotella, Actinomyces, Propionibacterium, Gemella, Fusobacterium, Leptotrichia, Veillonella, Haemophilus* and *Alloprevotella* were the more abundant in duodenal samples in infected and uninfected patients (Schulz et al., 2018). In studies that have included patients with gastritis and celiac disease, *Acinetobacter* and *Neisseria* have been described (D'Argenio et al., 2016; H. Han et al., 2019). In healthy individuals *Acinetobacter*, *Bacteroides*, *Prevotella*, *Stenotrophomonas*, and *Streptococcus* were the more prominent genera (G. Li et al., 2015). In the present study, in addition to the genera mentioned above, the most abundant were *Ralstonia*, *Herbaspirillum*, *Pseudomonas*, *Phorphyromonas, Rothia, Actinobacillus, Cutibacterium and Burkholderia-Caballeronia-Paraburkholderia*.

The presence of bacteria (undefined phylum) and unassigned OTUs can be a consequence of the limitations of databases as only bacteria present in the database can be identified (D'Argenio, 2018). Additionally, duodenum is an anatomic site with only few studies. Therefore, it is possible that many bacteria found in this place have not been properly characterized so far. Additionally, databases have an intrinsic error when assigning taxonomy. In the case of SILVA database error rate is of 15% (Edgar, 2018).

The small relative abundance of *H. pylori* described in the results is consistent with previous observations. When there is an infection with *H. pylori*, it tends to predominate in stomach mucosa. In these cases, the relative abundance can be between 40% to 95% (Pereira-Marques, Ferreira, Pinto-Ribeiro, & Figueiredo, 2019; Schulz et al., 2018). In duodenal samples (biopsies and aspirates) has also been reported the presence of *H. pylori*, but with relative abundances less than in the stomach. When comparing duodenal biopsies and aspirates, in the former the abundance is greater (Schulz et al., 2018).

H. pylori OTUs phylogeny and distribution in the study samples.

The current results suggest that a different *H. pylori* OTU was in every sample, except for HDRP0079 that had six different OTUs in the same sample. Furthermore, not all OTUs are in the same clade (figure 19 and 20). All this suggests the presence of different *Helicobacter* and a high diversity, which is consistent with previous reports. *H. pylori* is considered a bacterium with a panmictic population structure (Patra et al., 2012). This means high levels of recombination (Fernandez-Gonzalez & Backert, 2014; Patra et al., 2012) which cause high genetic diversity to the point that each clinical isolate can be genetically unique (Liu, Wei, Li, & Lei, 2011). Also, *H. pylori* has natural competence and the recombination is high compared with other bacteria. The presence of multiple strains in the same host is common (except in Europe and North America) and the distribution of the strains is different according to the biopsied place. Infections with multiple strains of *H. pylori* are important and favor horizontal gene transfer and adaptation (Patra et al., 2012). In our study, only one sample had more than one OTU (HDRP0079). However, it is possible that many OTUs were present in the stomach of all the patients studied but, it was not possible to isolate them in the duodenum. *Helicobacter* OTUs 18 and 19 had an identity less than 97% with the best matches in BLAST

analysis. This suggest that this OTUs could be other *Helicobacter* species (Chan, Halachev, Loman, Constantinidou, & Pallen, 2012). This possibility is also suggested by phylogenetic trees, which show that OTUs 18 and 19 are separated from the other OTUs.

Differential abundance

In the present study, it was possible to identify genera/species that were more abundant in samples positive for *H. pylori.* The taxa detected were *Haemophilus, Neisseria, Prevotella pallens, Prevotella 7, Streptococcus* and Bacteria (undefined phylum). In duodenal tissue, there is not previous evidence of changes in the relative abundance of the mentioned taxa when *H. pylori* positive and negative groups are compared (Schulz et al., 2018). On the other hand, there is evidence of changes in abundance in the stomach. *Haemophilus, Neisseria, Prevotella* and *Streptococcus* tend to be more abundant in the group without *H. pylori* infection compared with the group infected (Klymiuk et al., 2017; Schulz et al., 2018). Moreover, in the stomach a negative interaction (mutual exclusion) has been described between *H. pylori* and *Neisseria*, *Prevotella* and *Streptococcus* (Das et al., 2017). All taxa more abundant in the presence of *H. pylori* infection found in this study can be associated with pathology. *Haemophilus, Neisseria* and *Streptococcus* are nitrosating bacteria; as a result, can produce N-nitroso compounds (carcinogens) from nitrites and secondary amines (Sohn et al., 2017). Additionally, *Haemophilus* species (particularly *H. parainfluenzae*) have been associated with the accumulation of nitrites (Gantuya et al., 2019). In the following lines more details of the association between disease and differentially abundant taxa will be described.

Genus *Haemophilus* has been associated with Irritable bowel syndrome, eosinophilic esophagitis, celiac disease, gastric cancer, gastritis (in *H. pylori* negatives), peritonitis,

cholecystitis and liver abscess. In the case of latter five, *Haemophilus parainfluenzae* is probably the etiology (Frankard, Rodriguez-Villalobos, Struelens, & Jacobs, 2004; Gantuya et al., 2019; H. Han et al., 2019; Harris et al., 2015; Saulnier et al., 2011). It is possible that an ascending infection from the gastrointestinal tract (for example the duodenum) to the biliary tract is responsible for the hepatobiliary pathology (Frankard et al., 2004). If *H. pylori* increases *Haemophilus* in the duodenum, this could increase the risk of infections in the liver, gallbladder and biliary ducts. *Haemophilus* has the ability to penetrate the epithelial barrier, especially if it is inflamed. This contributes to maintain the inflammatory response (Harris et al., 2015). However, there are reports that genus *Haemophilus* tends to decrease in gastric cancer (Dias-jácome, Libânio, Borges-canha, Galaghar, & Pimentel-nunes, 2016).

An increase in the abundance of *Neisseria* has been associated with gastritis, duodenal ulcer and celiac disease. It was also isolated form antral ulcers (X. Chen et al., 2018; D'Argenio et al., 2016; Gantuya et al., 2019). Most studies have found that this genus is depleted in the presence of gastric adenocarcinoma or gastric neoplasia. However, Dias-jácome et al., Hu et al. and Gantuya et al. found a positive association (Dias-jácome et al., 2016; Gantuya et al., 2019; Hu, Pang, Huang, Zhang, & Zhang, 2018). The effects of *Neisseria* could depend on the bacterial species and the anatomic location. For example, the presence of *Neisseria enlongata* in the stomach was related with nausea, vomiting and lower abdominal pain. On contrast, *Neisseria perflava* in duodenum was considered a protective factor: reduces the risk of lower abdominal pain (H. Han et al., 2019).

The increase in abundance of *Prevotella* can produce a mucosal inflammatory response which is mediated by lymphocyte Th17. This is usually associated with metabolic syndrome and inflammatory bowel disease (Larsen, 2017). During the stages of gastric

carcinogenesis induced by *H. pylori*, it has been found that genus *Prevotella* increases its quantity. This effect is notorious in the passage from intestinal metaplasia to adenocarcinoma (Noto & Peek, 2017). Though, when an analysis at species level is made, it has been described that *Prevotella pallens* decreases in gastric cancer (Dias-jácome et al., 2016). Moreover, *P. pallens* has been defined as protective factor for heartburn and bloating when is present in the stomach and duodenum. In contrast, *P. pallens* has been related with *H. pylori* negative gastritis. The carcinogenic potential of *Prevotella* is not restricted to stomach, there are reports of the increase of this taxon in the fecal microbiome of patients with colorectal cancer (Gantuya et al., 2019; Weng et al., 2019). As a result, it is possible that the increase in *Prevotella* can contribute to the association between *H. pylori* infection and colon cancer (Teimoorian et al., 2018). *Prevotella* has been considered one of the dominant genus in duodenal ulcer of patients infected with *H. pylori* (X. Chen et al., 2018). In the case of *Prevotella 7,* its increase in stool samples was related with an increase in cardiovascular risk (Ascher & Reinhardt, 2018). Considering this finding, it is important to do future research that determines the role of microbiome in the *H. pylori* related cardiovascular disease.

Streptococcus like *Prevotella,* tends to increase in the transition from intestinal metaplasia to gastric adenocarcinoma (Noto & Peek, 2017). It is important to consider that *Streptococcus* associated gastric carcinogenesis could depend on the species. *S. mitis* and *S. sinensis* decrease in cancer while *S. anginosus* has a positive association (Dias-jácome et al., 2016; Gao et al., 2018). It has been described in vitro (not confirmed in vivo) a communication between *Streptococcus* and *H. pylori.* For example, when *Streptococcus mitis* and *H. pylori* are co-cultured, the first increases its culturability and the second changes into the coccoidal form (viable non-culturable) (Khosravi, Loke, Goh, & Vadivelu, 2016). These

effects could be related with a change in the expression of proteins. In *H. pylori* there is an increase in proteins associated with DNA synthesis and a decrease in metabolic enzymes, which has been previously described in the coccoid form (Khosravi et al., 2016). The coccoid form has been related with treatment failures, recurrences and epithelial cell proliferation which is associated with cancer (N. Li et al., 2013; Sarem & Corti, 2016). In the case of Streptococcal proteins, phosphoglycerate kinase increases, which favors its adhesion and proliferation (Khosravi et al., 2016). In contrast, one of the proteins that decreases its expression in the co-cultured *H. pylori* is thioredoxin. The latter may be associated with a lower risk of cancer (Khosravi et al., 2016). Other important association is that *Streptococcus* is more abundant in duodenal ulcers of patients with *H. pylori* infection in comparison with antral ulcers (X. Chen et al., 2018). Finally, in the gastrointestinal tract *Streptococcus* could be related with other diseases such as gastritis (*H. pylori* negative), colorectal cancer, and cirrhosis (Gantuya et al., 2019; Weng et al., 2019).

The only genus that increased its abundance in negative samples was *Ralstonia*. This genus is composed of gram negative, non-fermentative rods. They are encountered in many environments such as water (including potable water) and soil (Ryan & Adley, 2014). Therefore, its presence in the samples may also be related with contamination or with ingested food. In the group there are plant pathogens such as *R. solanareasum* and opportunistic human pathogens (*R. pickettii, R. insidiosa* and *R. mannitolilytica*) (Ryan & Adley, 2014; Yuliar, Nion, & Toyota, 2015). The diseases associated with these bacteria are bacteremia, septicemia, meningitis, septic arthritis, osteomyelitis and pneumonia (particularly in cystic fibrosis patients) (Prior, Gunaratnam, & Humphreys, 2017; Ryan & Adley, 2014; Ryan, Pembroke, & Adley, 2006). In relation to the interaction with *H. pylori*,

there is a study conducted in India where it was observed that these bacteria have a negative interaction in the stomach; in other words, they are mutually exclusive (Das et al., 2017). Intestinal *R. pickettii* is associated with obesity and glucose intolerance (Udayappan et al., 2017). Finally, *Ralstonia* has been encountered in the stomach of patient with gastric cancer in the absence of *H. pylori* (Weng et al., 2019).

The fact that *H. pylori* infection can potentially decrease the abundance of a bacterium associated with pathology (*Ralstonia*), shows us that it does not necessarily all interactions with the microbiome are harmful. Therefore, changes in the microbiome are complex and may be associated with multiple effects on human health.

The precise mechanisms by which *H. pylori* produces changes in the microbiome are not known yet. In the stomach possible explanations are the induction of hypochlorhydria and/or hypergastrinemia, chronic inflammation and the secretion of antimicrobial peptides by lymphoid tissue-inducing cells in response to *H. pylori* infection (Luyi Chen et al., 2018; Sgouras et al., 2015). However, *H. pylori* can affect gut microbiota in distal uninflamed regions of the gastrointestinal tract (Luyi Chen et al., 2018). Therefore, there must be other mechanisms. In an animal model made in Mongolian gerbils, the effect of the infection with *H. pylori* B8 (wildtype) and *ΔcagY* mutant (defective of type IV secretion system) was compared. The presence *H. pylori* B8 was associated with microbiota alterations in the distal uninflamed gastrointestinal tract (Heimesaat et al., 2014). *E. coli*, *Enterococcus*, Bacteroidetes, *Prevotella* and *Akkermansia* were more abundant in the large intestine of the animals infected with the wildtype compared with mutant strain.

Linear models with Gneiss

Linear models using Gneiss showed the small contribution that the presence of *H. pylori* had in the changes of the microbial community. As is known, there are many factors that modify the intestinal microbiota. For example, individual genetics, environment, diet, the use of medicines, diseases, gender and age (Wen & Duffy, 2017). Although the influence that *H. pylori* had in the duodenum seems to be small, it can be important from the point of view of human pathology. This is because it increases the abundance of bacteria that can be associated with disease.

Functional prediction with PICRUSt2 and DESeq2

With the use of PICRUSt2 and DESeq2 it was possible to identify genes that possibly expressed more in the positive group. Although it is a prediction, the information can be confirmed in future research.

Genes that code for proteins of the ComB system (ComB4, ComB8 and ComB9) were identified. These proteins are part comB-type 4 secretion system (ComB-T4SS) which permit the uptake of DNA and makes possible the natural transformation of *H. pylori.* In addition, in animal models performed in gerbils, it has been determined that ComB3 and ComB4 are involved in gastric colonization (Yuan, Wang, & Wang, 2018).

Components of restriction-modification (R-M) systems (restriction enzyme type II and DNA adenine methylase) had a higher expression in samples of the *H. pylori* positive group. Classically, these systems are a defense mechanism of bacterium against foreign DNA. However, they have other important functions. For example, recombination, stabilization of genomic islands, nutrition, regulation of conjugation, control of DNA replication, mismatch repair, gene regulation (housekeeping and virulence genes), etc. As a result, it could

contribute to the adaptation to the environment (Vasu & Nagaraja). In the case of *H. pylori* there are R-M systems that are related with virulence; for instance, the one constituted by iceA1 (type II restriction enzyme) and hpyIM (adenine-specific DNA methyltransferase). iceA1 is a virulence factor induced for contact with the epithelium and is associated with peptic ulcer disease (Joosten et al., 2015). Other bacteria that have many R-M systems, and belong to the genera with higher abundance in the positive group are *Streptococcus pneumoniae, Neisseria gonorrhoeae, Haemophilus influenzae* (Vasu & Nagaraja). Many components of the R-M systems of these bacteria have been linked with virulence: DpnA (*S. pneumoniae*), HI0209 (*H. influenzae*) and *Nme*SI system (*N. meningitidis*) (Bart, Pannekoek, Dankert, & van der Ende, 2001; Johnston, Martin, Granadel, Polard, & Claverys, 2013; Watson, Jarisch, & Smith, 2004).

There were components of toxin-antitoxin systems with differential expression. The toxin YafQ (part of the system YafQ-DinJ) has been described in detail in *E. coli*. This toxin causes a cleavage in mRNA which results in translation inhibition. In *E. coli,* toxin YafQ is involved in antibiotic resistance (cefazolin and tobramycin) within biofilms. Also in this bacterium, all YafQ-DinJ system is related with biofilm formation (Harrison et al., 2009; Kędzierska & Hayes, 2016). In *H. pylori* has been described proteins homologous to YafQ; for example, HP0894 (which is part of the HP0894-HP0895 system) and HP0892 (HP0892-HP0893 system). These systems could have a role in the survival of *H. pylori* in gastric mucosa. The anti-toxin genes could express in response to interactions with mucosa, which favors the inhibition of the toxin (K.-D. Han et al., 2013). HicB is an anti-toxin part of the HicBA system which is encountered in many prokaryotes. It could increase bacterial fitness because it helps to maintain genomic island (superintegrons) (Yeung, Cameron, Desjardins,

& Lee, 2011). In *Streptococcus pneumoniae,* HicB has additionally a function of transcriptional regulator of all the HicBA system. As a result, HicB not only blocks HicA (toxin), it also represses its production. In other words, HicB could have an important role in stretococcal growth by inhibiting HicA (Kim et al., 2018). In *Haemophilus ducreyi,* HicB contributes with ulcer formation (Yeung et al., 2011).

Genes of enzymes related with thiamine metabolism expressed more in positive group. Hydroxyethylthiazole kinase has been implicated as a possible virulence factor in *Haemophilus parasuis* and in *Streptococcus suis* serotype 2, by the use of suppression subtractive hybridization. However, these finding has not been confirmed (Jiang, Fan, & Lu, 2009; Yu et al., 2014). Thiamine diphosphokinase (thiamine pyrophosphokinase) is related with H_2O_2 in genus *Streptococcus*. This enzyme transfers pyrophosphate from ATP to thiamine. This results in the formation of thiamine pyrophosphate, which is a cofactor of pyruvate oxidase, an enzyme essential for H_2O_2 production (L. Chen et al., 2011). H_2O_2 is a virulence factor in *Streptococcus* which is associated with rupture of the double strand of DNA and apoptosis (Rai et al., 2015).

In general purine metabolism is essential for bacterial growth and survival. GMP reductase, an enzyme part of purine metabolism, increased its expression in the positive group. This enzyme is part of the salvage pathway of purine nucleotide biosynthesis in *H. pylori* which is indispensable for this bacterium, because it lacks many elements of de novo synthesis pathway (Liechti & Goldberg, 2012). Additionally, GMP-reductase is upregulated in response to acidic stress in *H. pylori* and in response to alkaline stress in other bacteria (*Enterococcus faecalis*) (Ran, Liu, Jiang, Sun, & Liang, 2015; Zanotti & Cendron, 2010). In conclusion, the gene could contribute to bacterial adaptation.

Flagellin of *H. pylori* and *Campylobacter jejuni* needs to be glycosylated with pseudaminic acid so that the flagella assembly can be done. UDP-4-amino-4,6-dideoxy-L-N-acetyl-beta-L-altrosamine transaminase (PseC) is an enzyme necessary for pseudaminic acid synthesis, and, as a result, controls an important virulence factor such as motility (Guerry et al., 2006; Ménard et al., 2014; Schoenhofen et al., 2006).

Sorbitol-6-phosphate 2-dehydrogenase is an enzyme that permit the use of sorbitol as a carbon source (Roux, Salmon, & Verchère-Béaur, 2006). However, in literature review it was not possible to associate its activity with gastrointestinal human disease or bacterial virulence.

Lipases, including phospholipases, are considered virulence factors in many bacteria (Istivan, 2006). Phospholipases A1 and A2 in *H. pylori* cause epithelial barrier disruption and favors gastric mucosa colonization. Also in *H. pylori,* outer membrane phospholipases contribute to control membrane lipid composition (change from the L-variant to the S-variant) which is an adaptation mechanism for acid conditions. S-variant have more affinity for epithelial cells and releases more urease and VacA (Istivan, 2006). In group A *Streptococcus*, phospholipase A2 increase bacterial attachment, induces inflammatory response and is cytotoxic (Brouwer, Barnett, Rivera-Hernandez, Rohde, & Walker, 2016; Sitkiewicz, Stockbauer, & Musser, 2007). Biofilm stabilization in *Neisseria meningitis* depends on the activity of phospholipases A (Lappann et al., 2010). Finally, *Prevotella* has phospholipases that produces tissue destruction (Alauzet, Marchandin, & Lozniewski, 2010).

Asparaginases are present in many gram positive and negative bacteria (Batool, Makky, Jalal, & Yusoff, 2016). In *H. pylori* inhibit cell cycle of host cells, are essential for colonization, produce cytotoxicity, suppress immune response (inhibit T cell proliferation)

and contribute to acid resistance (Miller & Maier, 2014; Rimbara, Mori, Kim, & Shibayama, 2013). Asparaginases have also been described as factors associated with virulence in *Haemophilus ducreyi*, *Neisseria meningitidis*, *Prevotella nigrescens*, *Mycobacterium tuberculosis* and *Salmonella typhimurium* (Bauer et al., 2008; Joseph et al., 2010; Ren et al., 2018; Szafrański et al., 2015).

Analysis of the mycobiome

Fungal taxa are also present in duodenal biopsies; however, it was no possible to do an analysis of diversity and differential abundance. The high number of OTUs assigned as Fungi (undefined phylum) is a limitation to make conclusions about the composition of the duodenal mycobiome. For the taxonomic assignment was used an updated database; but is possible that does not contain information of all taxa present in the duodenum. This because the lack of information of the duodenal mycobiome. According to previous studies, fungal diversity in the gastrointestinal tract is lower that the bacterial (Hillman, Lu, Yao, & Nakatsu, 2017). Nevertheless, the proportion fungi:bacteria is higher in the stomach and duodenum than in colon. In general in all gastrointestinal tract *Candida* (especially *C. albicans*), *Saccharomyces* (especially *S. cerevisiae*), *Penicillium*, *Aspergillus*, *Cryptococcus*, *Malassezia* (especially *M. restricta*), *Cladosporium*, *Galactomyces*, *Debaryomyces* and *Trichosporon* were the more frequent (Richard & Sokol, 2019). In the present study was possible to identify Ascomycota that contain the following genera: *Saccharomyces, Penicillium, Aspergillus, Cladosporium, Galactomyces* and *Debaryomyces*. We also identify Basidiomycota that contain genera *Cryptococcus*, *Malassezia* and *Trichosporon.* More information is need to determine the impact of *H. pylori* infection in fungal communities.

CONCLUSIONS

This is the first study looking at duodenal microbiome differences in Ecuadorians with and without *H. pylori* infection. Although the effect was small, it is important because the changes can be associated with disease. *H. pylori* infection modifies diversity (alpha and beta diversity, except evenness) and increases the abundance of bacteria that can be related with disease (*Haemophilus, Neisseria, Prevotella* and *Streptococcus*). On the other hand, it is possible that not all interactions with the microbiome are harmful. This is suggested with the higher abundance of *Ralstonia,* a potential pathogen, when *H. pylori* is not present. Finally, according to in silico predictions, in infected patients increases the expression of genes coding for products that can be liked to virulence and bacterial adaptation: ComB4, ComB8, ComB9, restriction enzyme type II, DNA adenine methylase, YafQ, HicB, hydroxyethylthiazole, thiamine diphosphokinase, GMP-reductase, phospholipases, asparaginase and UDP-4-amino-4,6-dideoxy-L-N-acetyl-beta-L-altrosamine transaminase. In ITS analysis was possible to find a small number of fungal taxa. The more important were: Recurvomyces, Ascomycota and *Xenoacremonium falcatum*. Further studies are needed to confirm the findings of the present work and to understand the relationships between *H. pylori* and other microorganisms in the duodenum.

REFERENCES

Alauzet, C., Marchandin, H., & Lozniewski, A. (2010). New insights into Prevotella diversity and medical microbiology. *Future Microbiology*, *5*(11), 1695–1718. https://doi.org/10.2217/fmb.10.126

Allaband, C., McDonald, D., Vázquez-Baeza, Y., Minich, J. J., Tripathi, A., Brenner, D. A., … Knight, R. (2019). Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians. *Clinical Gastroenterology and Hepatology*, *17*(2), 218–230. https://doi.org/10.1016/j.cgh.2018.09.017

Altschup, S. F., Gish, W., Pennsylvania, T., & Park, U. (1990). Basic Local Alignment Search Tool. *J Mol Biol.*, *215*(3), 403–410.

Ansari, S., & Yamaoka, Y. (2018). Current understanding and management of Helicobacter pylori infection: an updated appraisal. *F1000Research*, *7*, 721. https://doi.org/10.12688/f1000research.14149.1

Ascher, S., & Reinhardt, C. (2018). The gut microbiota: An emerging risk factor for cardiovascular and cerebrovascular disease. *European Journal of Immunology*, *48*(4), 564–575. https://doi.org/10.1002/eji.201646879

Bart, A., Pannekoek, Y., Dankert, J., & van der Ende, A. (2001). NmeSI Restriction-Modification System Identified by Representational Difference Analysis of a Hypervirulent Neisseria meningitidis Strain. *Infection and Immunity*, *69*(3), 1816–1820. https://doi.org/10.1128/IAI.69.3.1816-1820.2001

Batool, T., Makky, E. A., Jalal, M., & Yusoff, M. M. (2016). A Comprehensive Review on l-Asparaginase and Its Applications. *Applied Biochemistry and Biotechnology*, *178*(5), 900–923. https://doi.org/10.1007/s12010-015-1917-3

Bauer, M. E., Fortney, K. R., Harrison, A., Janowicz, D. M., Munson, R. S., & Spinola, S. M. (2008). Identification of Haemophilus ducreyi genes expressed during human infection. *Microbiology*, *154*(4), 1152–1160. https://doi.org/10.1099/mic.0.2007/013953-0

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). https://doi.org/10.1186/s13104-018-3565-5

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., … Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, *6*(1), 1–17. https://doi.org/10.1186/s40168-018-0470-z

Boylen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Ghalith, G. A. Al, … Naimey, A. T. (2018). QIIME 2 : Reproducible , interactive , scalable , and extensible microbiome data science. *PeerJ*. https://doi.org/10.7287/peerj.preprints.27295

Brouwer, S., Barnett, T. C., Rivera-Hernandez, T., Rohde, M., & Walker, M. J. (2016). Streptococcus pyogenes adhesion and colonization. *FEBS Letters*, *590*(21), 3739–3757. https://doi.org/10.1002/1873-3468.12254

Bühling, A., Radun, D., Müller, W. A., & Malfertheiner, P. (2001). Influence of anti-Helicobacter triple-therapy with metronidazole, omeprazole and clarithromycin on intestinal microflora. *Alimentary Pharmacology and Therapeutics*, *15*(9), 1445–1452. https://doi.org/10.1046/j.1365-2036.2001.01033.x

Buitrón, P. C. V. (2013). Prevalencia de infección por Helicobacter pylori y asociación con

patologías gástricas en pacientes adultos de chequeo ejecutivo desde enero del 2010 hasta septiembre del 2012 del Hospital Metropolitano de Quito- Ecuador. Retrieved from http://repositorio.usfq.edu.ec/bitstream/23000/1503/1/104865.pdf

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

Chan, J. Z., Halachev, M. R., Loman, N. J., Constantinidou, C., & Pallen, M. J. (2012). Defining bacterial species in the genomic era: insights from the genus Acinetobacter. *BMC Microbiology*, *12*(1). Retrieved from https://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-12-302

Chen, L., Ge, X., Dou, Y., Wang, X., Patel, J. R., & Xu, P. (2011). Identification of hydrogen peroxide production-related genes in Streptococcus sanguinis and their functional relationship with pyruvate oxidase. *Microbiology*, *157*(1), 13–20. https://doi.org/10.1099/mic.0.039669-0

Chen, Luyi, Xu, W., Lee, A., He, J., Huang, B., Zheng, W., … Chen, S. (2018). The impact of Helicobacter pylori infection, eradication therapy and probiotic supplementation on gut microenvironment homeostasis: An open-label, randomized clinical trial. *EBioMedicine*, *35*, 87–96. https://doi.org/10.1016/j.ebiom.2018.08.028

Chen, X., Xia, C., Li, Q., Jin, L., Zheng, L., & Wu, Z. (2018). Comparisons Between Bacterial Communities in Mucosa in Patients With Gastric Antrum Ulcer and a Duodenal Ulcer. *Frontiers in Cellular and Infection Microbiology*, *8*(May), 1–7. https://doi.org/10.3389/fcimb.2018.00126

Cox, M. J., Cookson, W. O. C. M., & Moffatt, M. F. (2013). *Sequencing the human*

microbiome in health and disease. *22*(1), 88–94. https://doi.org/10.1093/hmg/ddt398

D'Argenio, V. (2018). Human microbiome acquisition and bioinformatic challenges in metagenomic studies. *International Journal of Molecular Sciences*, *19*(2), 1–12. https://doi.org/10.3390/ijms19020383

D'Argenio, V., Casaburi, G., Precone, V., Pagliuca, C., Colicchio, R., Sarnataro, D., … Jabrì, B. (2016). Metagenomics Reveals Dysbiosis and a Potentially Pathogenic N . fl avescens Strain in Duodenum of Adult Celiac Patients. *Am J Gastroenterol*, *111*, 879–890. https://doi.org/10.1038/ajg.2016.95

Das, A., Pereira, V., Saxena, S., Ghosh, T. S., Anbumani, D., Bag, S., … Mande, S. S. (2017). Gastric microbiome of Indian patients with Helicobacter pylori infection, and their interaction networks. *Scientific Reports*, *7*(1), 15438. https://doi.org/10.1038/s41598-017-15510-6

Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., & Xia, J. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Research*, *45*(W1), W180–W188. https://doi.org/10.1093/nar/gkx295

Dias-jácome, E., Libânio, D., Borges-canha, M., Galaghar, A., & Pimentel-nunes, P. (2016). Gastric microbiota and carcinogenesis : the role of non- Helicobacter pylori bacteria - A systematic review. *Revista Española de Enfermedades Digestivas*, *108*(9), 530–540. https://doi.org/10.17235/reed.2016.4261/2016

Douglas, G. M., Maffei, V. J., Zaneveld, J., Yurgel, S. N., Brown, J. R., Taylor, C. M., … Langille, M. G. I. (2019). PICRUSt2: An improved and extensible approach for metagenome inference. *BioRxiv*, 672295.

Egorov, A. I., Sempértegui, F., Estrella, B., Egas, J., Naumova, E. N., & Griffiths, J. K. (2010). The effect of Helicobacter pylori infection on growth velocity in young children from poor urban communities in Ecuador. *International Journal of Infectious Diseases*, *14*(9), 788–791. https://doi.org/10.1016/j.ijid.2010.03.013

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10. https://doi.org/10.1890/0012-9658(2006)87[1465:ATTFHF]2.0.CO;2

Felsenstein, J. (1985). CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. *Evolution*, *39*(4), 783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Fernandez-Gonzalez, E., & Backert, S. (2014). DNA transfer in the gastric pathogen Helicobacter pylori. *Journal of Gastroenterology*, *49*(4), 594–604. https://doi.org/10.1007/s00535-014-0938-y

Frankard, J., Rodriguez-Villalobos, H., Struelens, M. J., & Jacobs, F. (2004). Haemophilus parainfluenzae : an Underdiagnosed Pathogen of Biliary Tract Infections ? *European Journal of Clinical Microbiology and Infectious Diseases*, *23*(1), 46–48. https://doi.org/10.1007/s10096-003-1050-z

Gantuya, B., El-Serag, H. B., Matsumoto, T., Ajami, N. J., Oyuntsetseg, K., Azzaya, D., … Yamaoka, Y. (2019). Gastric Microbiota in Helicobacter pylori-Negative and -Positive Gastritis Among High Incidence of Gastric Cancer Area. *Cancers*, *11*(4), 504. https://doi.org/10.3390/cancers11040504

Gao, J.-J., Zhang, Y., Gerhard, M., Mejias-Luque, R., Zhang, L., Li, Z.-X., … Quante, M. (2018). Association Between Gut Microbiota and Helicobacter pylori-Related Gastric Lesions

in a High-Risk Population of Gastric Cancer. *Frontiers in Cellular and Infection Microbiology*, *8*(June), 1–12. https://doi.org/10.3389/fcimb.2018.00202

Goodrich, J. K., Di Rienzi, S. C., Poole, A. C., Koren, O., Walters, W. A., Caporaso, J. G., … Ley, R. E. (2014). Conducting a Microbiome Study. *Cell*, *158*(2), 250–262. https://doi.org/10.1016/j.cell.2014.06.037

Gravina, A. G., Zagari, R. M., De Musis, C., Romano, L., Loguercio, C., & Romano, M. (2018). Helicobacter pylori and extragastric diseases: A review. *World Journal of Gastroenterology*, *24*(29), 3204–3221. https://doi.org/10.3748/wjg.v24.i29.3204

Guerry, P., Ewing, C. P., Schirm, M., Lorenzo, M., Kelly, J., Pattarini, D., … Logan, S. (2006). Changes in flagellin glycosylation affect Campylobacter autoagglutination and virulence. *Molecular Microbiology*, *60*(2), 299–311. https://doi.org/10.1111/j.1365-2958.2006.05100.x

Han, H., Lee, S.-Y., Oh, S., Moon, H., Cho, H., & Kim, J.-H. (2019). Correlations of the Gastric and Duodenal Microbiota with Histological, Endoscopic, and Symptomatic Gastritis. *Journal of Clinical Medicine*, *8*(3), 312. https://doi.org/10.3390/jcm8030312

Han, K.-D., Ahn, D.-H., Lee, S.-A., Min, Y.-H., Kwon, A.-R., Ahn, H.-C., & Lee, B.-J. (2013). Identification of Chromosomal HP0892-HP0893 Toxin-Antitoxin Proteins in Helicobacter pylori and Structural Elucidation of Their Protein-Protein Interaction. *Journal of Biological Chemistry*, *288*(8), 6004–6013. https://doi.org/10.1074/jbc.M111.322784

Harris, J. K., Fang, R., Wagner, B. D., Choe, H. N., Kelly, C. J., Schroeder, S., … Fillon, S. A. (2015). *Esophageal Microbiome in Eosinophilic Esophagitis*. 1–13. https://doi.org/10.1371/journal.pone.0128346

Harrison, J. J., Wade, W. D., Akierman, S., Vacchi-Suzzi, C., Stremick, C. A., Turner, R. J., & Ceri, H. (2009). The Chromosomal Toxin Gene yafQ Is a Determinant of Multidrug Tolerance for Escherichia coli Growing in a Biofilm. *Antimicrobial Agents and Chemotherapy*, *53*(6), 2253–2258. https://doi.org/10.1128/AAC.00043-09

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174. Retrieved from https://link.springer.com/article/10.1007/BF02101694

Heimesaat, M. M., Fischer, A., Plickert, R., Wiedemann, T., Loddenkemper, C., Göbel, U. B., … Rieder, G. (2014). Helicobacter pylori induced gastric immunopathology is associated with distinct microbiota changes in the large intestines of long-term infected Mongolian gerbils. *PLoS ONE*, *9*(6), 1–11. https://doi.org/10.1371/journal.pone.0100362

Hillman, E. T., Lu, H., Yao, T., & Nakatsu, C. H. (2017). Microbial Ecology along the Gastrointestinal Tract. *Microbes and Environments*, *32*(4), 300–313. https://doi.org/10.1264/jsme2.ME17017

Hiratsuka, T., Furihata, K., Ishikawa, J., Yamashita, H., Itoh, N., Seto, H., & Dairi, T. (2008). An Alternative Menaquinone Biosynthetic Pathway Operating in Microorganisms. *Science*, *321*(5896), 1670–1673. https://doi.org/10.1126/science.1160446

Hu, Y., Pang, W., Huang, Y., Zhang, Y., & Zhang, C. (2018). The Gastric Microbiome Is Perturbed in Advanced Gastric Adenocarcinoma Identified Through Shotgun Metagenomics. *Frontiers in Cellular and Infection Microbiology*, *8*(December), 1–11. https://doi.org/10.3389/fcimb.2018.00433

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science &*

Engineering, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Istivan, T. S. (2006). Phospholipase A in Gram-negative bacteria and its role in pathogenesis. *Microbiology*, *152*(5), 1263–1274. https://doi.org/10.1099/mic.0.28609-0

Jiang, H., Fan, H., & Lu, C. (2009). Identification and distribution of putative virulent genes in strains of Streptococcus suis serotype 2. *Veterinary Microbiology*, *133*(4), 309–316. https://doi.org/10.1016/j.vetmic.2008.07.014

Johnston, C., Martin, B., Granadel, C., Polard, P., & Claverys, J.-P. (2013). Programmed Protection of Foreign DNA from Restriction Allows Pathogenicity Island Exchange during Pneumococcal Transformation. *PLoS Pathogens*, *9*(2), e1003178. https://doi.org/10.1371/journal.ppat.1003178

Joosten, M., Lindén, S., Rossi, M., Tay, A. C. Y., Skoog, E., Padra, M., … Smet, A. (2015). Divergence between the Highly Virulent Zoonotic Pathogen Helicobacter heilmannii and Its Closest Relative, the Low-Virulence "Helicobacter ailurogastricus" sp. nov. *Infection and Immunity*, *84*(1), 293–306. https://doi.org/10.1128/IAI.01300-15

Jorgensen, J. H., Pfaller, M. A., Carroll, K. C., Landry, M. L., Funke, G., Richter, S. S., & Warnock, D. W. (2015). *Manual of Clinical Microbiology* (11th ed.). ASM Press.

Joseph, B., Schneiker-Bekel, S., Schramm-Gluck, A., Blom, J., Claus, H., Linke, B., … Schoen, C. (2010). Comparative Genome Biology of a Serogroup B Carriage and Disease Strain Supports a Polygenic Nature of Meningococcal Virulence. *Journal of Bacteriology*, *192*(20), 5363–5377. https://doi.org/10.1128/JB.00883-10

Kao, C. Y., Sheu, B. S., & Wu, J. J. (2016). Helicobacter pylori infection: An overview of bacterial virulence factors and pathogenesis. *Biomedical Journal*, *39*(1), 14–23.

https://doi.org/10.1016/j.bj.2015.06.002

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12136088%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC135756

Kędzierska, B., & Hayes, F. (2016). Emerging Roles of Toxin-Antitoxin Modules in Bacterial Pathogenesis. *Molecules*, *21*(6), 790. https://doi.org/10.3390/molecules21060790

Khosravi, Y., Loke, M. F., Goh, K. L., & Vadivelu, J. (2016). Proteomics Analysis Revealed that Crosstalk between Helicobacter pylori and Streptococcus mitis May Enhance Bacterial Survival and Reduces Carcinogenesis. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01462

Kim, D.-H., Kang, S.-M., Park, S. J., Jin, C., Yoon, H.-J., & Lee, B.-J. (2018). Functional insights into the Streptococcus pneumoniae HicBA toxin–antitoxin system based on a structural study. *Nucleic Acids Research*, *46*(12), 6371–6386. https://doi.org/10.1093/nar/gky469

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

Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., & Knight, R. (2012). Experimental and analytical tools for studying the human microbiome. *Nature Reviews Genetics*, *13*(1), 47–58. https://doi.org/10.1038/nrg3129

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lane, D. J. (1991). 16S/23S rRNA Sequencing. In E. Stackebrandt & M. Goodfellow (Eds.), *Nucleic Acid Techniques in Bacterial Systematic (Modern Microbiological Methods)* (pp. 115–175). New York: John Wiley & Sons.

Lappann, M., Claus, H., van Alen, T., Harmsen, M., Elias, J., Molin, S., & Vogel, U. (2010). A dual role of extracellular DNA during biofilm formation of Neisseria meningitidis. *Molecular Microbiology*, *75*(6), 1355–1371. https://doi.org/10.1111/j.1365-2958.2010.07054.x

Larsen, J. M. (2017). The immune response to Prevotella bacteria in chronic inflammatory disease. *Immunology*, *151*(4), 363–374. https://doi.org/10.1111/imm.12760

Li, G., Yang, M., Zhou, K., Zhang, L., Tian, L., Lv, S., … Hou, X. (2015). Diversity of duodenal and rectal microbiota in biopsy tissues and luminal contents in healthy volunteers. *Journal of Microbiology and Biotechnology*, *25*(7), 1136–1145. https://doi.org/10.4014/jmb.1412.12047

Li, N., Han, L., Chen, J., Lin, X., Chen, H., & She, F. (2013). Proliferative and apoptotic effects of gastric epithelial cells induced by coccoid Helicobacter pylori. *Journal of Basic Microbiology*, *53*(2), 147–155. https://doi.org/10.1002/jobm.201100370

Liechti, G., & Goldberg, J. B. (2012). Helicobacter pylori Relies Primarily on the Purine Salvage Pathway for Purine Nucleotide Biosynthesis. *Journal of Bacteriology*, *194*(4), 839–854. https://doi.org/10.1128/JB.05757-11

Liu, L. A., Wei, D., Li, Y., & Lei, H. (Eds.). (2011). *Handbook of Research on Computational*

and Systems Biology. https://doi.org/10.4018/978-1-60960-491-2

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

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

Maldonado-Contreras, A., Goldfarb, K. C., Godoy-Vitorino, F., Karaoz, U., Contreras, M., Blaser, M. J., … Dominguez-Bello, M. G. (2011). Structure of the human gastric bacterial community in relation to Helicobacter pylori status. *ISME Journal*, *5*(4), 574–579. https://doi.org/10.1038/ismej.2010.149

Mandal, S., Van Treuren, W., Eggesbø, M., Knight, R., Peddada, S. D., & White, R. A. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health & Disease*, *1*, 1–7. https://doi.org/10.3402/mehd.v26.27663

Matlock, B. (2015). *Assessment of Nucleic Acid Purity*. Retrieved from https://assets.thermofisher.com/TFS-Assets/CAD/Product-Bulletins/TN52646-E-0215M-NucleicAcid.pdf

Ménard, R., Schoenhofen, I. C., Tao, L., Aubry, A., Bouchard, P., Reid, C. W., … Logan, S. M. (2014). Small-Molecule Inhibitors of the Pseudaminic Acid Biosynthetic Pathway: Targeting Motility as a Key Bacterial Virulence Factor. *Antimicrobial Agents and Chemotherapy*, *58*(12), 7430–7440. https://doi.org/10.1128/AAC.03858-14

Miller, E. F., & Maier, R. J. (2014). Ammonium Metabolism Enzymes Aid Helicobacter pylori Acid Resistance. *Journal of Bacteriology*, *196*(17), 3074–3081. https://doi.org/10.1128/JB.01423-13

Morgan, X. C., & Huttenhower, C. (2012). Chapter 12: Human Microbiome Analysis. *PLoS Computational Biology*, *8*(12). https://doi.org/10.1371/journal.pcbi.1002808

Morton, J. T., Sanders, J., Quinn, R. A., Mcdonald, D., Gonzalez, A., Vázquez-baeza, Y., & Navas-molina, J. A. (2017). *Balance Trees Reveal Microbial Niche Differentiation*. *2*(1), 1–11.

Noto, J. M., & Peek, R. M. (2017). The gastric microbiome, its interaction with Helicobacter pylori, and its potential role in the progression to stomach cancer. *PLOS Pathogens*, *13*(10), e1006573.

Patra, R., Chattopadhyay, S., De, R., Ghosh, P., Ganguly, M., Chowdhury, A., … Mukhopadhyay, A. K. (2012). Multiple Infection and Microdiversity among Helicobacter pylori Isolates in a Single Host in India. *PLoS ONE*, *7*(8), e43370. https://doi.org/10.1371/journal.pone.0043370

Pereira-Marques, J., Ferreira, R. M., Pinto-Ribeiro, I., & Figueiredo, C. (2019). Helicobacter pylori Infection, the Gastric Microbiome and Gastric Cancer. In *Advances in Experimental Medicine and Biology*. https://doi.org/10.1007/5584_2019_366

Portal-Celhay, C., & Perez-Perez, G. I. (2006). Immune responses to *Helicobacter pylori* colonization: mechanisms and clinical outcomes. *Clinical Science*, *110*(3), 305–314. https://doi.org/10.1042/CS20050232

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, *5*(3).

https://doi.org/10.1371/journal.pone.0009490

Prior, A., Gunaratnam, C., & Humphreys, H. (2017). Ralstonia species - do these bacteria matter in cystic fibrosis? *Paediatr Respir Rev*, *23*, 78–83.

Qiime2. (2019). Differential abundance analysis with gneiss. Retrieved from https://docs.qiime2.org/2019.7/tutorials/gneiss/

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

Rai, P., Parrish, M., Tay, I. J. J., Li, N., Ackerman, S., He, F., … Engelward, B. P. (2015). Streptococcus pneumoniae secretes hydrogen peroxide leading to DNA damage and apoptosis in lung cells. *Proceedings of the National Academy of Sciences*, *112*(26), E3421–E3430. https://doi.org/10.1073/pnas.1424144112

Ran, S., Liu, B., Jiang, W., Sun, Z., & Liang, J. (2015). Transcriptome analysis of Enterococcus faecalis in response to alkaline stress. *Frontiers in Microbiology*, *6*, 795. https://doi.org/10.3389/fmicb.2015.00795

Ren, W., Rajendran, R., Zhao, Y., Tan, B., Wu, G., Bazer, F. W., … Yin, Y. (2018). Amino Acids As Mediators of Metabolic Cross Talk between Host and Pathogen. *Frontiers in Immunology*, *9*. https://doi.org/10.3389/fimmu.2018.00319

Richard, M. L., & Sokol, H. (2019). The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases. *Nature Reviews Gastroenterology & Hepatology*, *16*(6), 331–345. https://doi.org/10.1038/s41575-019-0121-2

Rideout, J. R., Chase, J. H., Bolyen, E., Ackermann, G., González, A., Knight, R., & Caporaso,

G. G. (2016). Keemei: Cloud-based validation of tabular bioinformatics file formats in Google Sheets. *GigaScience*, *5*(1). https://doi.org/10.1186/s13742-016-0133-6

Rimbara, E., Mori, S., Kim, H., & Shibayama, K. (2013). Role of γ-glutamyltranspeptidase in the pathogenesis of Helicobacter pylori infection. *Microbiology and Immunology*, *57*(10), 665–673. https://doi.org/10.1111/1348-0421.12089

Roux, C., Salmon, L., & Verchère-Béaur, C. (2006). Preliminary studies on the inhibition of D-sorbitol-6-phosphate 2–dehydrogenase from Escherichia coli with substrate analogues. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *21*(2), 187–192. https://doi.org/10.1080/14756360500535260

Ryan, M. P., & Adley, C. C. (2014). Ralstonia spp.: emerging global opportunistic pathogens. *European Journal of Clinical Microbiology & Infectious Diseases*, *33*(3), 291–304. https://doi.org/10.1007/s10096-013-1975-9

Ryan, M. P., Pembroke, J. T., & Adley, C. C. (2006). Ralstonia pickettii: a persistent Gram-negative nosocomial infectious organism. *Journal of Hospital Infection*, *62*(3), 278–284.

Sarem, M., & Corti, R. (2016). Role of Helicobacter pylori coccoid forms in infection and recrudescence. *Gastroenterología y Hepatología (English Edition)*, *39*(1), 28–35. https://doi.org/10.1016/j.gastre.2015.12.005

Sasaki, T., Hirai, I., Izurieta, R., Kwa, B. H., Estevez, E., Saldana, A., … Yamamoto, Y. (2009). Analysis of *Helicobacter pylori* Genotype in Stool Specimens of Asymptomatic People. *Laboratory Medicine*, *40*(7), 412–414. https://doi.org/10.1309/LMZ2WWCD2A9MFTNW

Saulnier, D. M., Riehle, K., Mistretta, T. A., Diaz, M. A., Mandal, D., Raza, S., … Versalovic, J.

(2011). Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome. *Gastroenterology*, *141*(5), 1782–1791. https://doi.org/10.1053/j.gastro.2011.06.072

Schoenhofen, I. C., Lunin, V. V., Julien, J.-P., Li, Y., Ajamian, E., Matte, A., … Young, N. M. (2006). Structural and Functional Characterization of PseC, an Aminotransferase Involved in the Biosynthesis of Pseudaminic Acid, an Essential Flagellar Modification in Helicobacter pylori. *Journal of Biological Chemistry*, *281*(13), 8907–8916. https://doi.org/10.1074/jbc.M512987200

Schulz, C., Schütte, K., Koch, N., Vilchez-Vargas, R., Wos-Oxley, M. L., Oxley, A. P. A., … Pieper, D. H. (2018). The active bacterial assemblages of the upper Gi tract in individuals with and without Helicobacter infection. *Gut*, *67*(2), 216–225. https://doi.org/10.1136/gutjnl-2016-312904

Sgouras, D. N., Thi, T., Trang, H., & Yamaoka, Y. (2015). *Pathogenesis of Helicobacter pylori Infection*. *20*, 8–16. https://doi.org/10.1111/hel.12251

Shiotani, A., Matsumoto, H., Fukushima, S., Katsumata, R., Kawano, M., & Saito, M. (2018). *H . pylori and Human Gut Microbiota*. *5*(8), 1–5.

Sitkiewicz, I., Stockbauer, K. E., & Musser, J. M. (2007). Secreted bacterial phospholipase A2 enzymes: better living through phospholipolysis. *Trends in Microbiology*, *15*(2), 63–69. https://doi.org/10.1016/j.tim.2006.12.003

Sohn, S.-H., Kim, N., Jo, H. J., Kim, J., Park, J. H., Nam, R. H., … Lee, D. H. (2017). Analysis of Gastric Body Microbiota by Pyrosequencing: Possible Role of Bacteria Other Than Helicobacter pylori in the Gastric Carcinogenesis. *Journal of Cancer Prevention*, *22*(2), 115–125. https://doi.org/10.15430/JCP.2017.22.2.115

Szafrański, S. P., Deng, Z.-L., Tomasch, J., Jarek, M., Bhuju, S., Meisinger, C., … Wagner-Döbler, I. (2015). Functional biomarkers for chronic periodontitis and insights into the roles of Prevotella nigrescens and Fusobacterium nucleatum; a metatranscriptome analysis. *Npj Biofilms and Microbiomes*, *1*(1), 15017. https://doi.org/10.1038/npjbiofilms.2015.17

Teimoorian, F., Ranaei, M., Hajian Tilaki, K., Shokri Shirvani, J., & Vosough, Z. (2018). Association of Helicobacter pylori Infection With Colon Cancer and Adenomatous Polyps. *Iranian Journal of Pathology*, *13*(3), 325–332. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/30636955

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Turner, S., Pryer, K. M., Miao, V. P. W., & Palmer, J. D. (1999). Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis. *The Journal of Eukaryotic Microbiology*, *46*(4), 327–338. https://doi.org/10.1111/j.1550-7408.1999.tb04612.x

Udayappan, S. D., Kovatcheva-Datchary, P., Bakker, G. J., Havik, S. R., Herrema, H., Cani, P. D., … Nieuwdorp, M. (2017). Intestinal Ralstonia pickettii augments glucose intolerance in obesity. *PLOS ONE*, *12*(11), e0181693. https://doi.org/10.1371/journal.pone.0181693

UNITE Community. (2019). *UNITE QIIME release for Fungi 2. Version 18.11.2018.*. Retrieved from https://doi.org/10.15156/BIO/786349

Vasu, K., & Nagaraja, V. Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiology and Molecular Biology Reviews*, *77*(1), 53–72. https://doi.org/10.1128/MMBR.00044-12

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., & Knight, R. (2013). EMPeror: A tool for visualizing high-throughput microbial community data. *GigaScience*, *2*(1), 2–5. https://doi.org/10.1186/2047-217X-2-16

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Watson, M. E., Jarisch, J., & Smith, A. L. (2004). Inactivation of deoxyadenosine methyltransferase (dam) attenuates Haemophilus influenzae virulence. *Molecular Microbiology*, *53*(2), 651–664. https://doi.org/10.1111/j.1365-2958.2004.04140.x

Wen, L., & Duffy, A. (2017). Factors Influencing the Gut Microbiota , Inflammation , and Type 2 Diabetes. *The Journal of Nutrition*, *147*(7), 1468S-1475S. https://doi.org/10.3945/jn.116.240754

Weng, M.-T., Chiu, Y.-T., Wei, P.-Y., Chiang, C.-W., Fang, H.-L., & Wei, S.-C. (2019). Microbiota and gastrointestinal cancer. *Journal of the Formosan Medical Association*, *118*, S32–S41. https://doi.org/10.1016/j.jfma.2019.01.002

Yang, B., Wang, Y., & Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, *17*(1), 135. https://doi.org/10.1186/s12859-016-0992-y

Yeung, A., Cameron, D. W., Desjardins, M., & Lee, B. C. (2011). Use of signature-tagged

mutagenesis to identify virulence determinants in Haemophilus ducreyi responsible for ulcer formation. *Journal of Microbiological Methods*, *84*(2), 290–298. https://doi.org/10.1016/j.mimet.2010.12.022

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., … Glöckner, F. O. (2014). The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, *42*(D1), 643–648. https://doi.org/10.1093/nar/gkt1209

Yu, J., Wu, J., Zhang, Y., Du, Y., Peng, J., Chen, L., … Wang, J. (2014). Identification of putative virulence-associated genes among Haemophilus parasuis strains and the virulence difference of different serovars. *Microbial Pathogenesis*, *77*, 17–23. https://doi.org/10.1016/j.micpath.2014.10.001

Yuan, X.-Y., Wang, Y., & Wang, M.-Y. (2018). The type IV secretion system in Helicobacter pylori. *Future Microbiology*, *13*(9), 1041–1054. https://doi.org/10.2217/fmb-2018-0038

Yuliar, Nion, Y. A., & Toyota, K. (2015). Recent Trends in Control Methods for Bacterial Wilt Diseases Caused by Ralstonia solanacearum. *Microbes and Environments*, *30*(1), 1–11. https://doi.org/10.1264/jsme2.ME14144

Zamani, M., Ebrahimtabar, F., Zamani, V., Miller, W. H., Alizadeh-Navaei, R., Shokri-Shirvani, J., & Derakhshan, M. H. (2018). Systematic review with meta-analysis: the worldwide prevalence of Helicobacter pylori infection. *Alimentary Pharmacology and Therapeutics*, *47*(7), 868–876. https://doi.org/10.1111/apt.14561

Zanotti, G., & Cendron, L. (2010). Functional and structural aspects of helicobacter pylori acidic stress response factors. *IUBMB Life*, *62*(10), 715–723. https://doi.org/10.1002/iub.382