The active bacterial assemblages of the upper gastrointestinal tract in individuals with and without *Helicobacter* infection

Short title: bacterial assemblages of the human intestine

Christian Schulz[1,2], Kerstin Schütte[1♯], Nadine Koch[2], Ramiro Vilchez-Vargas[1], Melissa L. Wos-Oxley[2], Andrew P. A. Oxley[2&], Marius Vital[2], Peter Malfertheiner[1] and Dietmar H. Pieper[2*].

[1]Otto-von-Guericke-University Magdeburg, Department of Gastroenterology, Hepatology and Infectious Diseases, Magdeburg, Germany.
[2]Helmholtz Centre for Infection Research, Microbial Interactions and Processes (MINP) Research Group, Braunschweig, Germany.

[♯]Current address: Department of Internal Medicine and Gastroenterology, Niels-Stensen-Kliniken, Marienhospital, Osnabrück, Germany.
[&]Current address: Molecular Sciences Laboratory, SARDI Aquatic Sciences, 2 Hamra Ave, West Beach, South Australia, Australia

*Corresponding author: Dietmar H. Pieper. Email: dpi@helmholtz-hzi.de. Postal address: Helmholtz-Zentrum für Infektionsforschung GmbH. Inhoffenstraße 7. 38124 Braunschweig. Germany. Telephone number: 49-531-61814200

word count: 3980

Keywords: *Helicobacter pylori*; gastric microbiome; human gastrointestinal tract; high-throughput 16S rRNA sequencing.

Abbreviations: ANOSIM, Analysis of Similarity; DB, duodenum biopsy; DA, duodenum aspirate; GI tract, gastrointestinal tract; H. pylori, Helicobacter pylori; nMDS, non-metric multidimensional scaling; OA, Oral aspirate (= saliva); PCR, polymerase chain reaction; PERMANOVA, Permutational Multivariate Analysis of Variance; rDNA, ribosomal deoxyribonucleic acid; rRNA, ribosomal ribonucleic acid; SB, stomach biopsy; SA, stomach aspirate; SD, standard deviation; TD, taxonomic distinctness.

Abstract

Objective

Patients infected with *Helicobacter pylori* develop chronic gastritis with a subgroup progressing to further complications. The role of microbiota from the oral cavity swallowed with saliva and either transiting the stomach or persisting in the gastric mucosa is uncertain. It is also not known whether the bacterial community differs in luminal and mucosal niches. A key question is whether *H. pylori* influences the bacterial communities of gastroduodenal niches.

Design

Saliva, gastric and duodenal aspirates as well as gastric and duodenal biopsies were collected during esophagogastroduodenoscopy from 24 patients (m:9, f:15, mean age 52.2 ± SD 14.5 years). RNA was extracted and the V1-V2 region of the retrotranscribed bacterial 16S rRNA amplified and sequenced on the Illumina MiSeq platform.

Results

Overall, 687 bacterial phylotypes that belonged to 95 genera and 11 phyla were observed. Each individual comprised a unique microbiota composition that was consistent across the different niches. However, the stomach fluid enriched for specific microbiota components. *Helicobacter* spp. were shown not only to dominate the mucosa-associated community in the stomach, but to also significantly influence duodenal and oral communities.

Conclusions

The detailed analysis of the active global bacterial communities from the five distinct sites of the upper gastrointestinal tract allowed for the first time the differentiation between host effects and the influence of sampling region on the bacterial community. The influence of *Helicobacter* spp. on the global community structures is striking.

Keywords

Helicobacter pylori; gastric microbiome; human gastrointestinal tract; high-throughput 16S rRNA sequencing.

Summary box

What is already known about this subject?

- The upper gastrointestinal tract is inhabited by a complex microbiota
- The oral cavity may be the source of the gastric microbiota
- The presence of *H. pylori* diminishes species richness and may further influence bacterial community structure

What are the new findings?

- Using reverse transcribed 16S rRNA as the amplification template, precise insights into the metabolically active bacteria of the upper gastrointestinal tract were obtained
- Each subject has their own individual microbiota composition that is consistent throughout the regions of the upper gastrointestinal tract
- The stomach fluid enriches for specific microbiota components
- The presence of *H. pylori* influences the community composition of the duodenum and oral cavity

How might it impact on clinical practice in the foreseeable future?

- The bacterial community of the oral cavity may play a crucial role in determining susceptibility to *H.pylori* infection and must be considered in further analyses
- Stomach aspirate microbiota are very distinct from biopsies and do not provide a reliable depiction of the mucosal associated bacteria
- Changes of the microbiota in the duodenum due to *H. pylori* infection may influence duodenal diseases

INTRODUCTION

Until the discovery of *Helicobacter pylori*, the human stomach was considered to be a sterile site due to its highly acidic nature. Following this discovery, it was assumed that *H. pylori* was the only bacterium able to colonize the gastric epithelium[1]. However, in the proceeding years, members of the phyla Firmicutes, Proteobacteria, and Bacteroidetes were cultured from gastric juice[2] with more recent culture-independent work revealing a complex bacterial community inhabiting the upper gastrointestinal tract[3].

H. pylori is usually acquired during childhood[4, 5], is persistent throughout life and associated with various degrees of gastric mucosal inflammation[4]. Infected individuals develop chronic gastritis and a subgroup progresses to further complications such as peptic ulcer disease or cancer[6]. *H. pylori* gastritis leads to changes in acid secretion of varying degrees depending on the topographic and phenotypic expression of gastritis[6, 7].

Recent studies have indicated saliva as a source for bacteria present in stomach aspirates[8,9]. However, there is a lack of knowledge on whether these bacteria are able to colonize the gastric mucosa. The gastric mucosa is covered by a viscous mucus layer with a gradient between the physiologic acidic luminal and the neutral epithelial pH. *H. pylori* utilizes this pH gradient to colonize its ecological niche on the gastric epithelium. This difference in pH is assumed to result in differences in the transient luminal and the mucosa-associated microbiota where recent analyses using limited numbers of individuals have shown differences in the phylogenetic composition of bacterial communities[10]. There is also limited knowledge on the biodiversity in the gastric mucosa of individuals infected with *H. pylori* compared to non-infected individuals even though it has been suggested that the carcinogenic role of *H. pylori* is partly due to its impact on the gastric commensal microbiota[5, 11]. However, *H. pylori* obviously dominates in the stomach mucosa of infected individuals while uninfected individuals show higher biodiversity[12, 13]. Changes of the microbiota in the duodenum and the proximal small bowel due to *H. pylori* infection have yet to be studied despite the causal relationship between a gastric infection with *H. pylori* and duodenal ulcer disease[14, 15].

In this study we aimed to characterize non-*H. pylori* bacteria, to identify whether they are transients or residents of the intestinal tract and to assess to which extent *H. pylori* affects the biodiversity of the whole human upper gastrointestinal (GI) tract. Thus, the microbiota of saliva, stomach aspirates, stomach biopsies, duodenum aspirates and duodenum biopsies of 24 individuals were analyzed by high-throughput sequencing. Most sequence-based studies of bacterial communities have used libraries constructed by amplifying DNA extracted from samples, which include DNA from active or viable cells, but also DNA from dead lysed or degraded cells. Due to the short lifespan of extracellular RNA compared with DNA, sequence libraries constructed from reverse transcripts of RNA describe the active component of a bacterial community[16, 17]. Thus, the biodiversity of GI samples was analyzed based on the RNA profiles.

MATERIALS AND METHODS

Study cohort

Individuals undergoing scheduled esophagogastroduodenoscopy at the University of Magdeburg were enrolled between October 2014 and March 2015. The protocol was performed in accordance with current good clinical practice guidelines, the declaration of Helsinki, and was approved by the local ethic committee (operation number 31/11). All individuals gave their written informed consent. The cohort of 24 individuals with histopathological diagnosis of chronic gastritis comprised 9 males and 15 females with a mean age of 52±14 years (Supplementary Table 1). Individuals had not received antibiotic treatment for at least four weeks prior to sampling. From each patient, samples from saliva as well as both stomach and duodenal aspirates and biopsies were obtained. *H. pylori* status was identified based on the rapid urease test and histopathological assessment (see Supplementary Materials). In addition, serum samples were analyzed to detect anti-H. pylori IgG (see Supplementary Materials).

RNA extraction, cDNA synthesis and library preparation

RNA was extracted using the RNeasy kit (Qiagen) following manufacturer's instructions, but including a mechanical lysis step as detailed in the Supplementary Materials. After DNA digestion, first-strand cDNA was synthesized using SuperScript® III First-Strand Synthesis System (Invitrogen™, Carlsbad, California, USA) and random primers, following manufacturer´s instructions. Amplicon libraries were generated as previously described[18] where the V1-V2 region of the 16S rRNA was amplified after 20 cycles PCR reaction using the 27F and 338R primers[19] and sequenced on a MiSeq (2x250 bp, Illumina, Hayward, CA, USA).

Bioinformatic and statistical analysis

Bioinformatic processing was performed as previously described[18]. Raw reads were merged with the RDP assembler[20]. Overall 3,677,068 paired-ends reads were obtained with a mean of 30,900±16,467 reads per sample. Sequences were aligned within MOTHUR (gotoh algorithm using the SILVA reference database) and subjected to pre-clustering (diffs=2)[21] yielding so-called phylotypes that were filtered for an average abundance of ≥0.001% and a sequence length ≥250bp before analysis. All samples were re-sampled to equal the smallest library size of 10,217 reads using the phyloseq package[22] returning 687 phylotypes (Supplementary Table 2). Phylotypes were assigned to a taxonomic affiliation based on the naïve Bayesian classification[23] with a pseudo-bootstrap threshold of 80%. Phylotypes were then manually analyzed against the RDP database using the Seqmatch function as well as against the NCBI database to define the discriminatory power of each sequence read. A species name was assigned to a phylotype when only 16S rRNA gene fragments of previously described isolates of that species showed ≤2 mismatches with the respective representative sequence read[18]. Similarly, a genus name was assigned to a phylotype when only 16S rRNA gene fragments of previously described isolates belonging to that genus and of 16S rRNA gene fragments originating from uncultured representatives of that genus

showed ≤2 mismatches. Relative abundances (in percentage) of phylotypes, genera and phyla were used for downstream analyses (Supplementary Table 3-5). The vegan package from R[24] was used to generate rarefaction curves as well as Shannon´s diversity and Pielou's evenness indices, while multivariate analyses were performed using PRIMER (v.7.0.11, PRIMER-E, Plymouth Marine Laboratory, UK), and univariate analyses using Prism 7 (Graphpad Software, Inc.).

Taxonomic diversity was calculated using algorithms for taxonomic distinctness (TD): average taxonomic distinctness (delta+) and variation in taxonomic distinctness (lambda+). Delta+ represents the average taxonomic distance between all pairs of species within each sample and thus is a summary of average taxonomic breadth of each sample, while lambda+ reports how consistent each level of organization within the Linnaean classification is represented[25, 26].

The data matrices comprising either the 687 phylotypes or the 95 genera were used to construct sample-similarity matrices using the Bray-Curtis algorithm[27], where samples were ordinated using non-metric multidimensional scaling (nMDS) with 50 random restarts[28]. Significant differences between *a priori* predefined groups of samples were evaluated using Analysis of Similarity (ANOSIM) (9999 permutations)[28] and/or Permutational Multivariate Analysis of Variance (PERMANOVA), allowing for type III (partial) sums of squares, fixed effects sum to zero for mixed terms, and Monte Carlo p-values generated using unrestricted permutation of raw data[29]. Groups of samples were considered significantly different if the p-value was <0.05. The abundances of phyla, genera and of those phylotypes that represent ≥0.5% of the total community in at least one sample (379 phylotypes), were compared by the Mann-Whitney test with Benjamini-Hochberg correction[30] for multiple comparisons. Groups of samples were considered significantly different if the adjusted p-value was <0.05. For analyzing the effect of *H. pylori* on the upper GI tract, communities of saliva (OA), stomach aspirates (SA), duodenum aspirates (DA), stomach biopsies (SB) and duodenum biopsies (DB) of individuals infected with and without *H. pylori* (H and X respectively), were compared: OAX/OAH, SAX/SAH, DAX/DAH, SBX/SBH, and DBX/DBH respectively. Differences in

bacterial communities of aspirates versus biopsies and between biopsies were identified by comparing SAX/SBX, DAX/DBX, SBX/DBX, SBX/DAX as well as SAH/SBH, DAH/DBH and SBH/DBH. Differences across different regions were identified by comparing OAX/SAX, SAX/DAX, OAH/SAH and SAH/DAH. Differences in the abundance of genera and phylotypes in comparisons comprising SAH and DAH were performed excluding *Helicobacter* reads.

Factors ascribed to individuals in this study (age, gender, use of proton pump inhibitors and suffering from intestinal metaplasia or not) were used to test for differences between groups of individuals being infected with and without *H. pylori* using the unpaired t-test with Welch's correction or the Fisher's exact test (two-tailed, confidence intervals at 95%) (see Supplementary Table 1).

RESULTS

Active bacterial communities present in saliva, gastric and duodenal fluids and gastric and duodenal biopsies from 24 individuals comprising eight individuals with a clinical and histopathological diagnosis of *H. pylori* induced gastritis and sixteen individuals with chronic gastritis but without *H. pylori* infection were compared to identify individual as well as site specific bacterial signatures and to identify the influence of *H. pylori* on the global community structures. Following sequencing, rarefaction analysis (Supplementary Figure 1) and rarefying of library sizes, 687 phylotypes were observed (Supplementary Table 2 and 3). The phylotypes belonged to 95 genera and 11 phyla, where sequences of Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria and Fusobacteria comprised approximately 99% of the total bacterial community (Supplementary Table 4 and 5).

Inter-individual variation in the global bacterial assemblages

The global bacterial profiles at phylotype level were compared using group-average agglomerative hierarchical clustering following sample-pairwise comparisons (Figure 1). Assemblages across the upper gastrointestinal tract from each individual typically clustered together irrespective of sample origin. Only communities obtained from stomach biopsies

from *H. pylori* positive individuals clustered based on sample origin, due to the high abundance of *H. pylori*.

This strong host effect at phylotype level is supported by PERMANOVA confirming a strong and statistically significant difference between individuals overall (Pseudo-F=4.46, p=0.001), where 230 of the 253 possible sample-pairwise comparisons were significantly different ($p<0.05$) (Supplementary Table 6). As an example of the uniqueness of samples from a single individual, the communities obtained from P5 are dominated by a single phylotype (Phy10, *Streptococcus salivarius*) accounting for 53% of reads obtained from duodenal aspirates, where only three further individuals contained this phylotype in an abundance >1% in any community (Figure 1). P2 was the only individual, where microorganisms of the candidate phylum SR1 (Phy312) reached a high relative abundance (2.5%).

Similarities in the bacterial community assemblages between locations of the upper human GI tract

Formal comparisons between the global bacterial assemblages across the three different regions of the upper GI tract and between biopsies and individuals infected with or without *H. pylori* revealed that the global bacterial structures were indeed significantly different (PERMANOVA pseudo-F=2.71, p=0.001) as also indicated in the nMDS plot (Figure 2A and B). Bacterial communities of stomach biopsies and stomach aspirates were significantly different in both infected and non-infected individuals as assessed by PERMANOVA and ANOSIM (Figure 2C and Supplementary Table 7). Communities associated with saliva and stomach aspirates of non-infected individuals showed very similar community structures (Figure 2).

Comparisons between the global bacterial assemblages at genus level (Supplementary Table 6 and Supplementary Figure 2) as compared to the phylotype level showed a less pronounced influence of the host as only 87 pairwise comparisons (34%) between individuals showed the presence of distinct communities, indicating that individual differences are highly

governed by differences in phylotypes rather than genera. PERMANOVA analysis of bacterial assemblages across the three different regions of the human upper gastrointestinal tract at genus level showed the same statistically significant differences between those regions where differences had been observed at phylotype level (Supplementary Table 7).

The phylotype richness differed only to a minor extent between the different sampled sites and a significant difference was observed only in biopsies of *H. pylori* infected individuals (Figure 3) where SBH samples showed a much lower phylotype diversity and phylotype evenness compared to all other samples. Neither a reduction in the phylotype richness nor evenness was observed in DBH samples (Figure 3), indicating that the effect of *H. pylori* on community diversity is restricted to the stomach. Diversity measures of communities obtained from individuals without *H. pylori* infection along the upper gastrointestinal tract were all similar (Figure 3).

In contrast, there were no statistical significant differences in average taxonomic distinctness (delta+) and variation in taxonomic distinctness (lambda+) between groups of samples. Thus, while the presence of *H. pylori* changed the phylotype diversity and richness, it did not change the taxonomic representation (Figure 3 and Supplementary Figure 3).

Bacterial community in the fluids of the upper human GI tract of individuals without *H. pylori* infection

Differences in bacterial community composition between sample sites were already evident at the phylum level (Figure 4A). While communities of saliva (OAX) and stomach aspirates (SAX) did not differ significantly, clear differences were observed between SAX and duodenum aspirates (DAX). The relative abundance of Actinobacteria and Firmicutes was higher in DAX (p= 0.03 and 0.004, respectively, Supplementary Table 8), whereas the relative abundance of Bacteroidetes and Fusobacteria was lower (p= 0.0018 and 0.0011, respectively). At the genus level, the upper human gastrointestinal tract communities in *H. pylori* negative individuals were dominated by *Streptococcus* and *Prevotella* (Figure 4B). Seven genera differed in abundance between OAX and SAX, with *Veillonella* being more

abundant in OAX and *Eubacterium, Fusobacterium, Lachnoanaerobaculum, Olsenella*, *Parvimonas* and *Tessarococcus* being more abundant in SAX (Figure 4C, Supplementary Table 9).

Differences were more pronounced when SAX and DAX samples were compared. Twenty-two genera were differently abundant where the higher abundance of Firmicutes in DAX can be attributed to a higher abundance of *Streptococcus*, and the lower abundance of Bacteroidetes and Fusobacteria to a lower abundance of *Prevotella* and *Fusobacterium* (Figure 4C, Supplementary Table 9), respectively, among others. Different niches for different genera became apparent with various Proteobacteria, Actinobacteria and *Staphylococcus* being virtually absent from SAX (Figure 4C).

Analysis at phylotype level revealed that Phy44 (*Fusobacterium*) had the significantly highest abundance in SAX (Figure 4B and C). Also Phy269 (*Campylobacter gracilis*) was specifically abundant in SAX and other *Campylobacter* phylotypes showed the same trend (Figure 4C and D). Of the Bacteroidetes phylotypes, especially those indicating the presence of *Prevotella* showed a higher abundance in SAX compared to DAX (Figure 4C and D). In contrast, *Staphylococcus* phylotypes showed the lowest abundance in SAX compared to DAX (see Figure 4C). A similar low abundance specifically in SAX, indicating vulnerability of the respective organism to acidic conditions, was observed for various actinobacterial and γ-proteobacterial phylotypes (Figure 4C and Supplementary Table 10).

Bacterial communities in the fluids of the stomach and duodenum compared to biopsies

An analysis at phylum level showed a clear difference between SAX and SBX in the abundance of Fusobacteria, Bacteroidetes and Firmicutes (Figure 5A and Supplementary Table 8) whereas those between DAX and DBX were restricted to Spirochaetes (Figure 5B). Accordingly, differences in community composition between DAX and DBX samples at genus level showed *Treponema* to be significantly more abundant in DAX (Figure 5C). In addition, four proteobacterial and two actinobacterial genera differed significantly in abundance. In

contrast, 22 genera were differently abundant in SBX versus SAX samples and *Fusobacterium* strains were responsible for the higher abundance of Fusobacteria in SAX while *Streptococcus, Staphylococcus* and *Enterococcus* for the higher abundance of Firmicutes in SBX (Figure 5C-D and Supplementary Table 9). Members of the genera *Alloprevotellla, Porphyromonas* and *Tannerella* were responsible for the higher abundance of Bacteroidetes in SAX whereas *Bacteroides* followed a different trend being more abundant in SBX (Figure 5C). Both genera of higher abundance in SAX and those of higher abundance in SAX were observed among the Actinobacteria and Proteobacteria (Figure 5C). Fourteen of the 22 genera detected in significantly different abundance in SBX compared to SAX were also in significantly different abundance in DAX compared to SAX (see Figure 5C), which may indicate SAX to be distinct from SBX, DAX and DBX with regard to the genera distribution pattern. Distributions observed at genus level were also evident at the phylotype level (Figure 5C and Supplementary Table 10).

H. pylori infection affects the stomach microbiota

Infection with *H. pylori* is typically manifested by a high abundance of this organism on the mucosa, and accordingly, SBH samples revealed the presence of *Helicobacter* in amounts exceeding 50% in all infected individuals (Figure 6A). Due to the high relative abundance of *H. pylori,* the relative abundance of all four major phyla, i.e. Actinobacteria, Bacteroidetes, Firmicutes and Fusobacteria, was low in SBH compared to SBX.

Three different *H. pylori* phylotypes were dominant in stomach biopsies with Phy1 dominating in six individuals and Phy27 and Phy36 dominating in one individual each. The same *Helicobacter* phylotypes dominating in biopsies were also observed in the respective stomach aspirates, however to a much lesser extent and seven of the eight individuals were infected by only one *Helicobacter* phylotype.

Analysis of the community composition of SAH versus SAX revealed significant differences only in the relative abundance of Proteobacteria, which were of higher abundance in SAH, concomitant with a lower relative abundance of all nine other phyla

detected in this niche (Figure 6A). Similarly, at genus level, besides the increase in *Helicobacter* only *Haemophilus* decreased significantly in abundance in SAH samples (Figure 6B) while at phylotype level only Phy25 (*Haemophilus*) and Phy80 (*Campylobacter conciscus*) showed a significant decrease (Supplementary Table 9 and 10).

Influence of *H. pylori* on duodenal and oral communities

The only significant influence of *H. pylori* infection on duodenal communities at the phylum level was the increased abundance of Proteobacteria in DAH compared to DAX probably due to transfer of *Helicobacter* from the stomach to the duodenal lumen (Figure 7A). Infection with *H. pylori* may have resulted in a lower abundance of Firmicutes and a higher abundance of Bacteroidetes in DBH biopsies of infected compared to DBX biopsies of uninfected individuals, however, the differences were statistically not significant. Except for the different abundances of *Helicobacter* in aspirates, no statistically significant differences were observed at genus level (Figure 7B). Differences became apparent at phylotype level. The differences in biopsies seem to be minor and only Phy78 (*Rhodobacteriaceae*) and Phy269 (*Campylobacter gracilis*) were significantly reduced in abundance in biopsies ($p=0.05$, Figure 7C). Other phylotypes were only observed in biopsies of non-infected individuals (Phy140, *Streptococcus infantis*; Phy168, *Actinomyces*; Phy81, *Enterococcus* and Phy478, *Lachnospiraceae*), indicating some influence of *H. pylori*. Besides Phy1 (*H. pylori*), which was observed in high abundance in the duodenal aspirates of infected individuals, only Phy103 (*Staphylococcus aureus*) was observed in significantly different abundances in DAH versus DAX samples, being more abundant in DAX (Figure 7C and Supplementary Table 10).

There was no significant difference in oral communities between individuals infected with and without *H. pylori*, neither at phylum (Figure 7A) nor at genus level. However, of the phylotypes three were significantly more abundant in non-infected individuals (Phy18, *Propionibacterium acnes*; Phy25, *Haemophilus* and Phy62, *Prevotella oris*) while Phy475 (*Treponema*) was absent from all non-infected individuals but observed in 4 of the 8 infected individuals (Figure 7D).

DISCUSSION

Despite the high number of studies addressing the bacterial biodiversity of the human upper gastrointestinal tract[31], there is still a lack of knowledge regarding the role of the commensal bacterial community in this niche. In this study, the bacterial communities of five sites from 24 individuals were defined where the use of reverse transcribed 16S rRNA instead of 16S rDNA as template allowed the detection of the metabolically active component of the community [16, 17].

Previous studies on the gastric bacterial communities of biopsies or aspirates have not permitted the characterization of different ecological niches in their anatomical continuity [3, 10-12]. By analyzing throat and stomach biopsies from 3 individuals each it was reported that prominent genera of the stomach were also abundant in the throat and may be swallowed microorganisms[12]. Further studies have compared the saliva and gastric microbiome of individuals characterized by an exceptional abundance of Enterobacteriaceae[9] or were biased by the inclusion of individuals infected with *H. pylori*, who dominated gastric mucosal communities[32] allowing only very limited conclusions. The identification of host specific phylotypes in our study evidences the saliva as main source for the gastric microbiome. In fact, each subject has an indigenous microbiota composition, which is consistent throughout the investigated regions that are connected by a constant fluid drain. This underlines the concept of a continuous bacterial migration through the upper GI tract with the oral cavity as the dominant source of active bacteria. Our analyses of the active microbiota of the stomach confirms that *Veillonella* is depleted and *Fusobacterium* enriched in aspirates compared to saliva. Various groups of organisms were depleted in stomach aspirates, for example *Propionibacterium.* This shows the analysis of the active microbiota to be advantageous compared to the analysis of bacterial presence[8] and indicates the stomach environment to select against various organisms, which are active in either saliva or in the duodenum such as *Staphylococcus* or *Streptococcus*.

Despite the considerable variation in the gastric microbiota composition between individuals and also between studies, studies usually state *Streptococcus, Lactobacillus, Bacteroides, Staphylococcus, Prevotella, Fusobacterium* and *Veillonella* among others[3, 10, 33] as prominent genera. Information on intra-individual differences between luminal and mucosal samples, however is scarce [2, 10, 13] and studies were usually performed on a small number of individuals and biased by the inclusion of individuals infected with *H. pylori.* In this work, the mucosal samples comprised relatively lower concentrations of phylotypes belonging to the Bacteroidetes genera *Prevotella, Alloprevotella, Porphyromonas* and *Tannerella*, and relatively higher concentrations of phylotypes belonging to the Firmicutes genera *Streptococcus* and *Staphylococcus* as well as the Actinobacteria genera *Corynebacterium, Kocuria* and *Propionibacterium.* This provides evidence for a major difference between the luminal community, probably adapted to a low pH value and the mucosa adherent community of the stomach.

In the duodenum, *Streptococcus* and *Prevotella* were described as the most prominent genera[34]. However, a recent sequencing analysis on five healthy subjects indicated *Prevotella* and generally Bacteroidetes to be virtually absent from duodenum mucosa[35]. More recently the Proteobacteria genera *Acinetobacter, Neisseria* and *Haemophilus* but also *Prevotella* were reported to be the most predominant genera[36]. The observation here of *Streptococcus, Prevotella, Actinomyces* and also of *Propionibacterium, Gemella* and *Fusobacterium* as abundant in duodenum biopsies is based on a cohort of 24 individuals and, thus gives for the first time a representative overview of this environment. Only minor differences between the mucosal communities were evident, which is in strong contrast to the clear preference of *Helicobacter* to colonize the stomach mucosa. More generally, *Campylobacter* phylotyopes were preferentially found in aspirates and *Enhydrobacter* in mucosal samples. Similarly, only minor differences were noted between duodenal aspirate and mucosal communities. One recent study also had compared the communities of duodenal biopsies and luminal in nine volunteers noting that *Acinetobacter, Bacteroides* and *Prevotella* were the most abundant genera in biopsies, and *Prevotella, Stenotrophomonas*

and *Streptococcus* the most abundant ones of the lumen[37]. *Stenotrophomonas* was absent from all active communities analyzed here and can be assumed to be present only in exceptional cases.

In accordance with previous reports[3, 13, 32], the gastric mucosa of *Helicobacter* infected individuals was highly dominated by this organism concomitant with the relative depletion of other genera. Even though the presence of *H. pylori* had a significant influence on species richness and evenness, it did not change the taxonomic diversity. Interestingly, a recent report[38] claimed that *H. pylori* infected adults are likely to have higher abundances of Spirochetes, Acidobacteria and non-*Helicobacter* Proteobacteria, of which none were detected in high abundance in this study. Strikingly, the influence of *Helicobacter* on the community was more evident in the duodenal samples and respective influences have, to our knowledge, yet to be reported. Duodenal ulcers and gastric cancer as complications of *H. pylori* infection have been reported to coexist in a small number of patients[39] and it is assumed that undetected factors such as different compositions of the bacterial community beyond *H. pylori* contribute to this pathology.

Interestingly, infection with *H. pylori* influences the composition of the oral bacterial community and/or vice versa. Assuming an oral to oral transmission route o*f H. pylori*[4] differences in the active oral community might influence the susceptibility of the host to *H. pylori* infection and a better understanding on the interaction between *H. pylori* and the oral community might give insights into the mode of transmission.

Acknowledgements

The authors like to thank Iris Plumeier and Silke Kahl for technical assistance, Robert Geffers and Michael Jarek for sequencing support and Diego Chaves-Moreno and Robert Thänert for critical discussions.

Competing Interests

None

Funding

This study was supported by iMed, the Helmholtz Association's Initiative on Personalized Medicine. C.S. was supported by CRC854, a research program by the German funding organization DFG.

Author Contributions

D.H.P., P.M. and C.S. contributed to the study design, analysis and interpretation of the data. D.H.P. and C.S. drafted the manuscript. C.S., K.S. and P.M. collected samples and identified suitable subjects. D.P.H. and P. M. supervised the study procedures. C.S., N.K. and A.P.O. established analyses and performed laboratory work-up. M.V., D.H.P. M.W-O and R.V.V. performed bioinformatic and statistical analyses. All authors read and approved the final version of the manuscript.

REFERENCES

1. Malfertheiner P, Link A, Selgrad M. *Helicobacter pylori*: perspectives and time trends. Nat Rev Gastroenterol Hepatol 2014;11:628-638.
2. Sanduleanu S, Jonkers D, De Bruine A, et al. Non-*Helicobacter pylori* bacterial flora during acid-suppressive therapy: differential findings in gastric juice and gastric mucosa. Aliment Pharmacol Ther 2001;15:379-388.
3. Bik EM, Eckburg PB, Gill SR, et al. Molecular analysis of the bacterial microbiota in the human stomach. Proc Natl Acad Sci U S A 2006;103:732-737.
4. Kusters JG, van Vliet AH, Kuipers EJ. Pathogenesis of *Helicobacter pylori* infection. Clin Microbiol Rev 2006;19:449-490.
5. He C, Yang Z, Lu N. Imbalance of gastrointestinal microbiota in the pathogenesis of *Helicobacter pylori*-associated diseases. Helicobacter 2016;21:337-348.

6. Amieva MR, El-Omar EM. Host-bacterial interactions in *Helicobacter pylori* infection. Gastroenterology 2008;134:306-323.
7. Malfertheiner P. The intriguing relationship of *Helicobacter pylori* infection and acid secretion in peptic ulcer disease and gastric cancer. Dig Dis 2011;29:459-464.
8. Bassis CM, Erb-Downward JR, Dickson RP, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. MBio 2015;6:e00037.
9. Ye F, Shen H, Li Z, et al. Influence of the biliary system on biliary bacteria revealed by bacterial communities of the human biliary and upper digestive tracts. PLoS One 2016;11:e0150519.
10. Delgado S, Cabrera-Rubio R, Mira A, et al. Microbiological survey of the human gastric ecosystem using culturing and pyrosequencing methods. Microb Ecol 2013;65:763-772.
11. Eun CS, Kim BK, Han DS, et al. Differences in gastric mucosal microbiota profiling in patients with chronic gastritis, intestinal metaplasia, and gastric cancer using pyrosequencing methods. Helicobacter 2014;19:407-416.
12. Andersson AF, Lindberg M, Jakobsson H, et al. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One 2008;3:e2836.
13. Sung J, Kim N, Kim J, et al. Comparison of gastric microbiota between gastric juice and mucosa by next generation sequencing method. J Cancer Prev 2016;21:60-65.
14. Wacklin P, Laurikka P, Lindfors K, et al. Altered duodenal microbiota composition in celiac disease patients suffering from persistent symptoms on a long-term gluten-free diet. Am J Gastroenterol 2014;109:1933-1941.
15. Malfertheiner P, Chan FK, McColl KE. Peptic ulcer disease. Lancet 2009;374:1449-1461.
16. Gaidos E, Rusch A, Ilardo M. Ribosomal tag pyrosequencing of DNA and RNA from benthic coral reef microbiota: community spatial structure, rare members and nitrogen-cycling guilds. Environ Microbiol 2011;13:1138-1152.

17. Frias-Lopez J, Shi Y, Tyson GW, et al. Microbial community gene expression in ocean surface waters. Proc Natl Acad Sci U S A 2008;105:3805-3810.
18. Camarinha-Silva A, Jauregui R, Chaves-Moreno D, et al. Comparing the anterior nare bacterial community of two discrete human populations using Illumina amplicon sequencing. Environ Microbiol 2014;16:2939-2952.
19. Chaves-Moreno D, Plumeier I, Kahl S, et al. The microbial community structure of the cotton rat nose. Environ Microbiol Rep 2015;7:929-935.
20. Cole JR, Wang Q, Fish JA, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 2014;42:D633-642.
21. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009;75:7537-7541.
22. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 2013;8:e61217.
23. Wang Q, Garrity GM, Tiedje JM, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007;73:5261-5267.
24. Oksanen J, Guillaume-Blanchet F, Kindt R, et al. vegan: Community Ecology Package. R package version 2.3-2. http://CRAN.R-project.org/package=vegan, 2015.
25. Pienkowski MW, Watkinson AR, Kerby G, et al. A taxonomic distinctness index and its statistical properties. J Appl Ecol 1998;35:523-531.
26. Warwick R, Clarke K. New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. Mar Ecol Prog Ser 1995;129:301-305.
27. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecological monographs 1957;27:325-349.
28. Clarke KR, Warwick RMRM, Laboratory PM. Change in marine communities : an approach to statistical analysis and interpretation: Plymouth, U.K. : PRIMER-E Ltd, 2001.

29. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol 2001;26:32-46.

30. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9:811-818.

31. Schulz C, Koch N, Schutte K, et al. *H. pylori* and its modulation of gastrointestinal microbiota. J Dig Dis 2015;16:109-117.

32. Stearns JC, Lynch MD, Senadheera DB, et al. Bacterial biogeography of the human digestive tract. Sci Rep 2011;1:170.

33. Sheh A, Fox JG. The role of the gastrointestinal microbiome in Helicobacter pylori pathogenesis. Gut Microbes 2013;4:505-531.

34. Nadal I, Donat E, Ribes-Koninckx C, et al. Imbalance in the composition of the duodenal microbiota of children with coeliac disease. J Med Microbiol 2007;56:1669-1674.

35. Angelakis E, Armougom F, Carriere F, et al. A metagenomic investigation of the duodenal microbiota reveals links with obesity. PLoS One 2015;10:e0137784.

36. D'Argenio V, Casaburi G, Precone V, et al. Metagenomics reveals dysbiosis and a potentially pathogenic *N. flavescens* strain in duodenum of adult celiac patients. Am J Gastroenterol 2016;111:879-890.

37. Li G, Yang M, Zhou K, et al. Diversity of duodenal and rectal microbiota in biopsy tissues and luminal contents in healthy volunteers. J Microbiol Biotechnol 2015;25:1136-1145.

38. Maldonado-Contreras A, Goldfarb KC, Godoy-Vitorino F, et al. Structure of the human gastric bacterial community in relation to *Helicobacter pylori* status. ISME J 2011;5:574-579.

39. Ubukata H, Nagata H, Tabuchi T, et al. Why is the coexistence of gastric cancer and duodenal ulcer rare? Examination of factors related to both gastric cancer and duodenal ulcer. Gastric Cancer 2011;14:4-12.

Figure Legends

Figure 1: Group-average agglomerative hierarchical clustering of 119 samples, based on global bacterial profiles (phylotype-level) along the upper gastrointestinal tract (saliva, O; stomach, S; and duodenum, D) where aspirates (A) and biopsies (B) from 24 individuals infected with (H) or without (X) *H. pylori*. Individuals are denoted by unique symbols. Four selected phylotypes, which were present in high abundance in samples from single individuals are shown as inserts. The y-axis shows the relative abundance of the phylotype in a given sample, while the x-axis shows samples originating from individual 1 – 24, which in each case are given in the order OA, SA, DA, SB and DB.

Figure 2. Differences in global bacterial community structures along theupper gastrointestinal tract (saliva, O; stomach, S; and duodenum, D) of aspirates (A) and biopsies (B) from 24 individuals infected with (H) or without (X) *H. pylori,* as assessed by non-metric multidimensional scaling (nMDS). (A) Global community structure based on standardized phylotype abundance data. (B) Global community structure excluding stomach biopsies of *H. pylori* infected individuals based on standardized genus abundance data. (C) Differences between groups of samples as evaluated using ANOSIM with the R statistic measuring the degree of separation between groups where ns denotes “not significant”.

Figure 3. Upper gastrointestinal tract bacterial community diversity indicated by total phylotype number, Shannon diversity (H´), Pielou's evenness (J´) taxonomic diversity (delta+) and taxonomic evenness (lambda+), respectively. Statistically significant differences between groups of samples from adjacent sites or from the same region but between individuals infected with or without *H. pylori* are indicated by **** $p<0.0001$; *** $p<0.001$. * $p<0.05$. The mean and SD are shown.

Figure 4. Differences in bacterial community structures along the upper gastrointestinal tract of aspirates from individuals not infected with *H. pylori*. (A) Relative mean abundance of phyla and (B) relative abundance of genera predominant in oral (OAX), stomach (SAX) and duodenal aspirates (DAX) where significantly differently distributed phyla and genera are indicated with asterisks * $p<0.05$, ** $p<0.01$, *** $p<0.001$. (C) Cladogram showing the differently distributed genera and phylotypes (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). (D) Abundance of selected phylotypes which were differently distributed between OAX, SAX and DAX. Colors of lines used in (C) and colors in (D) correspond to the color code used for phyla given in (A).

Figure 5. Differences in bacterial community structures between stomach and duodenum biopsies and aspirates in individuals not infected with *H. pylori***.** (A) Relative mean abundance of phyla and (B) relative mean abundance of low abundant phyla. (C) Cladogram showing the differently distributed genera and phylotypes. Colors of lines in (C) correspond to the color code used for phyla given in (A). If a genus or phylotype is differently distributed between SAX/SBX, SAX/DAX, DAX/DBX or SBX/DAX ($p>0.05$), the respective genus or phylotype name is colored, with the color code given at the bottom of (C). The level of statistical significance is indicated above the respective color code. (D) Relative mean abundance of prominent genera in stomach aspirates (SAX) and biopsies (SBX) as well as in duodenal aspirates (DAX) and biopsies (DBX). Significantly differently distributed phyla and genera are indicated with asterisks * $p<0.05$, ** $p<0.01$ and *** $p<0.001$.

Figure 6. Differences in bacterial community structures between stomach aspirates and biopsies in individuals infected with and without *H. pylori*. (A) Relative mean abundance of phyla and (B) relative mean abundance of genera in stomach aspirates of individuals infected with (SAH) or without *H. pylori* (SAX) as well as in stomach biopsies of non-infected (SBX) or infected individuals (SBH).

Figure 7. Differences in bacterial community structures between duodenum aspirates and biopsies and saliva in individuals infected with and without *H. pylori*. (A) Relative mean abundance of phyla and (B) relative mean abundance of genera in duodenum aspirates of individuals infected with (DAH) or without *H. pylori* (DAX) as well as in duodenum biopsies of non-infected (DBX) or infected individuals (DBH). (C) Abundance of those genera which were significantly differently distributed in duodenal samples between infected or non-infected individuals with asterisks denoting * $p<0.05$ and ** $p<0.01$. (D) Relative mean abundance of phyla which were differently distributed between OAX and OAH samples. Colors used in C and D correspond to the color code used for phyla given in (A).

Supplementary Figure 1: Rarefaction curves portraying the number of resolved phylotypes against sequencing depth for each sample. Samples from individuals infected with *H. pylori* are shown in red, those of non-infected individuals in black.

Supplementary Figure 2: Group-average agglomerative hierarchical clustering analysis of bacterial communities along the upper gastrointestinal tract at the genus level. The Bray-Curtis algorithm was used to assess the similarity between samples based on the relative abundance of genera. The sample regions are saliva (O), stomach (S), duodenum (D) and sample types are aspirates (A), biopsies (B), and *H. pylori* negative (X) or *H. pylori* positive (H). The 24 individuals are indicated by a unique symbol.

Supplementary Figure 3: Taxonomic distinctness and evenness of microbial communities along the gastrointestinal tract. Funnel plots chart the (A) average taxonomic distinctness (delta+) and (B) variation in taxonomic distinctness (lambda+) against the number of phylotypes within each sample. The funnel indicates the limits within 95% of the simulated TD values and the middle line represents the mean expected TD. (C) Charts average taxonomic distinctness (lambda+) against variation in taxonomic distinctness (delta+).