The ISME Journal (2011) 5, 574–579

www.nature.com/ismej

ORIGINAL ARTICLE

Structure of the human gastric bacterial community in relation to *Helicobacter pylori* status

Ana Maldonado-Contreras[1,5], Kate C Goldfarb[2,5], Filipa Godoy-Vitorino[1], Ulas Karaoz[2], Mónica Contreras[3], Martin J Blaser[4], Eoin L Brodie[2] and Maria G Dominguez-Bello[1]
[1]*Department of Biology, University of Puerto Rico, San Juan, PR, USA;* [2]*Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA;* [3]*Venezuelan Institute for Scientific Research (IVIC), Altos del Pipe, Miranda, Venezuela and* [4]*Departments of Medicine and Microbiology, New York University Langone Medical Center, New York, NY, USA*

The human stomach is naturally colonized by *Helicobacter pylori*, which, when present, dominates the gastric bacterial community. In this study, we aimed to characterize the structure of the bacterial community in the stomach of patients of differing *H. pylori* status. We used a high-density *16S rRNA* gene microarray (PhyloChip, Affymetrix, Inc.) to hybridize *16S rRNA* gene amplicons from gastric biopsy DNA of 10 rural Amerindian patients from Amazonas, Venezuela, and of two immigrants to the United States (from South Asia and Africa, respectively). *H. pylori* status was determined by PCR amplification of *H. pylori glmM* from gastric biopsy samples. Of the 12 patients, 8 (6 of the 10 Amerindians and the 2 non-Amerindians) were *H. pylori glmM* positive. Regardless of *H. pylori* status, the PhyloChip detected Helicobacteriaceae DNA in all patients, although with lower relative abundance in patients who were *glmM* negative. The G2-chip taxonomy analysis of PhyloChip data indicated the presence of 44 bacterial phyla (of which 16 are unclassified by the Taxonomic Outline of the Bacteria and Archaea taxonomy) in a highly uneven community dominated by only four phyla: Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes. Positive *H. pylori* status was associated with increased relative abundance of non-*Helicobacter* bacteria from the Proteobacteria, Spirochetes and Acidobacteria, and with decreased abundance of Actinobacteria, Bacteroidetes and Firmicutes. The PhyloChip detected richness of low abundance phyla, and showed marked differences in the structure of the gastric bacterial community according to *H. pylori* status.
The ISME Journal (2011) **5,** 574–579; doi:10.1038/ismej.2010.149; published online 7 October 2010
Subject Category: microbial population and community ecology
Keywords: gastric; microbiota; *H. pylori*; microarray; Amerindians

Introduction

Helicobacter species are natural colonizers of the mammalian stomach, and *H. pylori* has coevolved with its human host (Falush *et al.*, 2003; Linz *et al.*, 2007). In addition to *H. pylori*, the stomach can also contain transient oral, esophageal or intestinal bacteria. To date, the few studies that have explored the microbiota of the human stomach using molecular methods (Bik *et al.*, 2006; Andersson *et al.*, 2008) have shown that the gastric community is highly dominated by Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes, with *H. pylori* being the single dominant bacteria in patients of positive *H. pylori* status. Bik *et al.* (2006) sequenced 1833 bacterial clones from gastric biopsy samples from 23 US patients and found eight bacterial phyla (128 phylotypes) with no differences in richness by *H. pylori* status and with 7 of 11 patients of apparent negative *H. pylori* status having *H. pylori* clones.

In a more recent study using tagged 454 pyrosequencing, Andersson *et al.* (2008) produced 23 713 reads from gastric biopsy samples from six Swedish patients, finding 13 bacterial phyla and higher gastric diversity in patients of negative *H. pylori* status (262 phylotypes) in comparison with patients of positive *H. pylori* status (33 phylotypes).

Major problems of current molecular techniques include PCR biases (Farrelly *et al.*, 1995; Suzuki and Giovannoni, 1996; Polz and Cavanaugh, 1998), poor sampling by cloning (DeSantis *et al.*, 2007) and overestimation of richness by 454 sequencing (Kunin *et al.*, 2009). DNA microarrays overcome some of these biases to detect the presence and relative proportion of known bacteria (Brodie *et al.*, 2006; DeSantis *et al.*, 2007).

In this work, we used DNA microarrays to characterize the gastric bacterial community structure in patients differing in *H. pylori* status.

Correspondence: EL Brodie, Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, MS 70A-3317, Berkeley, CA 94720, USA or M Dominguez-Bello, University of Puerto Rico, Rio Piedras Campus, PO BOX 23360, Rio Piedras, JGD 224, San Juan, PR 00931, USA.
E-mails: ELBrodie@lbl.gov or maria.dominguez1@upr.edu
[5]These authors contributed equally to this work.
Received 26 July 2010; accepted 2 August 2010; published online 7 October 2010

Materials and methods

Patients

Corpus biopsy samples obtained from 12 adult patients were included in the study. Of them, 10 were Amerindians (5 Guahibo and 5 Piaroa, age 25–80 years) who underwent upper gastrointestinal endoscopy at the Clinica Ayacucho in Amazonas, Venezuela. The Guahibo occupy a large territory in both Colombia and Venezuela, in the latter within the Amazonas State. The Piaroa live in sylvatic areas close to the Orinoco river (Freire, 2007) but, like the Guahibo, also live in rural communities near Puerto Ayacucho, and are currently subject to rapid acculturation. Two patients were from other developing communities and were used as references. They were recent adult immigrants to the United States, one from Bangladesh and one from Rwanda (with 7 and 2 years in the United States, respectively), who consulted at Bellevue Hospital in New York City. All patients were fasting at least 12 h before the sampling. No information about previous antibiotic treatments was recorded. Patients provided signed informed consent to participate, and samples were managed without personal identifiers. The sampling protocols were approved by the IRBs at the Venezuelan Institute of Scientific research, IVIC (#0229/10), New York University (#12206) and University of Puerto Rico (#0809–051).

DNA extraction and H. pylori *status*

Gastric corpus biopsy DNA was extracted using the DNeasy tissue kit (Qiagen, Chatsworth, CA, USA), after homogenizing biopsy samples in 200 µl saline solution (0.9% NaCl) with ~0.1 ml glass beads (0.5 mm) in 1.5 ml tubes, mixing at high speed for 20 s in a bead beater. *H. pylori* status for each subject was determined by amplification of *glmM,* encoding a phosphotransferase conserved in *H. pylori* (Lu *et al.*, 1999), and all assays included negative and positive controls.

DNA preparation for the hybridization array

Bacterial *16S rRNA* gene amplicons from gastric biopsy samples of the 12 patients were hybridized onto *16S rRNA* gene microarrays (Brodie *et al.*, 2006). The accuracy of results obtained with the PhyloChip analysis at suprafamily levels has been previously validated using both quantitative PCR and *16S rRNA* gene clone libraries (Brodie *et al.*, 2007; DeSantis *et al.*, 2007). First, gastric biopsy DNA was amplified using primers specific for bacterial *16S rRNA*. Primers were the universal 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′) (Lane, 1991). Each PCR mix contained 50 units ml^{-1} of Taq DNA polymerase, 400 µM of each dNTP, 3 mM of $MgCl_2$ and 5 pmol of each primer. The gradient PCR followed these steps: 3 min at 95 °C, followed by 25 cycles of 95 °C for 30 s, gradient temperature of annealing from 48 to 58 °C for 25 s, 25 s at 72 °C, and final extension at 72 °C for 2 min. Pooled amplicons from the eight different annealing temperatures were purified using the QIAquick PCR Purification Kit, following the manufacturer's instructions (Qiagen, Chatsworth, CA, USA). *E. coli* genomic DNA was used as a PCR positive control and we also included a negative control without DNA.

The G2 PhyloChip

The bacterial *16S rDNA* amplicons were hybridized onto the G2 PhyloChip, a hybridization array developed by the Lawrence Berkeley National Laboratory (LBL) (Brodie *et al.*, 2007; DeSantis *et al.*, 2007). The Phylochip has been validated (Brodie *et al.*, 2007; DeSantis *et al.*, 2007), detecting 90% of the cloned subfamilies and unveiling 2.5-fold higher diversity than cloning. We believe that there is sufficient evidence to rely on the G2 PhyloChip at the subfamily level at higher taxonomic levels.

The G2 PhyloChip contains 297 851 probes targeting *16S rRNA* genes representing 8741 taxa. A taxon is defined as the result of grouping more than 30 000 records of *16S rRNA* gene sequences (at least 600 bp) reported within the 15 March 2002 release of the *16S rDNA* database, www.greengenes.lbl.gov. Each of the 8741 clusters on the PhyloChip represents a taxon, and in total includes all 121 demarcated bacterial and archaeal orders. For each taxon, the PhyloChip has 11 probe pairs of 25-mers. Taxons belong to subfamily taxonomic levels. As there can be overlapping probes for each taxon, we counted unique records at suprasubfamily and higher levels. The purified product of the amplification of the *16S rRNA* gene (200 ng) was fragmented using DNase I, biotin-labeled and then hybridized onto the PhyloChip, as described (Brodie *et al.*, 2006). After overnight hybridization at 48 °C and 60 r.p.m., the PhyloChips were washed and stained according to standard Affymetrix protocols, as described (Masuda and Church, 2002). The PhyloChips were scanned and recorded as pixel images using Gene Array Scanning (Affimetrix Inc., Santa Clara, CA, USA), and initial data acquisition and intensity determinations were performed using the standard Affymetrix software GeneChip microarray analysis suite, version 5.1. Background probes, noise and standard deviation of intensities were determined, as described by the developers (Brodie *et al.*, 2007). Probe pairs were positive when (i) the intensity of fluorescence from the perfectly matched probe was $\geqslant 1.3 \times 10^2$ times higher than that obtained from the mismatched control and (ii) perfectly matched probe intensity minus mismatched control was 130 times greater than the squared noise value (Brodie *et al.*, 2007; DeSantis *et al.*, 2007). A bacterial taxon was considered to be present in a sample when $\geqslant$90% of the probe sets designed for it were positive (positive fraction $\geqslant$0.9) (Brodie *et al.*, 2007; DeSantis *et al.*, 2007). To obtain relative abundances of each

taxon, probe intensities were first trimmed eliminating the highest and lowest values. The mean of the remaining values was normalized to the intensities of the control probes, using a maximum likelihood method to account for variation in PhyloChip processing. After normalization, mean values were scaled by the average overall microarray intensity to account for variation in amplicon quantification and then log-transformed to reduce differential variance at higher concentrations. A Bray–Curtis distance matrix was constructed in R, based on the intensity values of all taxa detected. The function 'adonis' from the R 'vegan' package (Jari Oksanen, 2008) was used to determine the partitioning among sources of variance within this distance matrix, using permutational multivariate analysis of variance (R-project, 2008). The PhyloChip taxonomy is based on Hugenholtz taxonomy (as of April 2007) (Hugenholtz, 2002; DeSantis *et al.*, 2006a, b) and we reclassified the detected taxa using RDP to compare with results in other publications (Eckburg *et al.*, 2005; Gill *et al.*, 2006; Andersson *et al.*, 2008; Dethlefsen *et al.*, 2008; Keijser *et al.*, 2008).

Non-metric multidimensional scaling was used to visualize the variation in two dimensions. Non-metric multidimensional scaling was used because it does not assume linearity of the data and does not require data transformation, which represents advantages over other classical ordination methods (that is, principal component analysis) for assessing community structures (Clarke, 1993). On the basis of ranked similarity distances, an iterative search for the least stress position of data in *k*-dimensions was conducted (Clarke, 1993). In addition, linear regressions were performed to assess the relationship between relative abundance of *H. pylori* and all other taxa.

Results

Bacterial diversity

The PhyloChip detected substantial bacterial richness in gastric communities (Supplementary Table S1). A total of 44 bacterial phyla were detected in an uneven ecosystem, with a strong dominance of only four phyla (in descending order: Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes). There was remarkable similarity in the representation of these four dominant phyla between the *H. pylori*-positive and -negative subjects (Supplementary Table S1). Of the 12 patients, eight had positive status for *H. pylori* (six Amerindians and two non-Amerindians), whereas four (all Amerindians) had negative status (Table 1), based on PCR of the *H. pylori* phosphotransferase gene *glmM*.

Subjects with positive and negative *H. pylori* status had similar phyla richness (Table 2), and the PhyloChip detected Helicobacteriaceae taxa in all 12 patients, including those 4 who failed to amplify *glmM*. However, the *glmM*-PCR-positive patients had five times higher signal for that taxon on the

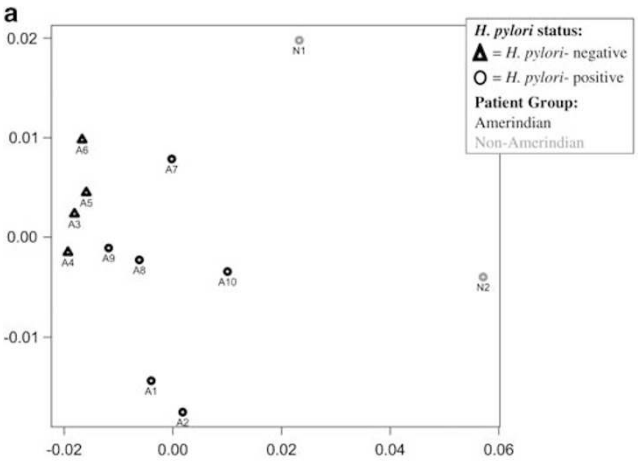
Table 1 Characteristics of 12 study subjects

Sample Code	*Sex*	*Age*	*Ethnic Group*	*Patient origin*	*Diagnosis*	H. pylori *status*[a]
A1	M	46	Piaroa	San Pedro del Orinoco, Amazonas, VZ	Erythematous pre-pyloric region	+
A2	F	58	Guahibo	La Reforma, Amazonas, VZ	Antral gastritis; Vesicular lithiasis	+
A3	M	60	Guahibo	La Reforma, Amazonas, VZ	Hiatal hernia; gastritis	–
A4	F	40	Guahibo	La Reforma, Amazonas, VZ	Gastritis	–
A5	M	80	Guahibo	La Reforma, Amazonas, VZ	Hiatal hernia; gastritis	–
A6	F	37	Guahibo	La Reforma, Amazonas, VZ	Gastritis	–
A7	F	59	Piaroa	La Reforma, Amazonas, VZ	Antral Gastritis	+
A8	M	44	Piaroa	Samaria, Amazonas, VZ	Antral Gastritis	+
A9	M	21	Piaroa	Samaria, Amazonas, VZ	Antral Gastritis	+
A10	F	25	Piaroa	Agua Linda, Amazonas, VZ	Severe inflammation; Erosive duodenitis	+
N1	M	31	SouthAsian	Bangladesh*	Heartburn/GERD symptoms	+
N2	F	39	African	Rwanda*	Dyspepsia	+

Abbreviation: VZ, Venezuela.
*Gastric biopsies obtained at the Bellevue Hospital, New York, NY.
[a]As determined by *glmM* PCR.

Table 2 Taxonomic complexity (mean ± s.d.) by classification of subjects

Taxa	H. pylori *status*		*Human group*		
	Positive (N = *8*)	*Negative* (N = *4*)	*Amerindian* (N = *10*)	*Bangladesh* (N = *1*)	*Rwanda* (N = *1*)
Phylum	39 ± 4	41 ± 4	41 ± 3	32	34
Class	47 ± 5	50 ± 3	50 ± 2	40	41
Order	81 ± 13	91 ± 3	88 ± 6	71	58
Family	137 ± 23	154 ± 3	150 ± 7	119	90



b

Comparison	p-value	% variation explained
***H. pylori* status:** positive vs. negative	0.015	27.9
Patient group: Amerindian vs. non-Amerindian	0.01	44.2
Residual		27.9

Figure 1 (**a**) Non-metric multidimensional scaling (NMDS) of community structure in gastric biopsy samples from the 12 studied patients. Triangles and circles indicate negative and positive *H. pylori* status determined by *glmM* PCR, respectively, and color indicates ethnicity/origin of the subjects (Amerindians in black and non-Amerindians in light gray). (**b**) *P*-value of the variation of the bacterial community as explained by *H. pylori* status and host ethnicity/origin (A = Amerindians; N = Non-Amerindians). The color reproduction of this figure is available on the html full text version of the manuscript.

PhyloChip (translated as relative abundance) than negative patients.

Individual Amerindian patients had 41 ± 3 bacterial phyla in their stomach, whereas the Bangladeshi and Rwandan patients had 32 and 34 phyla, respectively (Table 2). The two non-Amerindians lacked nine phyla that were present in at least one Amerindian subject, namely, Deferribacteres, LD1PA group, NC10, OD1, OP8, SPAM, SR1, Thermotogae and TM6. Two additional phyla, Dictyoglomi and WS5, were absent in the patient from Rwanda, and Fusobacteria was absent in the patient from Bangladesh.

Bacterial community structure

A non-metric multidimensional scaling clustering analysis based on normalized intensities of the hybridized probe-sets of the overall bacterial communities showed that about 28% of the total variance in the gastric microbiota of the 12 subjects was explained by *H. pylori* status (Figure 1). There were significant differences in the microbial community structure between Amerindians and non-Amerindians, and in particular, the Rwandan patient is an outlier, but higher numbers of non-Amerindians would be needed to assess the significance of this finding.

The differences in the gastric communities of patients with positive and negative *H. pylori* status could be explained by relative abundance differences in 152 taxa (analysis of variance, $P \leqslant 0.05$, Benjamini–Hochberg corrected). A heatmap (Figure 2a) shows that the gastric bacterial communities in *H. pylori*-negative patients (triangles) had greater relative abundance of Actinobacteria and Firmicutes (Figure 2b), whereas *H. pylori*-positive subjects had higher abundances of non-*H. pylori* Proteobacteria and Acidobacteria (Figure 2c). These results were confirmed by regression analysis of an *H. pylori* taxon (10534, based on nine *H. pylori* sequences) and other taxa (Supplementary Figure S1). We found that members of Proteobacteria (classes Alpha, Delta and Epsilonproteobacteria), Acidobacteria (class Acidobacteria) and Spirochaetae (class Spirochetes) were cocorrelated with the presence of *H. pylori* taxa, whereas Actinobacteria (class Actinobacteria), Firmicutes (classes Bacilli and Mollicutes), Bacteroidetes (classes Sphingobacteria and Flavobacteria), Chloroflexi (class Anaerolineae), Cyanobacteria (class Cyanobacteria), Fusobacteria (class Fusobacteria), Planctomycetes (class Planctomycetacea), Proteobacteria (classes Beta and Gammaproteobacteria) and Verrucomicrobia (class Verrucomicrobiae) showed an inverse correlation with *H. pylori* (Supplementary Figure S1). Regression analysis involving the other two taxa that contain *H. pylori* (taxa 10442 and 10443) showed similar results (data not shown).

Discussion

This work provides an early view of the microbiota of the human stomach from individuals living in developing countries. Consistent with previous studies (Bik *et al.*, 2006; Andersson *et al.*, 2008), the human gastric bacterial community is very rich but uneven, strongly dominated by only four phyla, namely, Proteobacteria, Firmicutes, Bacteroidetes and Actinobacteria.

Comparisons of PhyloChip results in this study and results from previous reports are not direct owing to the use of different taxonomy systems. The PhyloChip has its own taxonomic schema based on a previous version of the Hugenholtz taxonomy (Greengenes) with 56 bacterial phyla/divisions, compared with the 35 in the Taxonomic Outline of the Bacteria and Archaea (TOBA) taxonomy (http://www.taxonomicoutline.org/). DeSantis *et al.* (2006a, b) has previously highlighted this incongruence among taxonomies.

In all, 20 of the 44 phyla reported here have been reported in the human GI tract before, using the TOBA taxonomy of the Ribosomal Database Project (RDP) (Eckburg *et al.*, 2005; Bik *et al.*, 2006; Gill *et al.*, 2006; Andersson *et al.*, 2008; Dethlefsen *et al.*, 2008; Keijser *et al.*, 2008); 13 (excluding unclassified bacteria) were either absent in the RDP classification or included in a different phylum. For example, RDP classifies as Firmicutes, the Greengenes divisions, Natronoanaerobium, OP9/JS1 and NC10. Greengenes OD1 corresponds to OP11 in RDP, OP8 to Acidobacteria and TM6 to Proteobacteria. After

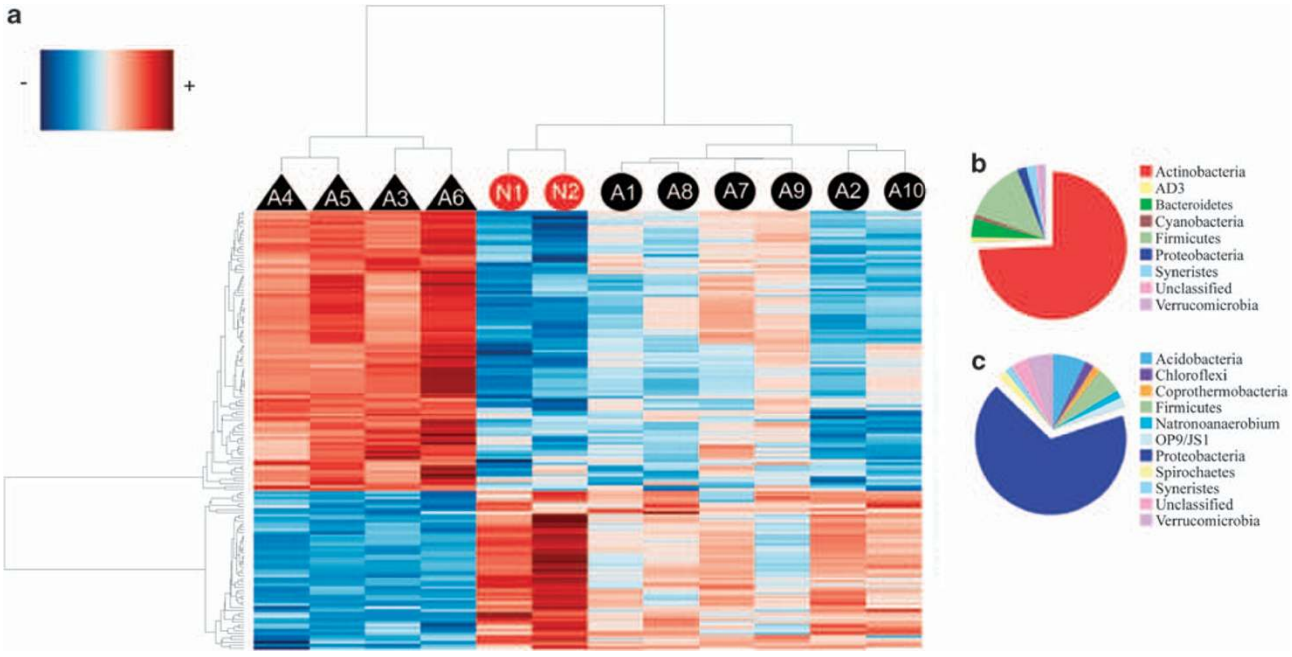


Figure 2 Heatmap with bidirectional clustering, displaying the relationship between 12 patient samples and 152 significantly different taxa. (**a**) Negative or positive *H. pylori-glmM* status is annotated with triangles or circles, respectively. Non-Amerindians are represented in red and Amerindians in black. Bacterial taxa are clustered to the left and patient cluster appears at the top of the heatplot according to their intensity profile similarity. The relative abundance of the three Helicobacter taxa was not included in this analysis. Pie charts depicting phylum level distribution between bacterial taxa that inversely correlated (**b**; $n = 97$) or co-correlated (**c**; $n = 55$) with *H. pylori* positivity.

reclassifying the sequences using RDP taxonomy, only four phyla had not been previously reported in the human GI tract: Thermotogae, Chlorobi, BRC1 and Nitrospira. These four phyla represent the new gastric bacterial diversity found by the PhyloChip.

The presence of *H. pylori* DNA in patients who were negative by PCR detection has also has been reported before (Bik *et al.*, 2006), and might reflect an *H. pylori* load below the sensitivity level of the method used.

Our results suggest remarkable changes in the structure of the gastric bacterial community, based on *H. pylori* status as determined by PCR. Of the gastric bacterial diversities, *H. pylori* is the bacterium considered to be indigenous to the stomach and has coevolved for at least 50 000 years with humans (Linz *et al.*, 2007). Therefore, the interactions between *H. pylori* and the other bacteria detected in the stomach might be indirect, likely mediated by the host response. *H. pylori* presence affects the gastric environment, hormones and immunity (Atherton and Blaser, 2009). The bacterium promotes density-dependent humoral and cellular immune responses (Plebani *et al.*, 1996), which might affect homeostasis of leptin (Nishi *et al.*, 2005; Pacifico *et al.*, 2008), a hormone that modulates immunity and gastric acid secretion and promotes a Th1 response (Faggioni *et al.*, 2001; Perry *et al.*, 2010) The intimate contact of *H. pylori* with epithelial cells (through adhesin molecules such as BabA or through Cag PAI proteins) augments immune responses (Rad *et al.*, 2002), presence of inflammatory cells (Atherton *et al.*, 1997; Bodger and Crabtree, 1998) and cytokines (IL-1b, -2, -6, -8 and TNF-α) (Yamaoka *et al.*, 1997), which also might affect other bacterial species. More research is needed to characterize the physiological differences related to *H. pylori* status, including variation in the gastric microbiota, as well as its clinical implications.

The variation in the gastric microbiota among humans from different origins or ethnicity was even more significant than differences associated with *H. pylori* presence. This trend is consistent with observation of host differences in communities at other body locations (Sepp *et al.*, 1997; Zhou *et al.*, 2007), and stresses the need for conducting larger studies of different human populations.

Acknowledgements

This work was supported by UPR grant FIPI 880314. Part of this work was performed at the Lawrence Berkeley National Laboratory under the auspices of the University of California under contract number DOE DE-AC02-05CH11231 and of the Diane Belfer Program in Human Microbial Ecology. We thank Lyd Marie Rodriguez for technical assistance.

References

Andersson AF, Lindberg M, Jakobsson H, Backhed F, Nyren P, Engstrand L. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS ONE* **3**: e2836.

Atherton JC, Blaser MJ. (2009). Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. *J Clin Invest* **119**: 2475–2487.

Atherton JC, Peek Jr RM, Tham KT, Cover TL, Blaser MJ. (1997). Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. *Gastroenterology* **112**: 92–99.

Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F *et al.* (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci USA* **103**: 732–737.

Bodger K, Crabtree JE. (1998). *Helicobacter pylori* and gastric inflammation. *Br Med Bull* **54**: 139–150.

Brodie EL, Desantis TZ, Joyner DC, Baek SM, Larsen JT, Andersen GL *et al.* (2006). Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl Environ Microbiol* **72**: 6288–6298.

Brodie EL, DeSantis TZ, Parker JP, Zubietta IX, Piceno YM, Andersen GL. (2007). Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci USA* **104**: 299–304.

Clarke KR. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecology* **18**: 117–143.

DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM, Andersen GL. (2007). High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* **53**: 371–383.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006a). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

DeSantis Jr TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM *et al.* (2006b). NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* **34**: W394–W399.

Dethlefsen L, Huse S, Sogin ML, Relman DA. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**: e280.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al.* (2005). Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.

Faggioni R, Feingold KR, Grunfeld C. (2001). Leptin regulation of the immune response and the immunodeficiency of malnutrition. *Faseb J* **15**: 2565–2571.

Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M *et al.* (2003). Traces of human migrations in *Helicobacter pylori* populations. *Science* **299**: 1582–1585.

Farrelly V, Rainey FA, Stackebrandt E. (1995). Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl Environ Microbiol* **61**: 2798–2801.

Freire GN. (2007). Indigenous Shifting Cultivation and the New Amazonia: A Piaroa Example of Economic Articulation. *Human Ecology* **35**: 681–696.

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS *et al.* (2006). Metagenomic analysis of the human distal gut microbiome. *Science* **312**: 1355–1359.

Hugenholtz P. (2002). Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3**: REVIEWS0003.

Jari Oksanen RK, Pierre Legendre, Bob O'Hara, Simpson GL, Peter Solymos, M. Henry, H. Stevens, Helene Wagner. (2008). http://cran.r-project.org/, http://vegan.r-forge.r-project.org/.

Keijser BJ, Zaura E, Huse SM, van der Vossen JM, Schuren FH, Montijn RC *et al.* (2008). Pyrosequencing analysis of the oral microflora of healthy adults. *J Dent Res* **87**: 1016–1020.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2009). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**: 118–123.

Lane DJ. (1991). *16S/23S rRNA sequencing*. Wiley: London, 115–175 pp.

Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P *et al.* (2007). An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**: 915–918.

Lu JJ, Perng CL, Shyu RY, Chen CH, Lou Q, Chong SK *et al.* (1999). Comparison of five PCR methods for detection of *Helicobacter pylori* DNA in gastric tissues. *J Clin Microbiol* **37**: 772–774.

Masuda N, Church GM. (2002). *Escherichia coli* gene expression responsive to levels of the response regulator EvgA. *J Bacteriol* **184**: 6225–6234.

Nishi Y, Isomoto H, Uotani S, Wen CY, Shikuwa S, Ohnita K *et al.* (2005). Enhanced production of leptin in gastric fundic mucosa with *Helicobacter pylori* infection. *World J Gastroenterol* **11**: 695–699.

Pacifico L, Anania C, Osborn JF, Ferrara E, Schiavo E, Bonamico M *et al.* (2008). Long-term effects of *Helicobacter pylori* eradication on circulating ghrelin and leptin concentrations and body composition in prepubertal children. *Eur J Endocrinol* **158**: 323–332.

Perry S, de Jong BC, Solnick JV, Sanchez M, Yang S, Lin PL *et al.* (2010). Infection with *Helicobacter pylori* is associated with protection against tuberculosis. *PLoS One* **5**: e8804.

Plebani M, Basso D, Cassaro M, Brigato L, Scrigner M, Toma A *et al.* (1996). *Helicobacter pylori* serology in patients with chronic gastritis. *Am J Gastroenterol* **91**: 954–958.

Polz MF, Cavanaugh CM. (1998). Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**: 3724–3730.

Rad R, Gerhard M, Lang R, Schoniger M, Rosch T, Schepp W *et al.* (2002). The *Helicobacter pylori* blood group antigen-binding adhesin facilitates bacterial colonization and augments a nonspecific immune response. *J Immunol* **168**: 3033–3041.

Rproject. (2008). *A language and environment for statistical computing. R. Development Core Team*. R Foundation for Statistical Computing, 1090 Vienna, Austria. URL http://www.R-project.org.

Sepp E, Julge K, Vasar M, Naaber P, Bjorksten B, Mikelsaar M. (1997). Intestinal microflora of Estonian and Swedish infants. *Acta Paediatr* **86**: 956–961.

Suzuki MT, Giovannoni SJ. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625–630.

Yamaoka Y, Kita M, Kodama T, Sawai N, Kashima K, Imanishi J. (1997). Induction of various cytokines and development of severe mucosal inflammation by *cagA* gene positive *Helicobacter pylori* strains. *Gut* **41**: 442–451.

Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P *et al.* (2007). Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. *ISME J* **1**: 121–133.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)