Molecular analysis of the bacterial microbiota in the human stomach

Elisabeth M. Bik*†‡, Paul B. Eckburg*†§, Steven R. Gill¶, Karen E. Nelson¶, Elizabeth A. Purdom‖, Fritz Francois**††, Guillermo Perez-Perez**††, Martin J. Blaser**††, and David A. Relman*†‡§

*Department of Microbiology and Immunology, Stanford University School of Medicine, Fairchild Science Building, 299 Campus Drive, Stanford, CA 94305; †Veterans Affairs Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304; §Department of Medicine, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305; ¶The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850; ‖Department of Statistics, Sequoia Hall, 390 Serra Mall, Stanford University, Stanford, CA 94305; and Departments of **Medicine and ††Microbiology, New York University School of Medicine, New York, NY 10016

The microbiota of the human stomach and the influence of *Helicobacter pylori* colonization on its composition remain largely unknown. We characterized bacterial diversity within the human gastric mucosa by using a small subunit 16S rDNA clone library approach and analyzed 1,833 sequences generated by broad-range bacterial PCR from 23 gastric endoscopic biopsy samples. A diverse community of 128 phylotypes was identified, featuring diversity at this site greater than previously described. The majority of sequences were assigned to the *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, and *Fusobacteria* phyla. Ten percent of the phylotypes were previously uncharacterized, including a *Deinococcus*-related organism, relatives of which have been found in extreme environments but not reported before in humans. The gastric clone libraries from 19 subjects contained *H. pylori* rDNA; however, only 12 of these subjects tested positive for *H. pylori* by conventional laboratory methods. Statistical analysis revealed a large degree of intersubject variability of the gastric ecosystem. The presence of *H. pylori* did not affect the composition of the gastric community. This gastric bacterial rDNA data set was significantly different from sequence collections of the human mouth and esophagus described in other studies, indicating that the human stomach may be home to a distinct microbial ecosystem. The gastric microbiota may play important, as-yet-undiscovered roles in human health and disease.

16S ribosomal rRNA clone library | *Helicobacter pylori* | microbial diversity | human indigenous microbiota

Helicobacter pylori is part of the gastric biota in a significant portion of the human population, and its role in the development of gastritis, peptic ulcer disease, and adenocarcinoma is well recognized (1). However, little is known about other members of the human gastric microbial ecosystem in disease and health. Characterization of the gastric microbiota has traditionally relied on cultivation of gastric juice or mucosal biopsies, and such studies have identified several members of the *Firmicutes*, *Proteobacteria*, *Actinobacteria*, and *Fusobacteria* phyla, as well as yeasts in relatively low abundance (2, 3). The relative paucity of cultivated bacteria and the presence of a harsh local environment have led to assumptions that the human stomach does not harbor a complex microbial biota. Molecular methods have the potential to offer more detailed insights into complex microbial communities (4–7). Analysis of PCR-amplified, small subunit rRNA genes (16S rDNA) is one such approach, because of the universal distribution of these genes among prokaryotes, the presence of conserved and variable domains within the gene, and its reliability for inferring phylogenetic relationships (6, 8, 9). This method has been used to characterize the microflora of the human oral cavity (10–13), esophagus (14), colon (7, 15–17), and stool (7, 15, 18), and it has revealed many novel, uncultivated bacterial species.

The human gastric microbiota has not been extensively characterized by using 16S rDNA sequence analysis. Most PCR-based studies of the gastric biota have focused on specific genera, e.g., *Helicobacter* (19) and *Lactobacilli* (20). In 2000, a combination of broad-range bacterial PCR and temporal temperature gradient gel electrophoresis suggested the presence of a complex gastric mucosal biota (21). However, this study relied on a small number of short (200 bp) 16S rDNA fragments and lacked the resolution to define sequence differences at the species level (only 11 genera were described). It remains unclear whether the gastric mucosal microbiota varies between different anatomic sites (fundus, corpus, antrum, and pylorus), is distinct from the oral and esophageal microbiota, and whether the presence of *H. pylori* influences the composition of the gastric microbial community. To address these issues, we characterized the gastric microbiota of *H. pylori*-positive and -negative subjects by using broad-range PCR and 16S rDNA sequence analysis.

Results

Phylogenetic Analysis. Fig. 1 shows a phylogenetic tree with the 128 phylotypes identified among the 1,833 clones in the combined gastric sequence data set. The gastric bacterial community was dominated by five major phyla: *Proteobacteria* (952 clones, 32 phylotypes, and 3 previously uncharacterized phylotypes), *Firmicutes* (low G+C Gram-positive bacteria) (464, 36, 2), *Bacteroidetes* (193, 35, 5), *Actinobacteria* (high G+C Gram-positive bacteria) (164, 12, 1), and *Fusobacteria* (56, 10, 1). The remaining gastric 16S rRNA sequences could be assigned to the TM7 (2, 1, 0), *Deferribacteres* (1, 1, 0), and *Deinococcus/Thermus* phyla (1, 1, 1). Thirteen (10%) phylotypes (24 clones) were previously uncharacterized (Table 1). Among these phylotypes, one clone had 93.6% sequence identity to a *Deinococcus* isolate from an environmental source. *H. pylori* was the only member of the genus *Helicobacter* found in these human stomach samples and was the most abundant phylotype within the libraries. *H. pylori* sequences were found in 19 of the 23 gastric samples (varying between 1% and 99% of clones per sample library) and constituted 42% of all sequences analyzed (777 clones). Only 12 of the 19 subjects in whom *H. pylori* sequences were detected tested positive for *H. pylori* by using conventional tests. The average percentage of *H. pylori* clones in the 12 corresponding samples was 72%, compared to an average of 11% among the 7 samples from *H. pylori*-negative subjects (Table 2). All four subjects whose gastric libraries contained no *H. pylori* clones
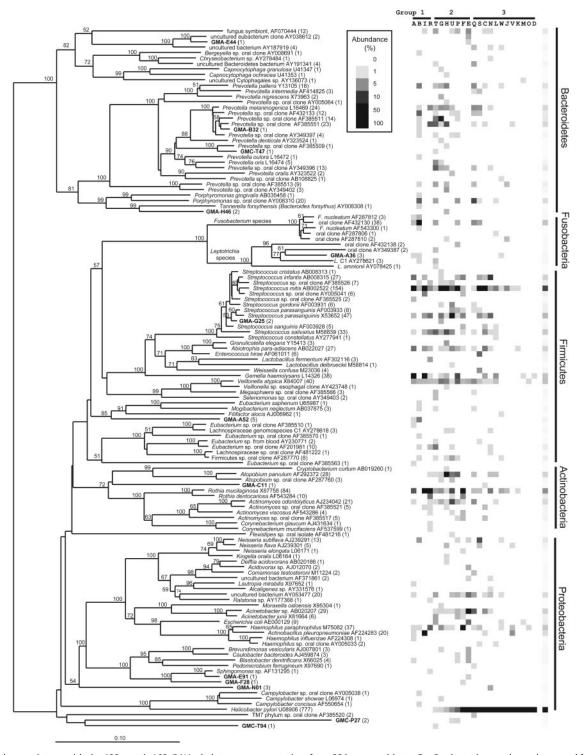
Fig. 1. Phylogenetic tree with the 128 gastric 16S rDNA phylotype representatives from 23 human subjects. GenBank entries are shown in normal font; names of previously uncharacterized phylotype representatives (<99% sequence identity to published sequences) are shown in bold. Numbers of clones within each phylotype are shown in parentheses. The tree was constructed by neighbor-joining analysis by using an Olsen correction. Bootstrap values >50 (expressed as percentages of 100 replications) are shown at branch points. The scale bar represents evolutionary distance (10 substitutions per 100 nucleotides). The right side of the figure shows the relative abundance of phylotypes per gastric specimen in gray values (white, 0% present; black, 100% of clone library). Letters above the abundance graph correspond to subjects A–W in Table 3, which is published as supporting information on the PNAS web site. Subjects are grouped according to *H. pylori* status as determined by conventional and molecular tests, as indicated in Table 2, in increasing order of percentage of *H. pylori* clones.

were considered to be negative for *H. pylori* by a combination of conventional methods (Table 3). The second and third most abundant genera found in the overall combined gastric library were *Streptococcus* (299 clones) and *Prevotella* (139 clones).

Species Richness and Diversity. Estimations of species coverage, richness, evenness, and diversity were calculated for the combined data set, as well as for three subsets of samples defined according to *H. pylori* status of the host, as determined by

Table 1. Previously uncharacterized phylotypes found in gastric clone libraries

	Gastric sequence		Closest neighbor present in public database	
Clone name	Assigned accession no.	Sequence similarity, %	Accession no.	Description
GMC-T94	AY582897	93.6	AJ549111	*Deinococcus indicus* from arsenic-contaminated water from aquifer
GMC-P27, GMC-U27	AY582895	94.2	AF173818	Uncultured Antarctic bacterium from permanent Antarctic lake ice
GMA-C11	AY582888	95.5	AY043855	Uncultured actinobacterium isolated from forest mineral soil
GMA-N01/N26/N40	AY582894	95.3	AY494619	Uncultured delta-proteobacterium from salmonid gill
GMC-T47	AY582896	97.2	AF385509	*Prevotella* sp. from tongue dorsa
GMA-H46/H63	AY582893	97.3	AJ318110	Uncultured Bacteroidetes from waste-gas biofilter
GMA-E44	AY582889	97.4	AY038612	Uncultured eubacterium from lacustrine subsurface sediments
GMA-E91	AY582890	98.2	AY162043	Uncultured alpha proteobacterium from soil
GMA-A52, GMA-B11/B19/B27/B81	AY582886	98.6	AY207059	*Peptostreptococcus* sp. from human mouth
GMA-A36 GMA-B65, GMC-W25	AY582885	98.7	AF385518	*Leptotrichia* sp. from tongue dorsa
GMA-F28	AY582891	98.8	AF131297	*Sphingomonas aquatilis*
GMA-B32	AY582887	98.9	L16469	*Prevotella melaninogenica*
GMA-G25/G61	AY582892	98.9	AF432137	*Streptococcus* sp. from tongue dorsa

Previously uncharacterized phylotypes were defined as sequences or groups of sequences having <99% sequence similarity to sequences present in public databases. The NCBI GenBank accession number and a short description of the closest neighbor is given, as well as sequence similarity (%) to that neighbor. Clone designations indicate site, patient, and clone number, e.g., GMA-E91 represents gastric mucosal biopsy from antrum, patient E, clone 91. GMC, gastric mucosal biopsy from corpus. From each phylotype, one representative sequence (underlined in the case of multiple clones) was deposited into the NCBI GenBank.

conventional techniques and according to the presence of *H. pylori* clones in the sequence library (Table 2). Good's coverage was 97% for the overall sequence set, indicating that three additional phylotypes would be expected for every 100 additional sequenced clones. This level of coverage indicated that the 16S rDNA sequences identified in these samples represent the majority of bacterial sequences present in the human stomach mucosa samples under study. Rarefaction curves were calculated for the overall combined data set by using the individual-based Coleman method and the sample-based Mao Tau method (Fig. 2); a nonrandomized phylotype abundance curve is provided in Fig. 4, which is published as supporting information on the PNAS web site. The Chao1 estimator of total species richness was 193; however, because it did not plateau with the current sequencing effort, it is likely to be a minimal estimate. Conventional diversity and evenness indices were lowest for samples that were positive for *H. pylori* by conventional methods (Table 2, group 3).

Influence of *H. pylori* and Anatomical Location on the Composition of the Gastric Microflora. To investigate the possible impact of *H. pylori* on the presence of other bacterial species, we first examined the phylogenetic tree and phylotype abundance plot (Fig. 1). Subjects in group 3 demonstrated a relative lack of non-*Proteobacteria* phylotypes, especially *Bacteroidetes* phylotypes, compared with *H. pylori*-negative subjects. This difference could largely be attributed to the relative low number of non-*H. pylori* clones among these libraries. When *H. pylori* sequences were left out of the analysis, phylotype evenness and diversity among libraries in which *H. pylori* clones were found (groups 2 and 3) were higher than that of libraries from *H. pylori*-negative subjects (group 1) (Table 2). Evaluation of the phylum distribution of all non-*H. pylori* phylotypes from each of the three groups of samples and subjects based on *H. pylori* status (Fig. 3) revealed no gross differences in the taxonomic patterns.

Subject and clone library variability were explored with double

Table 2. Sequence diversity and library coverage estimations

Measurement	Group 1 (n = 4)	Group 2 (n = 7)	Group 3 (n = 12)	Combined (n = 23)
H. pylori status (conventional testing)	(−)	(−)	(+)	
No. of clones	310	529 (471)	994 (275)	1833
H. pylori clones, %	0	11.0	72.3	42.4
Phylotypes	57	86 (85)	60 (59)	128
Singletons	29	27	20	49
Chao1 estimator of species richness (OTUs)	138.8	105.5	81.2	193.4
Chao1 standard deviation	66.3	10.8	15.1	27.8
Shannon's index for diversity	3.1	3.8 (3.8)	1.6 (3.6)	2.9
Simpson's index for diversity	12.9	26.6 (30.8)	1.9 (23.4)	5.2
Evenness	0.41	0.49 (0.54)	0.08 (0.60)	0.14
Good's estimator of coverage, %	90.6	94.9	98.0	97.3

Gastric biopsy specimens were divided into three groups. Group 1, samples without *H. pylori* sequences (all were from subjects negative for *H. pylori* by conventional testing, i.e., subjects A, B, I, and R). Group 2, samples containing *H. pylori* sequences from subjects negative for *H. pylori* by conventional testing (E, F, G, H, P, T, and U). Group 3, samples containing *H. pylori* sequences from subjects positive for *H. pylori* by conventional testing (C, D, J, K, L, M, N, O, Q, S, V, and W). Diversity and richness estimators were calculated with ESTIMATES. Numbers in parentheses indicate data when *H. pylori* sequences were left out of the analysis. Number of subjects in each group is in parentheses under each column label.
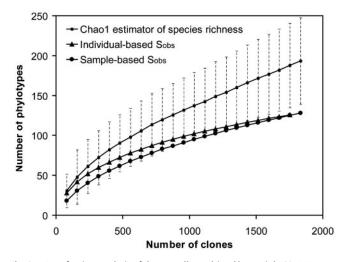
Fig. 2. Rarefaction analysis of the overall, combined bacterial 16S rRNA gene clone library recovered from human gastric biopsy specimens. The rarefaction curve, plotting the number of observed phylotypes (S_{obs}) as a function of the number of clones, was computed by ESTIMATES. The corrected Chao1 estimator for species richness (after 1,000 randomizations) was plotted as well. The dotted lines indicate the 95% confidence interval.

principal coordinate analysis (DPCoA) and redundancy analysis of non-*H. pylori* sequences. No significant association between phylotype distribution and *H. pylori* status ($P > 0.5$) or anatomical location ($P > 0.5$) was found. In addition, phylotype distribution was not significantly related to gastric pH levels ($P = 0.47$). Hierarchical clustering of the data did not reveal distinct clusters between *H. pylori*-negative and -positive subjects (Fig. 5, which is published as supporting information on the PNAS web site).

Pairwise comparisons of each gastric clone library to every other library by using ∫-LIBSHUFF revealed that the majority of
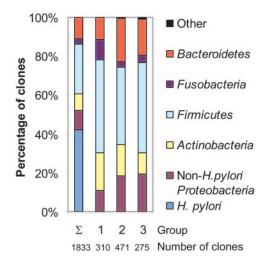

Fig. 3. Relative phylotype frequencies of clones isolated from gastric specimens. Sequences were assigned to a bacterial phylum according to their position in the phylogenetic tree in Fig. 1. Σ, combined libraries from all 23 subjects, including *H. pylori* sequences (1,833 sequences). Gastric libraries were divided into three groups; members are defined in Fig. 1 and Table 3. Groups: 1, clone libraries not containing *H. pylori* rDNA sequences from four subjects negative for *H. pylori* by conventional testing; 2, clone libraries containing *H. pylori* sequences from seven subjects negative for *H. pylori* by conventional testing; 3, clone libraries containing *H. pylori* sequences from 12 subjects positive for *H. pylori* by conventional testing. For groups 1–3, only non-*H. pylori* sequences are shown.

the libraries were distinct from any other (Table 4, which is published as supporting information on the PNAS web site). As predicted by the DPCoA results, no pattern of library relatedness between subject groups was evident in this analysis. However, the combined gastric sequence library (with or without *H. pylori* clones) was distinct from libraries derived from the subgingival crevice and esophageal mucosa obtained from other studies ($P < 0.00005$) (P. W. Lepp, K. Palm, and K. Lightfield, unpublished results; ref. 14). Caution must be taken in interpreting the latter ∫-LIBSHUFF results, given that the oral and esophageal libraries were not derived from the same individuals, the subjects in these studies had different clinical syndromes, and the clone libraries were constructed by using different PCR and sequencing protocols.

Discussion

Traditionally, the human stomach has been viewed as an inhospitable environment for microorganisms because of acidic conditions and other antimicrobial factors. With the discovery of *H. pylori* and other gastric helicobacters, and subsequent insight into the mechanisms by which these organisms adapt to the gastric environment (22), the existence of a bacterial community adapted to this human niche became more plausible. Despite the potential importance of a gastric microbial community in human health and disease, few data on the complexity of this biota have been available.

The current study used sensitive molecular methods to reveal previously uncharacterized features of the gastric bacterial biota among adults with symptomatic upper gastrointestinal disease. Interestingly, *H. pylori* was the most abundant phylotype in the human stomach in subjects who tested positive for this organism by using conventional clinical approaches. However, *H. pylori* DNA also was present in the biopsy samples of seven subjects that were considered to be *H. pylori*-negative by conventional methods. It is possible that some of our biopsies became cross-contaminated with *H. pylori* DNA during tissue processing, although abundant extraction controls did not yield visible amounts of DNA on gel electrophoresis after broad-range PCR. An alternative explanation is that the number of *H. pylori* bacteria in these samples (group 2) was too low to be detected by conventional methods and too few to elicit host antibody production. This hypothesis is supported by the low percentage of *H. pylori* clones in the libraries derived from samples and subjects that were negative by using conventional *H. pylori* tests. Detection of bacterial DNA does not necessarily indicate the presence of live, resident organisms. Bacterial DNA in stomach biopsy samples might reflect the presence of bacterial cell remnants or transient bacteria (3).

We found that the human gastric environment contains ribosomal DNA from a wealth of bacteria, in addition to *H. pylori* (Fig. 1). One hundred twenty-eight phylotypes among eight bacterial phyla were detected, suggesting much greater gastric ecosystem diversity than previously described; this higher-order taxon diversity is similar to that described for the lower gastrointestinal tract (15–17). Phylotypes were found that were affiliated with bacterial taxa not yet described with cultivation-based analyses of the gastric flora, including *Caulobacter*, *Actinobacillus*, *Corynebacterium*, *Rothia*, *Gemella*, *Leptotrichia*, *Porphyromonas*, *Capnocytophaga*, TM7, *Flexistipes*, and *Deinococcus*. Among the representative sequences of the 128 phylotypes found in this study, 64 (50%) were derived from uncultivated bacteria. Of these phylotypes, 43 (67%) have been previously described in specimens from the human mouth (10–13). This finding is not unexpected because it is likely that the composition of the gastric community is not only determined by niche-specific factors but also by stochastic colonization from upstream components of the alimentary tract.

Ten percent of the phylotypes found in this study were previously uncharacterized, and some of these sequences display >5% sequence dissimilarity from publicly available sequences (Table 1). Because little is known about their closest relatives, the biological and clinical significance of these putative organisms is unclear. Of the previously uncharacterized phylotypes found in this study, the sequence belonging to the *Deinococcus/Thermus* phylum is particularly interesting. *Deinococcus* species (specifically *D. radiodurans*) have been isolated from extreme environments, such as radioactive waste disposal sites and hot springs, as well as from animal fecal samples (23). To our knowledge, this *Deinococcus*-related sequence is the first identified from a human.

ʃ-LIBSHUFF analysis verified a large degree of intersubject variability, similar to that observed in other 16S rDNA surveys of the gastrointestinal microbiota (15–17). Given the predominance of men in this study, comparison of the gastric community of the only woman with those of the 22 male subjects would not be helpful. Based on visual inspection of the phylogenetic trees (Fig. 1), assessment of higher-order taxa distribution among samples (Fig. 2), and DPCoA results, the presence of *H. pylori*, gastric anatomical location, and gastric pH level did not significantly influence or alter local bacterial community composition. In addition, the presence of *H. pylori* did not have a negative effect on diversity and evenness of the other members of the gastric community (Table 2). Further molecular analyses are needed to understand better the impact of *H. pylori* infection on the gastric mucosal microbial community.

In this study, bacterial sequences in the stomach were not simply a random sampling of bacterial sequences from oral (subgingival) or esophageal communities. Given prior assumptions about limited microbial diversity in the human stomach, and the natural flow of oroesophageal secretions into the stomach, these findings are somewhat surprising and suggest the presence of distinct bacterial communities that have adapted to multiple specific environmental habitats in the stomach. These results must be interpreted with caution, because the libraries were derived from different subjects with differing medical conditions and constructed by using different DNA extraction methods and amplification protocols. Whether the stomach is home to distinct bacterial communities remains to be verified. Further studies should include a more detailed spatial (anatomical) and temporal analysis of gastric microbial community structure within subjects of different ethnic and racial backgrounds, as well as an analysis of the effect of *H. pylori* eradication therapy on community structure. A better understanding of the resident microbial communities at healthy and diseased sites should shed light on the pathogenesis, diagnosis, and treatment of gastrointestinal illnesses.

Materials and Methods

Gastric Biopsy Samples and *H. pylori* Testing. A single gastric mucosal biopsy was obtained from each of 23 adults (22 males, 1 female; 13 Caucasians, 5 Hispanics, and 5 African-Americans) during upper endoscopy at the New York Campus of the Veterans Affairs New York Harbor Healthcare System (Table 3). The use of these subjects was approved by the New York University Institutional Review Board and the Stanford University Administrative Panel on Human Subjects in Medical Research. All participants in the study provided their informed consent. Biopsies were obtained from the corpus of the stomach in 9 subjects and the gastric antrum in 14 subjects. Twelve subjects were determined to be *H. pylori*-positive by a positive result in two or more of the following conventional tests: culture of tissue, rapid urease test on tissue, serum *H. pylori* IgG ELISA, and histopathology (details in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site).

Construction of 16S rDNA Clone Libraries. Total genomic DNA was extracted from the biopsy samples by using a combination of the QIAamp DNA isolation kit (Qiagen, Valencia, CA) and a bead-beater method. Universal bacterial PCR primers were used to amplify the region corresponding to positions 8 to 806 of the *Escherichia coli* 16S rRNA gene by using a 25-cycle PCR. Experimental details are provided in *Supporting Materials and Methods*.

Amplification products were gel-purified, ligated into the pCR2.1 vector, and transformed into *E. coli* TOP10 cells by using the TOPO-TA cloning kit (Invitrogen) according to the manufacturer's instructions. From each gastric biopsy, 95 transformants were selected. Plasmid template DNA from each transformant was prepared by a modified alkaline lysis method. Sequencing reactions were carried out with BigDye terminator and run on ABI 3730xl sequencers (Applied Biosystems, Foster City, CA).

Phylogenetic Analysis. A total of 1,833 nonchimeric 16S rDNA sequences were aligned and analyzed by using the ARB software package (24); phylogenetic analysis details are available in *Supporting Materials and Methods*. A neighbor-joining tree containing all sequences was constructed by using Olsen correction and 738 nonambiguous filter positions. Phylotypes were defined as single clones or groups of sequences having 99% or more identity to a published sequence, or, in the case of a previously uncharacterized phylotype, <99% identity to any public database sequence. The reliability of phylogenetic tree topology was tested by bootstrap resampling with 100 replicates.

Estimation of Microbial Diversity. Rarefaction curves, richness estimations, and diversity indices were determined with ESTIMATES (Version 7, http://viceroy.eeb.uconn.edu/estimates). The species rarefaction curve of the entire data set was computed by using the individual-based Coleman method and the sample-based Mao Tau method. The bias-corrected Chao1 estimator of species richness was calculated after 1,000 randomizations of sampling without replacement. The percentage of coverage was calculated by Good's method with the formula $[1 - (n/N)] \times 100$, where n is the number of phylotypes in a sample represented by one clone (singletons) and N is the total number of sequences in that sample (25). Diversity of the sampled sequence set was estimated by using the Simpson and Shannon indices within the ESTIMATES application. The Shannon index of evenness was calculated with the formula $E = e^D/N$, where D is the Shannon diversity index.

Statistical Analysis. DPCoA was implemented to investigate the relationships between phylotype dissimilarities and Rao diversity data among the gastric communities (26) by using the ade4 and dpcoa functions within the R statistical package (www.r-project.org). Redundancy analysis (1,000 permutations) applied to the results of DPCoA was used to evaluate the statistical significance of categorical explanatory variables, such as positive or negative *H. pylori* status (27). Cluster analysis methods were used to analyze the relationships between sequence phylotypes as detected among each subject. More details are provided in *Supporting Materials and Methods*.

To determine whether differences in clone libraries were due to underlying variability in the microbial communities versus artifacts of subsampling, gastric clone libraries were compared with the ʃ-LIBSHUFF program (28, 29). This tool compares homologous and heterologous coverage curves by using the integral form of the Cramér-von Mises statistic and performs multiple pairwise comparisons among a set of libraries (details in *Supporting Materials and Methods*). ʃ-LIBSHUFF was used to compare individual clone libraries and combined clone libraries from the stomach, the oral cavity, and esophagus. Regarding the

latter, we compared our pooled gastric library to 1,563 subgingival 16S rDNA sequences derived from subjects with and without gingivitis or periodontitis (P. W. Lepp, K. Palm, and K. Lightfield, unpublished results) and 896 esophageal 16S rDNA sequences from subjects with normal esophageal histology (14).

1. Peek, R. M., Jr. & Blaser, M. J. (2002) *Nat. Rev. Cancer* **2,** 28–37.
2. Adamsson, I., Nord, C. E., Lundquist, P., Sjostedt, S. & Edlund, C. (1999) *J. Antimicrob. Chemother.* **44,** 629–640.
3. Savage, D. C. (1977) *Annu. Rev. Microbiol.* **31,** 107–133.
4. Amann, R. I., Ludwig, W. & Schleifer, K. H. (1995) *Microbiol. Rev.* **59,** 143–169.
5. Hugenholtz, P., Goebel, B. M. & Pace, N. R. (1998) *J. Bacteriol.* **180,** 4765–4774.
6. Pace, N. R. (1997) *Science* **276,** 734–740.
7. Zoetendal, E. G., Collier, C. T., Koike, S., Mackie, R. I. & Gaskins, H. R. (2004) *J. Nutr.* **134,** 465–472.
8. Wilson, K. H., Blitchington, R. B. & Greene, R. C. (1990) *J. Clin. Microbiol.* **28,** 1942–1946.
9. Woese, C. R. (1987) *Microbiol. Rev.* **51,** 221–271.
10. Kazor, C. E., Mitchell, P. M., Lee, A. M., Stokes, L. N., Loesche, W. J., Dewhirst, F. E. & Paster, B. J. (2003) *J. Clin. Microbiol.* **41,** 558–563.
11. Kroes, I., Lepp, P. W. & Relman, D. A. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 14547–14552.
12. Munson, M. A., Banerjee, A., Watson, T. F. & Wade, W. G. (2004) *J. Clin. Microbiol.* **42,** 3023–3029.
13. Paster, B. J., Boches, S. K., Galvin, J. L., Ericson, R. E., Lau, C. N., Levanos, V. A., Sahasrabudhe, A. & Dewhirst, F. E. (2001) *J. Bacteriol.* **183,** 3770–3783.
14. Pei, Z., Bini, E. J., Yang, L., Zhou, M., Francois, F. & Blaser, M. J. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 4250–4255.
15. Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005) *Science* **308,** 1635–1638.
16. Hold, G. L., Pryde, S. E., Russell, V. J., Furrie, E. & Flint, H. J. (2002) *FEMS Microbiol. Ecol.* **39,** 33–39.
17. Wang, X., Heazlewood, S. P., Krause, D. O. & Florin, T. H. (2003) *J. Appl. Microbiol.* **95,** 508–520.
18. Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D. & Dore, J. (1999) *Appl. Environ. Microbiol.* **65,** 4799–4807.
19. Pena, J. A., McNeil, K., Fox, J. G. & Versalovic, J. (2002) *J. Clin. Microbiol.* **40,** 1511–1513.
20. Roos, S., Engstrand, L. & Jonsson, H. (2005) *Int. J. Syst. Evol. Microbiol.* **55,** 77–82.
21. Monstein, H. J., Tiveljung, A., Kraft, C. H., Borch, K. & Jonasson, J. (2000) *J. Med. Microbiol.* **49,** 817–822.
22. Merrell, D. S., Goodrich, M. L., Otto, G., Tompkins, L. S. & Falkow, S. (2003) *Infect. Immun.* **71,** 3529–3539.
23. Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V. & Daly, M. J. (2001) *Microbiol. Mol. Biol. Rev.* **65,** 44–79.
24. Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., *et al.* (2004) *Nucleic Acids Res.* **32,** 1363–1371.
25. Good, I. J. (1953) *Biometrika* **40,** 237–264.
26. Pavoine, S., Dufour, A. B. & Chessel, D. (2004) *J. Theor. Biol.* **228,** 523–537.
27. Legendre, P. & Legendre, L. (1998) *Numerical Ecology* (Elsevier, Amsterdam).
28. Schloss, P. D., Larget, B. R. & Handelsman, J. (2004) *Appl. Environ. Microbiol.* **70,** 5485–5492.
29. Singleton, D. R., Furlong, M. A., Rathbun, S. L. & Whitman, W. B. (2001) *Appl. Environ. Microbiol.* **67,** 4374–4376.

MICROBIOLOGY