Phylogeographic Origin of *Helicobacter pylori* Is a Determinant of Gastric Cancer Risk

Thibaut de Sablet[1,4,*], **M. Blanca Piazuelo**[1,*], **Carrie L. Shaffer**[2,*], **Barbara G. Schneider**[1], **Mohammad Asim**[1,4], **Rupesh Chaturvedi**[1,4], **Luis E. Bravo**[6], **Liviu A. Sicinschi**[1], **Alberto G. Delgado**[1], **Robertino M. Mera**[1], **Dawn A. Israel**[1], **Judith Romero-Gallo**[1], **Richard M. Peek Jr.**[1,3,4], **Timothy L. Cover**[2,4,5], **Pelayo Correa**[1], and **Keith T. Wilson**[1,3,4]

[1]Division of Gastroenterology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, USA

[2]Department of Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, USA

[3]Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, USA

[4]Veterans Affairs Tennessee Valley Healthcare System, Nashville, USA

[5]Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, USA

[6]Department of Pathology, Universidad del Valle School of Medicine, Cali, Colombia

Abstract

Background and Aims—*Helicobacter pylori* colonises the stomach in half of all humans, and is the principal cause of gastric cancer, the second leading cause of cancer death worldwide. While gastric cancer rates correlate with *H. pylori* prevalence in some areas, there are regions where infection is nearly universal, but rates of gastric cancer are low. In the case of Colombia, there is a 25-fold increase in gastric cancer rate in the Andean mountain (high risk) region compared to the coastal (low risk) region, despite similarly high (~90%) *H. pylori* prevalence in the two locations. Our aim was to investigate the ancestral origin of *H. pylori* strains isolated from subjects in these high and low risk regions and to determine whether this is a predictive determinant of precancerous lesions.

Methods—Multi-locus sequence typing was used to investigate phylogeographic origins of infecting *H. pylori* strains isolated from subjects in the Pacific coast and Andean mountains in the state of Nariño, Colombia. We analysed 64 subjects infected with *cagA*[+] *vacA* s1m1 strains. Gastric biopsy slides from each individual were scored for histologic lesions and evaluated for DNA damage by immunohistochemistry.

Correspondence to: Professor Keith T. Wilson, Division of Gastroenterology, Hepatology, and Nutrition Vanderbilt University School of Medicine MRBIV, Room 1030C 2215 Garland Avenue Nashville, TN USA 37232 Phone: 615-343-5675, Fax: 615-343-6229 keith.wilson@vanderbilt.edu.
*These authors contributed equally.

Results—We show that strains from the high risk region were all of European phylogeographic origin, whereas those from the low risk region were of either European (34%) or African origin (66%). European strain origin was strongly predictive of increased premalignant histologic lesions and epithelial DNA damage, even in the low risk region; African strain origin was associated with reduced severity of these parameters.

Conclusion—The phylogeographic origin of *H. pylori* strains provides an explanation for geographic differences in cancer risk deriving from this infection.

Keywords

Gastric Cancer; *Helicobacter pylori*; multi-locus sequence typing; DNA damage; gastritis; intestinal metaplasia; gastric atrophy

INTRODUCTION

Helicobacter pylori, a bacterium infecting half of the world's population, is a causative agent of gastric and duodenal ulcers, and gastric cancer [1,2], and is classified as a Class I carcinogen [3]. The precursor lesions leading to development of gastric carcinoma are sequential from chronic active non-atrophic gastritis to multifocal atrophic gastritis, intestinal metaplasia, and dysplasia [4,5]. The outcome of the infection is determined by features of the pathogen in combination with both environmental and host factors [6]. For example, *cagA*[+] and *vacA* s1m1 genotypes have been correlated with an increased risk for gastric cancer, which increases further in some populations in the presence of certain host gene polymorphisms [7–9]. An important feature of the infection is the induction of oxidative DNA damage in gastric epithelial cells, which has been strongly linked to carcinogenesis [10,11]. Gastric cancer rates correlate with *H. pylori* prevalence rates in some areas of the world, but this relationship is not consistently observed [12]. An "African Enigma" has been described in which nearly 100% of the population in Africa is infected, but the rates of ulcers and cancer are quite low [13]. In Latin America, geographic differences in gastric cancer rates have been documented, despite similarly high *H. pylori* prevalence in different regions [14]. This has been well described in Colombia, where in the state of Nariño, inhabitants of the high altitude Andes mountains have very high incidence rates of gastric cancer estimated to be 150/100,000 people per year compared to 6/100,000 inhabitants of the Pacific coast, despite the fact that the two locations are only 200 km apart [15,16]. Furthermore, the prevalence of advanced gastric precancerous lesions is higher in persons from the Colombian mountain region than in those from the coast [17].

There are striking ethnic differences between the mountain region with a high risk (HR) of gastric cancer and the coastal region with a low risk (LR) of gastric cancer. People living in the mountains are mostly "Mestizos" of Amerindian descent with European admixture, dating from the Spanish colonisation period, whereas people living in the coastal region are predominantly "Mulattos" of mixed African and European ancestry [18]. Although strain-specific variations in *H. pylori* genetic content have been implicated in gastric carcinogenesis [7–9], the prevalence of *cagA*[+] *vacA* s1m1 *H. pylori* strains in the LR region of Colombia is only slightly (13%) lower than in the HR region [17], and this difference in strain characteristics is not sufficient to account for a 25-fold difference in cancer rate. We hypothesised that other bacterial factors are determinants of gastric cancer risk. To determine whether there are fundamental differences in *H. pylori* isolates from the LR versus HR regions, we analysed the ancestral haplotypes of these strains.

MATERIALS AND METHODS

Human Subjects

Males with dyspeptic symptoms who were 39–60 years of age and underwent upper gastrointestinal tract endoscopy, were consecutively recruited in 2006 in two public hospitals from two areas with contrasting gastric cancer risks in the State of Nariño, Colombia. The selection of male subjects for this study was based on findings that rates of gastric cancer are two-fold greater in males compared to females worldwide and in Colombia specifically (http://globocan.iarc.fr), and previous data that this difference occurs in Nariño [19]. The high risk (HR) Andean mountain region included the city of Túquerres and the low risk (LR) region on the Pacific coast was based in the city of Tumaco. People from the LR coastal region are of Mulatto ethnicity and people from the HR Andean region are Mestizos [18,20,21]. Data regarding diet in subjects from these two regions have been previously reported [20], and current smoking status was recorded. Exclusion criteria were previous gastrectomy, serious chronic diseases, or ingestion of any of the following in the 4 weeks prior to the endoscopic procedure: H_2 receptor antagonists, proton pump inhibitors, or antimicrobials. Endoscopies were performed by the same experienced gastroenterologist in both enrollment sites. Gastric mucosal biopsy samples were obtained from the antrum, incisura angularis, and corpus, and embedded in paraffin for histology. Two additional biopsies, one from antrum and one from corpus were immediately frozen in glycerol and thioglycolate for *H. pylori* culture. The samples were shipped on dry ice to Vanderbilt University (Nashville, TN, USA) for analysis.

From 89 subjects that were enrolled, there were 81 subjects from whom *H. pylori* could be isolated that were included in this study. Forty-two individuals (51.3%) were residents of Túquerres, in the Andes Mountains, where the gastric cancer incidence is high, and 39 (48.7%) were residents of Tumaco, on the Pacific coast, where the gastric cancer incidence is low [15,16]. Pathologists who scored the biopsies were blinded regarding geographic origins of the subjects, as well as *H. pylori* genotype and phylogeny. Ethics committees of the two local hospitals and of the Universidad del Valle in Cali, Colombia, as well as the Institutional Review Board at Vanderbilt University approved the protocol for this study. All participants provided informed consent.

Isolation and culture of *H. pylori* strains

Biopsy specimens from antrum and corpus were each homogenised under sterile conditions in 100 μL of sterile phosphate-buffered saline (PBS, pH 7.4) using a homogeniser (Kimble-Kontes). Homogenates were plated onto selective Trypticase soy agar (TSA) with 5% sheep blood, containing vancomycin (20 mg/L), bacitracin (200 mg/L), nalidixic acid (10 mg/L) and amphotericin B (2 mg/L) (Sigma-Aldrich), and 1:10 dilutions were plated on TSA plates (BBL; LABSCO) with no antibiotics. Agar plates were incubated under microaerobic conditions (6% O_2, 6% CO_2 and 88% N_2; Campy Pak Plus envelope, BBL) at 37°C for 4 to 8 days until small grey translucent colonies appeared. Gram stains and assays for oxidase and urease were performed to confirm bacteria as *H. pylori*. Colony morphology was consistent with the characteristic shape of *H. pylori* colonies. Swabs and single colonies were used for DNA extraction.

H. pylori DNA extraction and genotyping

DNA was extracted from bacterial pellets using Gentra Puregene reagents, according to the manufacturer's instructions (Qiagen). Quantification of DNA from all strains was performed spectrophotometrically. The primers used to amplify the 3′ region of the *cagA* gene were CAG1 and CAG2 [22], resulting in generation of fragments varying from 591 to 856 bp. To amplify *vacA* s regions, primers VA1F and VA1R were selected, resulting in generation of

fragments of 259 bp for the type s1 variants or fragments of 286 bp for type s2 variants [23]. For the detection of the m region of the *vacA* gene by PCR, a mixture of four forward primers, MF1.1, MF1.2, MF1.3, and MF1.4, and one reverse primer MR1, was used, resulting in amplification of fragments of 107 bp for m1 or 182 bp for m2 strains [24].

Histopathology score

The global diagnosis for each subject was determined independently by 2 pathologists (P.C. and M.B.P.) based on all gastric biopsies from antrum, incisura angularis, and corpus, according to the updated Sydney system for gastritis, including the degree of inflammation, atrophy, and intestinal metaplasia using the recommended visual analogue scales [25] and the Padova International Classification for dysplasia [26]. Global diagnoses were scored on an ordinal scale from 1 to 6, as follows: 1, mild to moderate non-atrophic gastritis; 2, severe non-atrophic gastritis; 3, multifocal atrophic gastritis without intestinal metaplasia (MAG); 4, intestinal metaplasia (IM); 5, dysplasia; and 6, carcinoma. Degrees of inflammation (acute and chronic) were used to determine scores in cases with only non-atrophic gastritis.

In addition, we used a detailed histopathology scoring system to quantify differences in morphological variables within each global diagnosis category, which we have shown to be a sensitive and reliable system in a long-term follow-up of patients treated for *H. pylori* infection [27]. This system assigns numerical values to the severity of the atrophic non-metaplastic changes, the extent and type of intestinal metaplasia (representing the proportion of complete and incomplete types) and the severity of the dysplastic changes. We used this system, because increased severity of gastric atrophy is associated with greater cancer risk [28], the incomplete type of intestinal metaplasia (IM) carries a higher risk than the complete type [29,30], and the histological grade of dysplasia correlates with gastric cancer risk [26]. The multifocal atrophic gastritis (MAG) score (3) was modified, adding the following values: indefinite for atrophy (0.25), mild (0.50), moderate (0.75), and severe (1.0). The IM score (4) was modified according to type and extension. Four different IM types were defined: complete type (0.1), mixed predominant complete type (0.2), mixed predominant incomplete type (0.3), and incomplete (colonic) type (0.4). Periodic acid Schiff/Alcian blue and high iron diamine/Alcian blue stains were used for the assessment of the type of IM [27]. For the extension of the IM, each biopsy was also scored according to the area of the histological section with IM, in a 0–3 scale as follows: negative for IM, < 30%, >30 to 60%, and >60%, respectively. The mean extension of the IM (calculated using the total number of biopsies per subject) was grouped by tertiles. Each tertile was given a value: 0.2, 0.4, or 0.6, respectively. In order to obtain a total score for IM, values for type and extension were added to the original score for IM (4). The dysplasia score (5) was modified adding the following values: indefinite (0.25), low grade (0.50), and high grade (0.75). This augmented histopathology score, rather than the global diagnosis, was used for statistical analyses.

We utilised the Operative Link for Gastritis Assessment (OLGA) scoring system [31] for our histologic sections as an alternative method in order to confirm our findings. The OLGA system uses the respective antral and corpus updated Sydney scores for atrophic gastritis and categorises patients into stages 0–4 based on the distribution and severity of atrophy (with or without metaplasia).

8-OHdG staining

Tissue sections from the biopsies exhibiting the most advanced histologic lesion from each case were selected for 8-hydroxy-2′-deoxyguanosine (8-OHdG) staining. Paraffin-embedded tissues were deparaffinised and antigen retrieval was performed as described [18]. Sections were incubated with a mouse monoclonal anti-8-hydroxy-2′-deoxyguanosine (8-OHdG)

antibody at a dilution of 1:4000 (Abcam) overnight at room temperature. After washing, incubation with biotinylated secondary antibody (dilution 1:1000) was performed for 30 minutes. Sections were rinsed and incubated with streptavidin–horseradish peroxidase (Biocare Medical) for 30 minutes. Diaminobenzidine (Sigma-Aldrich) was used as a chromogen, and the tissues were counterstained with hematoxylin [18]. 8-OHdG staining was quantified by a pathologist unaware of the origin of the specimens, as the number of positive epithelial nuclei/100 cells in each of the superficial, neck, and deep regions of the glands to yield a 0 – 300 score; data were expressed as % of positive nuclei.

Multi-locus sequence typing (MLST) analysis

Fragments of seven unlinked housekeeping genes (*atpA*, *efp*, *ureI*, *ppa*, *mutY*, *trpC* and *yphC*), ranging from 398 to 627 bp per gene, were PCR-amplified from each Colombian *H. pylori* strain, using a high fidelity Taq polymerase (Invitrogen), and sequenced on both strands [32–34]. In addition, corresponding sequences from 380 globally-distributed *H. pylori* isolates, previously assigned to one of seven *H. pylori* ancestral haplogroups [34], were retrieved from an MLST database (http://pubmlst.org/helicobacter). Concatenated nucleotide sequences were aligned using MUSCLE [35]. Phylogenetic analyses were conducted in MEGA4 [36] using the Kimura 2-parameter model of nucleotide substitution and neighbour-joining clustering, with all partitions containing gaps eliminated from the final data set of 3401 positions. Branches of the phylogeny corresponding to partitions reproduced in fewer than 50% of the bootstrap replicates are collapsed. Interior branch bootstrap values were significantly high, indicating correct branch topology within clades. Phylogenetic trees were drawn to scale with branch lengths corresponding to the evolutionary distances used to infer the tree.

Statistical analyses

For analysis of phylogenetic distributions, Fisher's exact test was used. For analysis of histopathology scores and 8-OHdG staining, normality testing was performed, and since data were normally distributed, a generalised multivariate linear model with a Gaussian link was built to analyse the dependent variables of histopathology and DNA damage scores based on the independent variables of the MLST results and the geographic risk region, after adjusting for age. Means are listed ± a 95% confidence interval, with corresponding p values when comparisons were made. STATA® 11 (StataCorp LP) was used for all computations.

RESULTS

H. pylori Colombian strains segregate into two phylogenetic groups

To analyse the phylogeographic origin of *H. pylori* isolates, we employed MLST, which is based on analysis of the sequences of housekeeping genes and has been previously used to assess phylogenetic relationships of *H. pylori* strains from different parts of the world [32–34]. A total of 89 age-matched adult male subjects were included in the study, 44 from the LR region and 45 from the HR region (table S1); 81/89 of these persons were positive for *H. pylori* in gastric biopsies by culture detection. By genotype analysis, 64 of these 81 isolates were positive for *cagA* and *vacA* s1m1 (35 HR, 29 LR; tables S1,S2), two virulence factors strongly associated with gastric cancer [7–9]. We therefore focused on the 64 *cagA*+ *vacA* s1m1 strains. Phylogenetic relationships among the 64 sets of concatenated housekeeping gene sequences were analysed (figure 1A). We identified two distinct groups of strains: Group 1 (20 strains), which were all from the LR region; and Group 2, which included all of the isolates (35/35) from the HR region, and the remaining 9 isolates from the LR region (p<0.0001 for group segregation; figure 1B).

H. pylori strains from the HR region are all of European phylogeographic origin

We then performed a second MLST analysis, using sequences from the 64 Colombian strains plus those from 380 reference strains, which have previously been classified into distinct ancestral haplogroups [34]. The resulting neighbor-joining tree (figure 2A) demonstrated that *H. pylori* strains from Colombia are classified exclusively as either hpEurope or hpAfrica1, and no strains were classified as hpEAsia (hspAmerind or hspEAsia). Notably, 35/35 (100%) of the isolates from the HR region were classified as hpEurope. Among strains from the LR region, 10/29 (34.5%) were classified as hpEurope, while the remaining 19 were classified as hpAfrica1 (3 hspSAfrica, 16 hspWAfrica). Thus, all of the Colombian strains classified as hpAfrica1 were from the LR region. The MLST classification of isolates was strongly associated with the geographic origin of the strains within Colombia (p<0.0001; figure 2B). Notably, segregation of the Colombian strains within the context of the reference strains mirrored the findings when only the Colombian strains were analysed, since 19/20 (95%) of the strains in Group 1 (figure 1) were classified as hpAfrica1 (figure 2) and 44/44 (100%) of the strains in Group 2 (figure 1) were classified as hpEurope (figure 2).

We also performed MLST analysis on the 17 remaining strains that were not *cagA*⁺ *vacA* s1m1 (table S2). All of the *cagA*⁻ strains were classified as hpEurope. When the additional strains were included in the analysis along with the 64 *cagA*⁺ *vacA* s1m1 strains, we found that all 41 strains (100%) from the HR region were classified as hpEurope, and the majority (56%) of strains from the LR region were classified as hpAfrica1. Thus, when all strains were considered, there was still a striking difference in the MLST classification of isolates from the HR region compared to the LR region (p<0.0001).

H. pylori strains of European phylogeographic origin are associated with more advanced histologic lesions

We assessed disease progression in subjects from whom strains were cultured, by scoring gastric biopsies with a validated scoring system that is based on the updated Sydney System, and the Padova classification, and includes the type and extent of histologic lesions observed [25–27]. Histologic damage was analysed with a linear multivariate model. Subjects from the HR region had more advanced lesions (mean 3.9, 95% CI 3.6–4.2) than subjects from the LR region (mean 3.2, 95% CI 2.9–3.6, p=0.016; figure 3A). When scores were analysed based on the MLST classification of *H. pylori* strains (figure 3B), an even greater difference in severity of pre-malignant histologic disease progression was detected (mean 3.9, 95% CI 3.6–4.2 for subjects with hpEurope isolates *versus* mean 2.9, 95% CI 2.5–3.3 for subjects with hpAfrica1 isolates, p=0.001). Because the European group included isolates from both the HR (mountain) and the LR (coastal) region, these findings raised the question of the relative importance of phylogeny *versus* geographic origin of isolates as predictors of clinical outcome of *H. pylori* infection. We therefore compared three groups: LR region hpEurope, LR region hpAfrica1, and HR region hpEurope (figure 4A). Within the LR region, individuals infected with hpAfrica1 strains had less advanced lesions (mean 2.9, 95% CI 2.5–3.3; figures 4A, 4B) than subjects infected with hpEurope strains (mean 3.9, 95% CI 3.3–4.4; p=0.032; figures 4A, 4B). The histopathology scores (figure 4A) were similar when comparing subjects from the LR region and HR region (figure 4B) who harboured hpEurope strains (HR hpEurope mean 3.7, 95% CI 3.4–4.1; LR hpEurope mean 3.9, 95% CI 3.3–4.4).

We also analysed all the gastric tissues in our study with the OLGA system [31] for assessing atrophy and intestinal metaplasia. In accordance with the results obtained with our histologic scoring system, the subjects from the HR region had a higher mean OLGA stage than the subjects from the LR region (figure S1A), and the subjects infected with hpEurope

strains had a higher mean OLGA stage than those infected with hpAfrica1 strains (figure S1B). Within the LR region, the subjects infected with hpAfrica1 strains had a lower mean OLGA stage than persons infected with hpEurope strains, and the latter group was not significantly different compared to the group from the HR region infected with hpEurope strains (figure S1C).

H. pylori strains of European phylogeographic origin are associated with more DNA damage in gastric epithelial cells

Induction of DNA damage has been shown to be an important step in the carcinogenic process in *H. pylori* infection, and oxidative DNA damage can result from inflammatory events occurring during this infection [10,11], in addition to direct generation of oxidative stress within gastric epithelial cells [37]. DNA damage was therefore assessed by immunohistochemistry using an antibody directed against the oxidative DNA damage marker 8-hydroxy-2′-deoxyguanosine (8-OHdG), followed by scoring of positively-stained epithelial nuclei. Using a multivariate linear model, we determined that biopsies from subjects in the HR region exhibited more 8-OHdG-positive cells (mean 45.6%, 95% CI 38.5–52.8%) than those from the LR region (mean 31.2%, 95% CI 23.5–38.4%; p=0.009; Fig. 5A). A greater difference was detected when all persons infected with hpEurope strains (mean 46.0%, 95% CI 40.1–51.9%; figure 5B) were compared to those infected with hpAfrica1 strains (mean 22.7%, 95% CI 13.4–31.9%; p=0.001; figure 5B). Within the LR region, persons infected with hpAfrica1 strains had less 8-OHdG-positive cells (mean 22.7%, 95% CI 13.3–32.1%) than those infected with hpEurope strains (mean 45.2%, 95% CI 33.0–57.4%; p=0.007; figure 6A, 6B). Amongst subjects harbouring European strains, the number of 8-OHdG-positive cells in the biopsies from the subjects in the HR region (mean 46.2%, 95% CI 39.5–53.0%) was not different from the number of positive cells from those in the LR region (mean 45.2%, 95% CI 33.0–57.4%; figure 6A, 6B).

DISCUSSION

Our findings that inhabitants of the Andean region of Colombia were all colonised by hpEurope strains and that individuals living on the Pacific coast were infected with strains of European or African ancestry, with the latter predominating, suggest that Mestizos from the HR mountain region have lost their ancestral Amerindian strains, and that the Mulattos of the LR coastal region [18,20,21], still carry *H. pylori* strains from their African ancestors brought to the New World during the slave trade. Amerindians are hypothesised to have been infected originally with hspAmerind strains that are part of the hpEAsia group [34], but the Mestizo subjects in the Andean region in our study were exclusively infected with hpEurope strains, suggesting that hpEurope strains possess a competitive advantage compared to indigenous hspAmerind strains. In a study of 22 Mestizos from Colombia and Venezuela, subjects harboured only hpEurope and hpAfrica1 strains [38], but the relationship between MLST classification of strains and cancer risk parameters was not assessed. It has been reported that humans can be colonised with more than one *H. pylori* strain [39,40]. However, when we tested for bacterial genetic diversity using PCR with enterobacterial repetitive intergenic consensus sequence primers, we found that in all cases randomly selected for analysis (n=28), colonies isolated from antrum and corpus for each subject possessed the same DNA fingerprint. In addition, in all of the 12 cases that were further studied, 4 colonies (2 from antrum and 2 from corpus) were all identical.

We have shown that as expected, persons from the HR region have more histologic and DNA damage than those from the LR region. The data presented herein suggest that these findings are attributable to differences in the phylogeographic origin of the strains harboured by these subjects. In accordance with this concept, we have studied 10 strains from each of the risk regions cocultured with gastric epithelial cells, and found a significant increase in

oxidative DNA damage (measured by flow cytometry for 8-oxoguanosine [37]) with the HR *versus* LR strains; similarly, we also detected increased DNA damage in cells infected by hpEurope strains compared to hpAfrica1 strains (data not shown). These data support our hypothesis that *H. pylori* strains from the two different risk regions exhibit substantial differences in pathogenicity that relate to their phylogeographic origin.

Our findings are in distinction to previous considerations that in addition to ethnic differences, the high altitude and starch-based diet in HR regions compared to the low altitude and seafood-based diet in LR regions may explain differences in gastric cancer rates in Latin America [18]. Our group has recently reported that there are currently no significant difference in household condition index, crowding, vitamin supplement consumption, salt intake, and smoking status between the LR and HR regions [20], indicating that these are not confounding variables in our study. Furthermore, we have analysed our data based on smoking history [20], and have found that histopathology score and 8-OHdG staining were not different between smokers and non-smokers. In the LR region, where inhabitants are infected with either hpAfrica1 strains or hpEurope strains, we have shown that the divergence in *H. pylori* genetic features, as reflected by strain phylogeographic origin, is a very important factor that could explain the distinct patterns of pre-malignant histologic injury and oxidative DNA damage that we observed. Because the subjects in the LR region are ethnographically similar, but their gastric tissues exhibited a significant increase in the frequency of atrophy, intestinal metaplasia, and dysplasia if they were infected with hpEurope strains rather than hpAfrica1 strains, this suggests the importance of phylogeographic strain origin. However, host genetics may also have an effect, and to this end we intend to investigate the mitochondrial DNA of Colombian subjects from the LR and HR regions in future studies.

It has been reported that there is a strong association between the presence of the *cag* PAI or the *vacA* s1m1 genotype and an increased likelihood of developing gastric cancer in Latin American populations [41]. Consistent with this, we found that patients in this study who were infected with *cagA*⁻ strains (tables S1, S2) had very low histopathology scores with non-atrophic gastritis only (data not shown). Interestingly, all of the *cagA*⁻ strains in this study were classified as hpEurope (table S2). The histologic comparisons in the current study focused on subjects who were infected with *cagA*⁺ *vacA* s1m1 strains, because including the *cagA*⁻ strains would be a confounding factor. Importantly, we have demonstrated that among subjects all harbouring *cagA*⁺ and *vacA* s1m1 strains, there are large differences in histologic disease progression and epithelial DNA damage that correlate with the MLST type of the strains. This provides evidence that other genetic characteristics of the strains (besides the *cagA* and *vacA* status) are strongly associated with disease risk. Our phylogenetic data provide a new strategy for stratifying risk, and should serve as a foundation for studies searching for specific *H. pylori* genetic factors, that vary between strains of European *versus* African origin in order to gain insights into gastric carcinogenesis. We expect that next generation sequencing technology used for whole genome sequencing of *H. pylori* could be a useful future approach to identify specific determinants of cancer risk.

For Summary Box

What is already known about this subject?

▶   *Helicobacter pylori* is the main causative agent for peptic ulceration and gastric cancer.

▶ High prevalence rates of *H. pylori* generally correlate with a high risk of gastric cancer, but there are areas of the world with low gastric cancer rates despite high *H. pylori* prevalence.

▶ *H. pylori* strains with *cagA* and *vacA* s1m1 virulence factors are the most commonly associated with gastric cancer risk.

▶ In Colombia, people living in the Andean mountain region have a 25-fold greater risk of developing gastric cancer than people living in the coastal region.

What are the new findings?

▶ *H. pylori* strains infecting human subjects from the Colombian mountain region are exclusively of European ancestry.

▶ *H. pylori* strains of European or African phylogeographic origin are found in the coastal region, with the latter predominating.

▶ European phylogeographic origin of the *H. pylori* strains is strongly associated with more advanced histologic lesions and increased DNA damage in gastric epithelial cells, regardless of whether the subjects were from the high risk mountain region or the low risk coastal region.

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

▶ We have found that the ancestral origin of *H. pylori* strains is a strong predictor of gastric cancer risk.

▶ Therefore, determination of phylogeographic origin of *H. pylori*-infecting strains may be a useful strategy for risk stratification and clinical management, including intensification of eradication programs and surveillance endoscopy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

1. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet. 1984; 1:1311–5. [PubMed: 6145023]

2. Uemura N, Okamoto S, Yamamoto S, et al. *Helicobacter pylori* infection and the development of gastric cancer. N Engl J Med. 2001; 345:784–9. [PubMed: 11556297]

3. IARC. Infection with *Helicobacter pylori*. IARC Monogr Eval Carcinog Risks Hum. 1994; 61:177–240. [PubMed: 7715070]

4. Correa P. A human model of gastric carcinogenesis. Cancer Res. 1988; 48:3554–60. [PubMed: 3288329]

5. Correa P, Haenszel W, Cuello C, et al. Gastric precancerous process in a high risk population: cohort follow-up. Cancer Res. 1990; 50:4737–40. [PubMed: 2369748]

6. Peek RM Jr. Fiske C, Wilson KT. Role of innate immunity in *Helicobacter pylori*-induced gastric malignancy. Physiol Rev. 2010; 90:831–58. [PubMed: 20664074]

7. Basso D, Zambon CF, Letley DP, et al. Clinical relevance of *Helicobacter pylori* cagA and vacA gene polymorphisms. Gastroenterology. 2008; 135:91–9. [PubMed: 18474244]

8. Blaser MJ, Perez-Perez GI, Kleanthous H, et al. Infection with *Helicobacter pylori* strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res. 1995; 55:2111–5. [PubMed: 7743510]

9. Figueiredo C, Machado JC, Pharoah P, et al. *Helicobacter pylori* and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. J Natl Cancer Inst. 2002; 94:1680–7. [PubMed: 12441323]

10. Baik S-C, Youn H-S, Chung M-H, et al. Increased Oxidative DNA Damage in *Helicobacter pylori*-infected Human Gastric Mucosa. Cancer Res. 1996; 56:1279–82. [PubMed: 8640814]

11. Farinati F, Cardin R, Degan P, et al. Oxidative DNA damage accumulation in gastric carcinogenesis. Gut. 1998; 42:351–6. [PubMed: 9577340]

12. Correa P, Fox J, Fontham E, et al. *Helicobacter pylori* and gastric carcinoma. Serum antibody prevalence in populations with contrasting cancer risks. Cancer. 1990; 66:2569–74. [PubMed: 2249197]

13. Holcombe C. *Helicobacter pylori*: the African enigma. Gut. 1992; 33:429–31. [PubMed: 1582581]

14. Torres J, Lopez L, Lazcano E, et al. Trends in *Helicobacter pylori* infection and gastric cancer in Mexico. Cancer Epidemiol Biomarkers Prev. 2005; 14:1874–7. [PubMed: 16103429]

15. Correa P, Cuello C, Duque E, et al. Gastric cancer in Colombia. III. Natural history of precursor lesions. J Natl Cancer Inst. 1976; 57:1027–35. [PubMed: 1003539]

16. Cuello C, Correa P, Haenszel W, et al. Gastric cancer in Colombia. I. Cancer risk and suspect environmental agents. J Natl Cancer Inst. 1976; 57:1015–20. [PubMed: 1003537]

17. Bravo LE, van Doom LJ, Realpe JL, et al. Virulence-associated genotypes of *Helicobacter pylori*: do they explain the African enigma? Am J Gastroenterol. 2002; 97:2839–42. [PubMed: 12425557]

18. Piazuelo MB, Camargo MC, Mera RM, et al. Eosinophils and mast cells in chronic gastritis: possible implications in carcinogenesis. Hum Pathol. 2008; 39:1360–9. [PubMed: 18614201]

19. Correa P, Cuello C, Duque E. Carcinoma and intestinal metaplasia of the stomach in Colombian migrants. J Natl Cancer Inst. 1970; 44:297–306. [PubMed: 11515071]

20. Camargo MC, Burk RF, Bravo LE, et al. Plasma selenium measurements in subjects from areas with contrasting gastric cancer risks in Colombia. Arch Med Res. 2008; 39:443–51. [PubMed: 18375257]

21. Whary MT, Sundina N, Bravo LE, et al. Intestinal helminthiasis in Colombian children promotes a Th2 response to *Helicobacter pylori*: possible implications for gastric carcinogenesis. Cancer Epidemiol Biomarkers Prev. 2005; 14:1464–9. [PubMed: 15941957]

22. Yamaoka Y, Kodama T, Kashima K, et al. Variants of the 3' region of the cagA gene in *Helicobacter pylori* isolates from patients with different H. pylori-associated diseases. J Clin Microbiol. 1998; 36:2258–63. [PubMed: 9666002]

23. Atherton JC, Cao P, Peek RM Jr. et al. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific vacA types with cytotoxin production and peptic ulceration. J Biol Chem. 1995; 270:17771–7. [PubMed: 7629077]

24. van Doorn LJ, Figueiredo C, Sanna R, et al. Expanding allelic diversity of *Helicobacter pylori* vacA. J Clin Microbiol. 1998; 36:2597–603. [PubMed: 9705399]

25. Dixon MF, Genta RM, Yardley JH, et al. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol. 1996; 20:1161–81. [PubMed: 8827022]

26. Rugge M, Correa P, Dixon MF, et al. Gastric dysplasia: the Padova international classification. Am J Surg Pathol. 2000; 24:167–76. [PubMed: 10680883]

27. Mera R, Fontham ET, Bravo LE, et al. Long term follow up of patients treated for *Helicobacter pylori* infection. Gut. 2005; 54:1536–40. [PubMed: 15985559]

28. Ohata H, Kitauchi S, Yoshimura N, et al. Progression of chronic atrophic gastritis associated with *Helicobacter pylori* infection increases risk of gastric cancer. Int J Cancer. 2004; 109:138–43. [PubMed: 14735480]

29. Filipe MI, Munoz N, Matko I, et al. Intestinal metaplasia types and the risk of gastric cancer: a cohort study in Slovenia. Int J Cancer. 1994; 57:324–9. [PubMed: 8168991]

30. Gonzalez CA, Pardo ML, Liso JM, et al. Gastric cancer occurrence in preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. Int J Cancer. 2010; 127:2654–60. [PubMed: 20178099]

31. Rugge M, Correa P, Di Mario F, et al. OLGA staging for gastritis: a tutorial. Dig Liver Dis. 2008; 40:650–8. [PubMed: 18424244]

32. Devi SM, Ahmed I, Francalacci P, et al. Ancestral European roots of *Helicobacter pylori* in India. BMC Genomics. 2007; 8:184. [PubMed: 17584914]

33. Devi SM, Ahmed I, Khan AA, et al. Genomes of *Helicobacter pylori* from native Peruvians suggest admixture of ancestral and modern lineages and reveal a western type cag-pathogenicity island. BMC Genomics. 2006; 7:191. [PubMed: 16872520]

34. Falush D, Wirth T, Linz B, et al. Traces of human migrations in *Helicobacter pylori* populations. Science. 2003; 299:1582–5. [PubMed: 12624269]

35. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–7. [PubMed: 15034147]

36. Tamura K, Dudley J, Nei M, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol Biol Evol. 2007; 24:1596–9. [PubMed: 17488738]

37. Xu H, Chaturvedi R, Cheng Y, et al. Spermine oxidation induced by *Helicobacter pylori* results in apoptosis and DNA damage: implications for gastric carcinogenesis. Cancer Res. 2004; 64:8521–5. [PubMed: 15574757]

38. Dominguez-Bello MG, Perez ME, Bortolini MC, et al. Amerindian *Helicobacter pylori* strains go extinct, as european strains expand their host range. PLoS One. 2008; 3:e3307. [PubMed: 18830403]

39. Jorgensen M, Daskalopoulos G, Warburton V, et al. Multiple strain colonization and metronidazole resistance in *Helicobacter pylori*-infected patients: identification from sequential and multiple biopsy specimens. J Infect Dis. 1996; 174:631–5. [PubMed: 8769626]

40. van der Ende A, Rauws EA, Feller M, et al. Heterogeneous *Helicobacter pylori* isolates from members of a family with a history of peptic ulcer disease. Gastroenterology. 1996; 111:638–47. [PubMed: 8780568]

41. Sugimoto M, Yamaoka Y. The association of vacA genotype and *Helicobacter pylori*-related disease in Latin American and African populations. Clin Microbiol Infect. 2009; 15:835–42. [PubMed: 19392900]
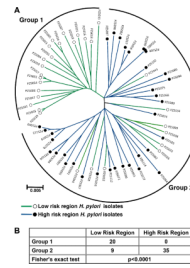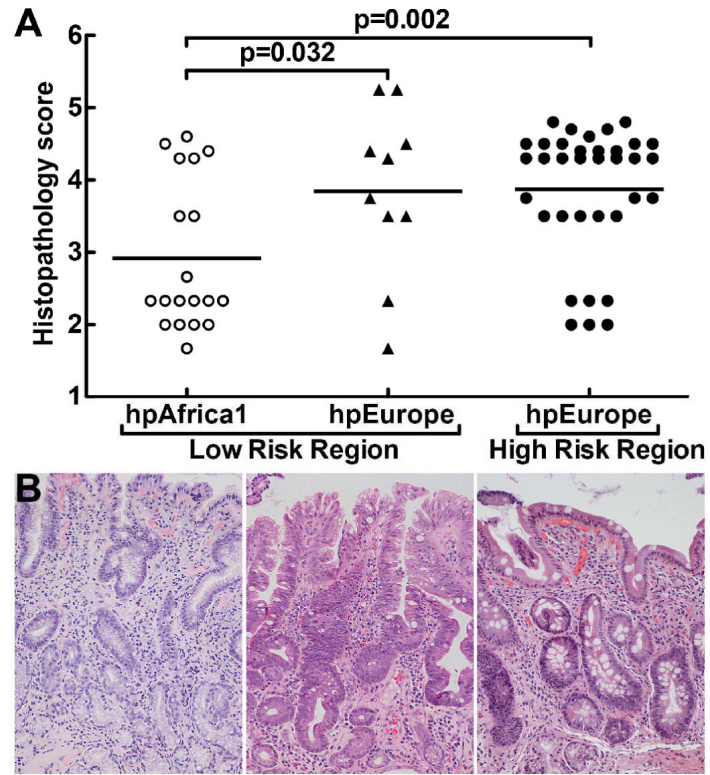
Figure 1.
MLST analysis of 64 *cagA*[+] *vacA* s1m1 *Helicobacter pylori* strains from Colombia. (A) Neighbor-joining tree; branches are drawn to scale to represent evolutionary distance. ◯ with green lines, isolates from LR region (low risk of gastric cancer); ● with blue lines, isolates from HR region (high risk of gastric cancer). The brackets denote the classification of the strains into two groups. (B) Distribution of strains into phylogenetic groups defined in (A), according to risk region of origin.
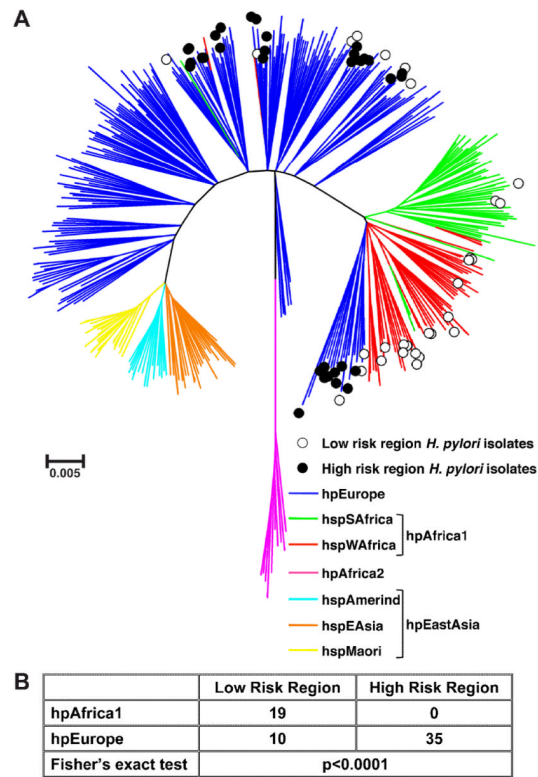
Figure 2.
MLST analysis of 64 *cagA*$^+$ *vacA* s1m1 Colombian isolates from the HR and LR regions, along with 380 reference strains that were previously classified into distinct ancestral haplogroups [34]. (A) Neighbor-joining tree built using bootstrapping with 10,000 trials. Branches are drawn to scale to represent evolutionary distance. ◯, isolates from LR region; ● isolates from HR region. The colours of the branches represent the classification of strains into distinct ancestral haplogroups [34]. (B) Distribution of strains into phylogeographic groups defined in (A), according to risk region of origin.
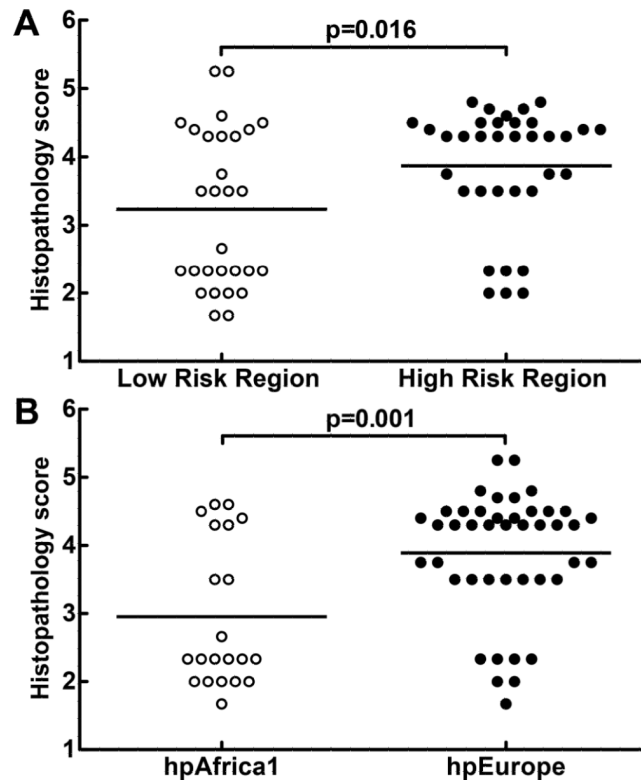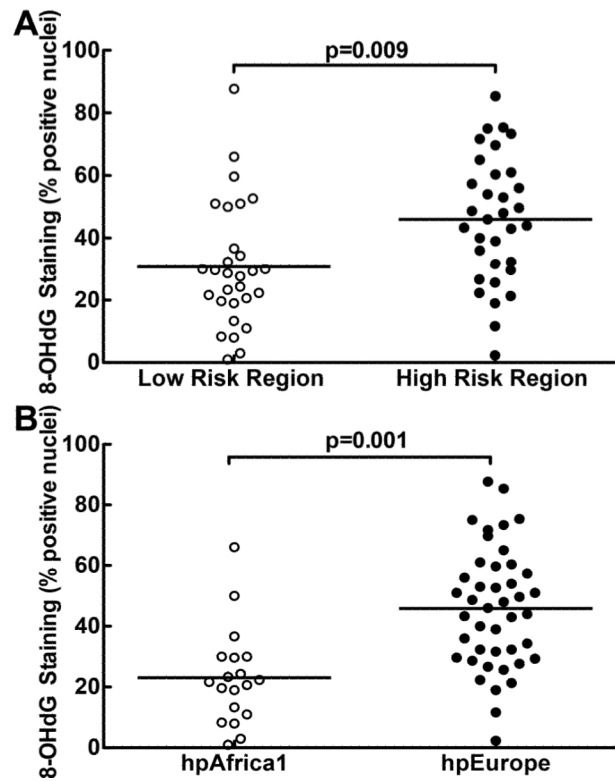
Figure 3.
Histopathologic analysis of biopsies from subjects infected with Colombian *H. pylori* strains. Histologic scoring on a 1 to 6 scale (defined in Methods; each symbol represents the score for an individual subject. (A) Histopathology scores by risk region of origin. (B) Scores grouped by phylogenetic category of the strains. Horizontal bars represent the mean; p values were obtained by a linear multivariate model analysis.
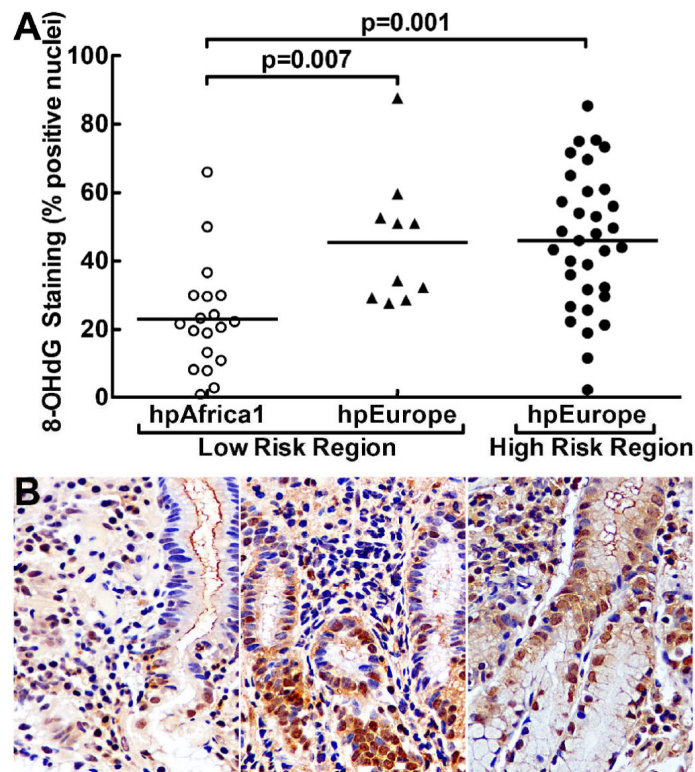
Figure 4.
Histopathologic analysis of biopsies from subjects infected with Colombian *H. pylori* strains. Histologic scoring on a 1 to 6 scale (defined in Methods; each symbol represents the score for an individual subject. (A) Scores grouped according to phylogeny within a given risk region of origin. Horizontal bars represent the mean; pairwise comparisons were made using Dunnett's test with Sidak's adjustment. (B) Representative photomicrographs of hematoxylin and eosin staining of gastric tissue as follows: (*left panel*) LR subject infected with hpAfrica1 strain, showing non-atrophic gastritis; (*center panel*) LR subject infected with hpEurope strain, demonstrating intestinal metaplasia with changes indefinite for dysplasia; and (*right panel*) HR subject infected with hpEurope strain with intestinal metaplasia. Magnification in B was 200×.

Figure 5.
Detection of oxidative DNA damage in gastric biopsies by quantification of 8-OHdG-positive epithelial cell nuclei, expressed as % positively-stained nuclei. (*A*) 8-OHdG levels, by risk region of origin. (*B*) Cases grouped according to strain phylogeny. Horizontal bars represent the mean; statistics were performed as in figure 3.

Figure 6.
Detection of oxidative DNA damage in gastric biopsies by quantification of 8-OHdG-positive epithelial cell nuclei, expressed as % positively-stained nuclei. (A) Cases grouped according to strain phylogeny within a given risk region of origin. Horizontal bars represent the mean; statistics were performed as in figure 4. (B) Representative photomicrographs of 8-OHdG staining on gastric tissues as follows: (*left panel*) LR subject infected with hpAfrica1 strain; (*center panel*) LR subject infected with hpEurope strain; and (*right panel*) HR subject infected with hpEurope strain. Nuclei showing DNA damage are stained brown. Magnification in B was 400×.