Published in final edited form as:

Infect Genet Evol. 2012 March ; 12(2): 203–213. doi:10.1016/j.meegid.2011.12.002.

Molecular epidemiology, population genetics, and pathogenic role of *Helicobacter pylori*

Rumiko Suzuki^a, Seiji Shiota^a, and Yoshio Yamaoka^{a,b}

^aDepartment of Environmental and Preventive Medicine, Oita University Faculty of Medicine, 1-1 Idaigaoka, Hasama-machi, Yufu-City, Oita 879-5593, Japan

^bDepartment of Medicine-Gastroenterology, Baylor College of Medicine and Michael E. DeBakey Veterans Affairs Medical Center, 2002 Holcombe Blvd. Houston, Texas 77030, USA

Abstract

Helicobacter pylori infection is linked to various gastroduodenal diseases; however, only approximately 20% of infected individuals develop severe diseases. Despite the high prevalence of *H. pylori* infection in Africa and South Asia, the incidence of gastric cancer in these areas is much lower than in other countries. Furthermore, the incidence of gastric cancer tends to decrease from north to south in East Asia. Such geographic differences in the pathology can be explained, at least in part, by the presence of different types of *H. pylori* virulence factors, especially *cagA*, *vacA*, and the right end of the *cag* pathogenicity island. The genotype of the virulence genes is also useful as a tool to track human migration utilizing the high genetic diversity and frequent recombination between different *H. pylori* strains. Multilocus sequence typing (MLST) analysis using 7 housekeeping genes can also help predict the history of human migrations. Population structure analysis based on MLST has revealed 7 modern population types of *H. pylori*, which derived from 6 ancestral populations. Interestingly, the incidence of gastric cancer is closely related to the distribution of *H. pylori* populations. The different incidence of gastric cancer can be partly attributed to the different genotypes of *H. pylori* circulating in different geographic areas. Although approaches by MLST and virulence factors are effective, these methods focus on a small number of genes and may miss information conveyed by the rest of the genome. Genome-wide analyses using DNA microarray or whole-genome sequencing technology give a broad view on the genome of *H. pylori*. In particular, next-generation sequencers, which can read DNA sequences in less time and at lower costs than Sanger sequencing, enabled us to efficiently investigate not only the evolution of *H. pylori*, but also novel virulence factors and genomic changes related to drug resistance.

Keywords

Helicobacter pylori; *cagA*; *vacA*; multilocus sequence typing; whole genome sequencing technology; next-generation sequencer

© 2011 Elsevier B.V. All rights reserved.

Corresponding author: Yoshio Yamaoka MD, PhD, Department of Environmental and Preventive Medicine, Oita University Faculty of Medicine, 1-1 Idaigaoka, Hasama-machi, Yufu-City, Oita 879-5593, Japan, Tel: +81-97-586-5740; Fax: +81-97-586-5749, yyamaoka@oita-u.ac.jp.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Potential competing interests: The authors declare that they have no competing interests.

1. Introduction

Helicobacter pylori is a gram-negative spiral bacterium whose ecological niche is the human stomach. *H. pylori* infection is now accepted as the major cause of chronic gastritis. Several epidemiological studies have shown that *H. pylori* infection is also linked to severe gastritis-associated diseases, including peptic ulcer and gastric cancer (Suerbaum and Michetti, 2002). In 1994, the International Agency for Research on Cancer categorized *H. pylori* infection as a group I carcinogen (1994). Although the incidence of gastric cancer is decreasing, it remains the fourth most common cancer and second leading cause of cancer-related deaths worldwide (access in 2010b). Interestingly, despite the high prevalence of *H. pylori* infection in Africa and South Asia, the incidence of gastric cancer in these areas is much lower than in other countries; these phenomena are called African enigmas or Asian enigmas (Malaty, 2007). Furthermore, the incidence of gastric cancer has a tendency to decrease from north to south even within East Asia. In addition to host, environmental, and dietary factors, another possible reason for the varying outcomes of *H. pylori* infection relates to differences in the virulence of *H. pylori* strains. Several virulence factors of *H. pylori* such as *cagA*, *vacA*, *oipA*, *babA*, *hopQ*, and *homaB* have been demonstrated to be predictors of gastric atrophy, intestinal metaplasia, and severe clinical outcomes (Jung et al., 2009; Lu et al., 2005; Ohno et al., 2009; Sugimoto et al., 2009a; Yamaoka et al., 2002a; Yamaoka et al., 1998b; Yamaoka et al., 2000a). Importantly, most of these virulence factors are associated with each other; *cagA*-positive strains also possess a *vacA* s1/m1 region type and they are further closely linked to the presence of the *babA* and *oipA* “on” status (Yamaoka, 2010).

H. pylori strains from different geographical areas show clear phylogeographic features (Figure 1); these features enabled us to assume the migration of human populations by phylogeographic analyses of *H. pylori*. In addition, the genetic diversity within *H. pylori* is greater than within most other bacteria (Achtman et al., 1999), and about 50-fold greater than that of the human population (Li and Sadler, 1991). Furthermore, frequent recombination between different *H. pylori* strains (Suerbaum and Josenhans, 2007) leads to only partial linkage disequilibrium between polymorphic loci, which provide additional information for population genetic analysis (Falush et al., 2003). In addition, multilocus sequence typing (MLST) analysis, which uses 7 housekeeping genes, can also help predict the history of human migrations (Achtman et al., 1999; Falush et al., 2003; Linz et al., 2007; Moodley et al., 2009; Wirth et al., 2004). This review describes the current knowledge about *H. pylori* typing (*cagA*, *vacA*, and housekeeping genes) from the point of virulence factors, as well as approaches for elucidating the history of human migration. In addition, we give an overview on the utilization of genome-wide information for advanced studies.

2. Virulence factors

2.1. CagA

cagA, which encodes a highly immunogenic protein (CagA), is the most extensively studied *H. pylori* virulence factor (Covacci et al., 1993; Tummuru et al., 1993). The *cagA* gene is located at one end of the *cag* pathogenicity island (PAI), an approximately 40-kbp-region that is thought to have been incorporated into the *H. pylori* genome by horizontal transfer from an unknown source (Censini et al., 1996). The *cag* PAI encodes a type IV secretion system, through which CagA is delivered into host cells (Asahi et al., 2000; Backert and Selbach, 2008). CagA has been reported to interact with various target molecules in host cells; the best studied is the cytoplasmic Src homology 2 domain of Src homology 2 phosphatase (SHP-2). Mutations of SHP-2 have been found in various human malignancies and mice that lacked the SHP-2-binding site developed hyperplastic antral tumors (Judd et al., 2004), indicating that SHP-2 plays an important role in gastric cancer. Gastric dysplasia

and cancer developed in Mongolian gerbils infected with the wild-type strain, but no occurrence was observed in animals infected with the *cagA* knockout strain (Franco et al., 2005; Franco et al., 2008). Another study showed that CagA transgenic mice had gastric epithelial hyperplasia and some of the mice developed gastric polyps and adenocarcinomas of the stomach and small intestine (Ohnishi et al., 2008). These results provide strong evidence for the role of CagA as a bacterium-derived oncoprotein. There are 2 types of clinical *H. pylori* isolate: CagA-producing (*cagA*-positive) strains and CagA-non-producing (*cagA*-negative) strains. Almost all *H. pylori* isolates from East Asia are *cagA*-positive, whereas approximately 20% to 40% of isolates from Europe and Africa are *cagA*-negative. In Western countries, it has been reported that individuals infected with *cagA*-positive *H. pylori* have a higher risk of peptic ulcer or gastric cancer than those infected with *cagA*-negative strains (van Doorn et al., 1998; Yamaoka et al., 2002a). However, in East Asian countries, it is difficult to prove the importance of *cagA* in clinical outcomes because almost all *H. pylori* strains are *cagA*-positive (Yamaoka et al., 1999c).

Depending on the strain, *cagA* has different numbers of repeat sequences located in the 3'-region of the *cagA* gene (Yamaoka et al., 1999a; Yamaoka et al., 1998a; Yamaoka et al., 2000b). The repeat regions contain the Glu-Pro-Ile-Tyr-Ala (EPIYA) motif. Recently, sequences have been annotated according to segments (20-50 amino acids) flanking the EPIYA motifs (ie, segments EPIYA-A, -B, -C, or -D) (Yamaoka, 2010). Recent studies show that the East-Asian-type CagA, containing EPIYA-D segments, exhibits a stronger binding affinity for SHP-2 and a greater ability to induce morphological changes in epithelial cells than Western-type CagA, which contains EPIYA-C segments (Hatakeyama, 2004). Some reports showed that individuals infected with East-Asian-type *cagA* strains have an increased risk of peptic ulcer or gastric cancer compared with those infected with Western-type *cagA* strains (Jones et al., 2009; Vilaichone et al., 2004a). Consistently, transgenic mice expressing the East-Asian-type CagA developed tumors more efficiently than did those expressing the Western-type CagA (Miura et al., 2009). In addition, in Western countries, the incidence of gastric cancer is higher in patients infected with strains carrying multiple EPIYA-C repeats compared to those infected with strains with a single repeat (Argent et al., 2004; Azuma et al., 2002; Xia et al., 2009; Yamaoka et al., 1999a; Yamaoka et al., 1998a). A recent large-scale study also showed that a higher number of EPIYA-C repeats was associated with gastric cancer and gastric precancerous lesions, as demonstrated by histological gastric atrophy/metaplastic changes and decreased serum levels of pepsinogen I (Batista et al., 2011). Although rare, some East Asian strains have a Western-type CagA sequence, especially strains isolated from Okinawa, the southern islands of Japan (Azuma et al., 2004; Satomi et al., 2006; Yamaoka et al., 2002b; Yamazaki et al., 2005). On the other hand, none of the Western strains have an East-Asian-type CagA sequence (Xia et al., 2009). We previously examined the frequency of each EPIYA motif using 3 data sources, ie, the National Center for Biotechnology Information (NCBI, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov), UniProtKB/Swiss-Prot (the Swiss Institute for Bioinformatics and the European Bioinformatics Institute, www.ebi.ac.uk/swissprot/), and the DNA Data Bank of Japan, the National Institute of Genetics, www.ddbj.nig.ac.jp) (Xia et al., 2009). In total, 1,796 EPIYA motifs were obtained from 560 CagA sequences. On average, each CagA sequence contained approximately 3 EPIYA motifs. The 3 most frequent EPIYA motifs were EPIYA (1,657/1,796 = 92.3%), EPIYT (92/1,796 = 5.1%), and ESIYA (24/1,796 = 1.3%) (Xia et al., 2009). Of note, most of the changes in the EPIYA motif were found in the EPIYA-B segment of Western-type CagA.

Interestingly, the incidence of gastric cancer in Okinawa (6.3 deaths/100,000 population) is the lowest in Japan (access in 2010a), although the prevalence of *H. pylori* in this area is not significantly different from that in other parts of Japan (Ito et al., 1996; Nobuta et al., 2004).

The low prevalence of gastric cancer in Okinawa is comparable to that of the U.S. (4.1/100,000) (access in 2010b). Okinawa was under the rule of the U.S. after World War II (WWII) until 1972, and there are still many U.S. populations (access in 2010c). A few previous reports showed a high prevalence of Western-type CagA strains in Okinawa compared with other areas in Japan (Azuma et al., 2004; Satomi et al., 2006; Yamazaki et al., 2005). Our preliminary data show that the prevalence of EPIYA-C motifs in Okinawa was approximately 10% (our unpublished data), which is comparable to that of the U.S. (Yamaoka et al., 1999b). It is still not clear where the Western-type CagA strains originated from, ie, whether they derived from the U.S. populations after WWII or not. Intriguingly, recent studies showed that the Western-type CagA detected in strains from Okinawa formed a different cluster compared to the original Western-type CagA and it was named the J-Western-type CagA subtype (Truong et al., 2009). We also reported that the J-Western-type CagA strains possess a 12-bp insertion in the *cagA* sequence compared to the original Western-type CagA strains (Shiota et al., 2010). Further studies are needed to clarify the role of the different Western-type CagA genes in the pathogenesis of the disease.

An additional CagA type, Amerindian CagA, was recently reported from populations of the Peruvian Amazon (Kersulyte et al., 2010; Suzuki et al., 2011a). Amerindian CagA can be divided into 2 types, ie, AM-I and AM-II (Suzuki et al., 2011a). Amerindian CagAs have altered or degenerated EPIYA-B motifs; ESIYT and GSIYD in AM-I and AM-II CagA, respectively. The N-terminus of AM-II CagA lacks 2 large internal segments totaling 180 amino acids. Interestingly, AM-II CagA has attenuated abilities to stimulate gastric epithelial proliferation and inflammation during infection compared to those of Western-type CagA or East-Asian-type CagA. The Amerindian CagA multimerization (CM) segment, which has been reported to promote proliferation and inflammation (Yamaoka, 2010), plays an important role in those findings (Suzuki et al., 2011a).

The right end of the *cag* PAI has been divided into 5 subtypes according to deletion, insertion, and substitution motifs (Kersulyte et al., 2000). Type I is most common in isolates from ethnic European groups and from Africa; type II is predominant in those from East Asia; and type III is predominant in isolates from South Asia (Kersulyte et al., 2000; Yamaoka et al., 2002b). Type IV is very rare and, therefore, has not been assigned to a specific geographical area (Kersulyte et al., 2000; Yamaoka et al., 2002b). Type V is found in a few strains from Calcutta, India (Kersulyte et al., 2000; Yamaoka et al., 2002b). Interestingly, our report showed that type V was present in 10% of isolates from patients of Thailand, and the ratio was especially high in strains obtained from ethnic Thai (21%) (Vilaichone et al., 2004b). The presence of this genotype in Thailand suggests that it migrated to the east of Calcutta. It will be necessary to examine strains from Myanmar as well as from ethnic minorities in Northern Thailand in order to investigate this possibility. In addition, examination of strains from ethnic Thai in Southern Thailand and Northern Malaysia will be needed to complete the distribution map and migration pattern of strains conveying the type V *H. pylori* genotype.

2.2. VacA

VacA is the second most extensively studied *H. pylori* virulence factor. Unlike *cagA*, virtually all *H. pylori* strains have a functional *vacA*, which encodes a vacuolating cytotoxin. VacA can induce vacuolation, multiple cellular activities, including membrane-channel formation, cytochrome *c* release from mitochondria leading to apoptosis, and binding to cell-membrane receptors followed by initiation of a proinflammatory response (Atherton, 2006; Cover and Blanke, 2005; Kusters et al., 2006). The differences in *vacA* structure at the signal (s) region (s1 and s2) and the middle (m) region (m1 and m2) (Atherton et al., 1995) led to variations in the vacuolating activity of different *H. pylori* strains (Cover and Blaser, 1992; Leunk, 1991). *In vitro* experiments showed that s1m1 strains are the most cytotoxic,

followed by s1m2 strains, whereas s2m2 strains have no cytotoxic activity; s2m1 strains are rare (Atherton et al., 1995). In agreement with *in vitro* data, many studies in Western countries showed that individuals infected with *vacA* s1 or m1 *H. pylori* strains have an increased risk of peptic ulcer or gastric cancer compared with those with s2 or m2 strains (Atherton et al., 1995; Cover and Blaser, 1992; Sugimoto and Yamaoka, 2009; Sugimoto et al., 2009b). The m1 strains are common in areas of Northeast Asia, eg, Japan and South Korea, whereas m2 strains are predominant in areas of Southeast Asia, eg, Taiwan and Vietnam (Nguyen et al., 2010; Yamaoka et al., 2002b). As the incidence of gastric cancer is higher in Northeast Asia than in Southeast Asia, the m region may play a role in the regional difference in the disease pattern. We recently reported that, in Vietnam, the *vacA* m2 strain is more prevalent in Ho Chi Minh City than in Hanoi, and the incidence of gastric cancer is lower in Ho Chi Minh City than in Hanoi (Nguyen et al., 2010). This finding also supports the possibility that the *vacA* m region is related to the clinical outcomes.

The *vacA* s1 and m1 type can be subdivided into 3 types, ie, s1a, s1b, and s1c (Atherton et al., 1995) and m1a, m1b, and m1c, respectively (Mukhopadhyay et al., 2000). The *vacA* s1c and m1b types are common in East Asia and the s1a and m1c types are common in South Asia (Mukhopadhyay et al., 2000; Yamaoka et al., 2002b). The *vacA* m1c genotype is predominant in strains from Central Asia (Caucasians and ethnic Kazakhs) (Yamaoka et al., 2002b). The m1a type is common in Africans and ethnic Europeans (Van Doorn et al., 1999; Yamaoka et al., 2002b). Both the s1a and s1b types are common in strains of ethnic Europeans, and s1b types are especially common in strains from the Iberian Peninsula and Latin America (Sugimoto and Yamaoka, 2009; Van Doorn et al., 1999). The s1b type is also predominant in Africa (Letley et al., 1999; Sugimoto and Yamaoka, 2009).

A third disease-related region of *vacA* named the intermediate (i) region was identified between the s region and the m region (Rhead et al., 2007). Although s1m1 and s2m2 strains were exclusively classified as i1 and i2, respectively, s1m2 strains were classified as either type i1 or i2, and i1 strains were shown to be more pathogenic. A very recent study showed that the amino acid at position 196 of the i region was significantly associated with severe outcomes (Jones et al., 2011). In addition, the authors found that the amino acid at position 231 of the i region was linked to severe disease pathologies only in non-East-Asian strains. A fourth and possibly disease-related region is the deletion (d) region located between the i region and the m region. The d region can also be subdivided into d1 (no deletion) and d2 (a 69- to 81-bp deletion). The role of the status of the i region or d region is still controversial (Ogiwara et al., 2008; Ogiwara et al., 2009; Rhead et al., 2007). At present, therefore, the type of the m region may be the best overall marker for gastric cancer, especially in East Asia.

3. Human migration accompanied by *H. pylori*

3.1. Virulence genes as a tool for tracking human migration

On the basis of the analysis of the *cagA*, *cag* right end junction, and *vacA* genotypes of more than 1,042 *H. pylori* strains collected from East Asia, Southeast Asia, South Asia, Central Asia, Europe, Africa, North America, and South America (Yamaoka et al., 2002b), *H. pylori* strains can be divided into 5 major groups (East Asian-type, South/Central Asian-type, Iberian/African-type, and European-type) according to geographical associations (Table 1). In these groups, *cagA*-negative and/or *vacA* m2 genotypes are not taken into account, but we can predict the geographical origins of each group using available genotypes (ie, *cagA*-negative strains with *vacA* s1a-m1a genotype are predicted to be of the European-type). Overall, the genotype of the virulence genes is useful for tracking human migration, as well as for investigating *H. pylori*-related gastroduodenal diseases.

The genotypes of the virulence genes revealed the human migration to the Americas. The Americas were populated by humans of East Asian ancestry approximately 15,000 years ago. Over the last 500 years, Europeans and Africans migrated to the Americas, leading to an increasing Mestizo population. Our group reported that isolates of native Colombians had similar but not identical structures of *vacA* m and *cagA* compared to isolates from East Asia (Yamaoka et al., 2002b). A *cagA*-negative native Alaskan strain also possessed specific *vacA* m structures which were closer to structures from East Asia than to those from non-Asian countries (Yamaoka et al., 2002b). Isolates of native Venezuelans were also reported to have a high frequency of the *vacA* s1c genotype (Ghose et al., 2002). These data confirmed that *H. pylori* “accompanied” humans when they crossed the Bering Strait from Asia to the New World. Importantly, none of the *H. pylori* strains from Mestizo populations possessed East Asian-like genotypes. Sequence analysis of *H. pylori* genomes has shown that East Asian-like Amerindian strains were the least genetically diverse, probably due to a genetic bottleneck, whereas European strains are the most diverse among European, African, Amerindian, and East Asian strains (Domínguez-Bello et al., 2008). If diversity is important for the success of *H. pylori* colonization, the East Asian-like Amerindian strains may lack the needed diversity to compete with the diverse *H. pylori* population brought to the New World by non-Amerindian hosts and has therefore disappeared (Yamaoka, 2009).

CagA sequence typing alone can serve as a tool for tracking human migration. As described above, Western-type CagA detected in strains from the Okinawa (J-Western-type CagA) formed a different cluster compared to the original Western-type CagA (Truong et al., 2009). The pre-EPIYA region of CagA also shows geographic divergence (Uchida et al., 2009). Most strains isolated from East Asia have a 39-bp deletion, but this deletion was absent in most strains from Western countries. On the other hand, an 18-bp deletion was common in Vietnamese strains. In addition, we found that the frequencies of the EPIYT and ESIYT motifs are relatively high among the sequences of the Okinawa strains (Xia et al., 2009). Amerindian CagA from AM-I also contained ESIYT motifs, which supports the possibility that these populations share the same origin (Suzuki et al., 2011a). A recent study revealed the recombination processes of *cagA* (Furuta et al., 2011a). Interestingly, the left half of the EPIYA-D segment of East Asian-type CagA was derived from the Western-type EPIYA, with the Amerindian-type EPIYA as intermediate, through rearrangement of specific sequences within the gene. J-Western-type EPIYA is phylogenetically located between Western-type EPIYA and Amerindian-type EPIYA. This finding suggests that the original *H. pylori* strain had a Western-type CagA sequence. Subsequently, they evolved to the J-Western-type CagA, Amerindian CagA, and then to the East Asian-type CagA.

3.2 MLST

MLST was proposed in 1998 as a tool for the epidemiological study of bacteria (Maiden et al., 1998). Recently, the genomic diversity within *H. pylori* populations was examined by employing the MLST method using 7 housekeeping genes (*atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, and *yphC*) (Figure 3) (Falush et al., 2003; Linz et al., 2007; Moodley et al., 2009). According to a study examining 769 isolates (Linz et al., 2007), the sequences from the 7 housekeeping gene fragments are concatenated to form a 3,406-bp haplotype, of which 45% (1,552 bp) is polymorphic, indicating that the polymorphic loci in the sequences can theoretically yield 4^{1552} distinct combinations. MLST analysis is reported to give more detailed information about human population structure than the method using human microsatellite or mitochondrial DNA (Wirth et al., 2004). At present, *H. pylori* strains can be divided into 7 population types on the basis of geographical associations and designated as follows: hpEurope, hpEastAsia, hpAfrica1, hpAfrica2, hpAsia2, hpNEAfrica, and hpSahul (Falush et al., 2003; Linz et al., 2007; Moodley et al., 2009). hpEurope includes almost all *H. pylori* strains isolated from ethnic Europeans, including people from countries colonized

by Europeans. hpEastAsia is common in *H. pylori* isolates from East Asia. hpEastAsia also includes subpopulations, ie, hspMaori (Polynesians, Melanesians, and native Taiwanese), hspAmerind (Amerindians), and hspEAsia (East Asians). hpAsia2 strains have been isolated in South, Southeast, and Central Asia. hpAfrica1 includes 2 subpopulations, hspWAfrica (West Africans, South Africans, and Afro-Americans) and hspSAfrica (South Africans); hpAfrica2 is very distinct and has only been isolated in South Africa. hpNEAfrica is predominant in isolates from Northeast Africa. hpSahul strains are isolated from aborigines of Australia and highlanders in New Guinea. All these modern populations derived from 6 ancestral populations which were designated ancestral European 1 (AE1), ancestral European 2 (AE2), ancestral EastAsia, ancestral Africa1, ancestral Africa2 (Falush et al., 2003), and ancestral Sahul (Moodley et al., 2009). *H. pylori* is predicted to have spread from East Africa over the same time period as anatomically modern humans (~58,000 years ago), and has remained intimately associated with their human hosts ever since (Linz et al., 2007; Moodley et al., 2009; Yamaoka, 2010).

Recently, MLST analyses were conducted in several regions such as India, Malaysia, China, Iran, and Cambodia (Devi et al., 2007; Latifi-Navid et al., 2010; Liao et al., 2009; Tay et al., 2009). In India, most strains initially belonged to hpAsia2 (Linz et al., 2007), whereas some strains belonged to hpEurope (Devi et al., 2007). Likewise, *H. pylori* from Malaysian Indians, known to have originated from India, consisted of hpEurope and hpAsia2 (Tay et al., 2009). *H. pylori* in the Indian population is more heterogeneous in origin, reflecting perhaps both earlier common ancestry and recent imports. It is notable that hpAsia2 strains from Ladakh Indians and Malaysian Indians can be divided into 2 subpopulations, hspLadakh and hspIndia (Tay et al., 2009). This may reflect regional differences in India as the Malaysian Indians mainly came from South India. Human migrations in Southeast Asia have also been clarified on the basis of MLST analyses from Cambodia (Breurec et al., 2011). MLST analyses from Iran also provided evidence that *H. pylori* strains from Iran are similar to other isolates from Western Eurasia and can be placed in the previously described hpEurope population (Latifi-Navid et al., 2010).

Our previous data showed that 4 strains isolated from the Ainu ethnic group, living in Hokkaido, a northern island of Japan, belong to the hspAmerind population (Gressmann et al., 2005b). Japanese aboriginal people, known as Jomon people, are thought to have migrated to the northern or southern area such as Hokkaido and Okinawa because of the immigration of Yayoi people from the Korean Peninsula (Ishida and Hinuma, 1986). Further analyses using isolates from Okinawa will clarify the human migration in Japan; these analyses are now in progress.

The relationships between the phylogeny of housekeeping genes and *cag* PAI or *VacA* phylogeny were reported (Gangwer et al., 2010; Olbermann et al., 2010). The phylogeny of most *cag* PAI genes was similar to that of MLST, indicating that *cag* PAI was probably acquired only once by *H. pylori*, and its genetic diversity reflects the isolation by distance which has shaped this bacterial species since modern humans migrated out of Africa (Olbermann et al., 2010). On the other hand, the overall topology of the *VacA* tree did not always match with that of MLST (Gangwer et al., 2010). Furthermore, rooting the *VacA* tree with outgroup sequences from the closely related *H. acinonychis* revealed that the ancestry of *VacA* is different from the African origin.

More interestingly, distribution of the incidence of gastric cancer is closely related to these *H. pylori* groups by MLST. A high incidence of gastric cancer was found in the regions prevailed by hpEastAsia strains (especially hspEAsia). On the other hand, the incidence of gastric cancer is very low in Africa, where most strains are hpNEAfrica, hpAfrica1, or hpAfrica2, and in South Asia, where most strains are hpAsia2. Overall, the African and

Asian enigmas might be explained, at least in part, by the different genotypes of *H. pylori* circulating in different geographic areas. Intriguingly, a recent report on *cagA*-positive strains in Colombia showed that all strains from high-risk regions of gastric cancer are hpEurope, whereas hpEurope and hpAfrica1 coexist in the low-risk regions (de Sablet et al., 2011). In addition, subjects infected with hpEurope strains of *H. pylori* showed higher histopathological scores than those infected with hpAfrica1 strains. They concluded that the phylogeographic origin by MLST can be used as a predictor of gastric cancer risk. Nevertheless, all *cagA*-negative strains belonged to hpEurope in their study. hpEurope strains without the presence of *cagA* can be less virulent. Thus, the *cagA* genotype rather than the phylogeographic origin is a better predictive factor of gastric cancer (Shiota et al., 2011).

4. Utilization of genome-wide information

Analysis of MLST data and virulence factors revealed much information about the pathogenicity and genealogy of *H. pylori*; however, these approaches focus on a small number of genes and may miss information conveyed by the rest of the genome. Genome-wide analyses using DNA microarray or whole-genome sequencing technology give a broad view on the genome of *H. pylori*.

Microarray analysis provides comprehensive information about gene contents of different strains and helps identify strain-specific genes as well as core genes shared by multiple strains. Gressman et al. investigated 56 strains of *H. pylori* and 4 strains of *H. acinonychis* by using whole genome microarrays (Gressmann et al., 2005a). They found that phylogenetic trees based on microarray data differed from those based on sequences of 7 genes from the core genome. They inferred that the phylogeny constructed by microarray data might be distorted due to homoplasies resulted from independent gene loss in multiple strains. Salama et al. examined the genomic content of 15 clinical isolates using a whole-genome DNA microarray and defined 1,281 genes as functional core genes (Salama et al., 2000). They identified candidates of virulence genes on the basis of coinheritance with PAI. A similar approach was used to elucidate the genomic diversity of isolates obtained from clinical patients in China (Han et al., 2007).

The whole-genome sequencing technology is another powerful tool to study the evolution and pathogenicity of *H. pylori*. Since the first release of the whole genome of strain 26695 (Tomb et al., 1997), the sequences of more than 20 genomes were determined by Sanger sequencing or the massively parallel sequencing technology. Accumulation of whole-genome data enables extensive sequence analyses of *H. pylori* strains. About 1,200 core genes were identified by comparison of peptic ulcer strain P12 and 6 other *H. pylori* genomes, which were in agreement with preceding studies (Fischer et al., 2010). The authors found that the P12 genome contains 3 plasticity zones and that 1 of them is capable of self-excision and horizontal transfer by conjugation. Their result suggests that conjugative transfer of genomic islands may contribute to the genetic diversity of *H. pylori*. Lara-Ramirez et al. analyzed genomes of 9 strains focusing on microevolution such as inversion, homologous recombination, and genesis of pseudogenes through homopolynucleotide mutations (Lara-Ramirez et al., 2011). They suggested that homopolynucleotide mutations are reversible and facilitate the control of gene expression through the change of DNA structure. Mobility of these pseudogenes within or between strains implies the possibility of pseudogenes as a reservoir of adaptation materials.

The detection of horizontal gene transfer is also empowered by whole-gnome analyses. Comprehensive gene-by-gene comparison was performed using genomes of strain J99 and strain 26695 (Saunders et al., 2005). The result indicated that a wide range of virulence

factors and strain-specific genes were acquired by gene transfer and that core metabolic genes such as *ftsK*, *xerD*, and *polA* were also exchanged.

Complete genomes of closely related strains provide a good basis to investigate changes in genome organization. Furuta et al. analyzed 9 whole genomes of *H. Pylori* strains of various geographical origin and investigated gain and loss of outer membrane genes (Furuta et al., 2011b). Sequence comparison revealed DNA duplication mechanisms associated with inversion. This finding helps understanding how genome organization changed through inversion events in the course of evolution.

5. Detection of genomic changes for clinical and evolutionary studies

Whole-genome analyses are also useful for the investigation of genetic factors related to differences in the virulence among strains. McClain et al. compared genome sequences of an isolate obtained from a patient with gastric cancer (strain 98-10) and an isolate from a patient with gastric ulcer (strain B128) (McClain et al., 2009). Strain 98-10 was found to be closely related to East Asian strains, while strain B128 was related to European strains. They determined strain-specific genes of strain 98-10 as candidate genes associated with gastric cancer. East Asian strains are known for their stronger carcinogenicity compared to strains of other areas. Kawai et al. investigated the evolution of East Asian strains using 20 whole genomes of Japanese, Korean, Amerindian, European, and West African strains (Kawai et al., 2011). Phylogenetic analysis revealed a greater divergence between the East Asian strains and the European strains in genes related to virulence factors, outer membrane proteins, and lipopolysaccharide synthesis enzymes. They examined positively selected amino acid changes and mapped the identified residues on CagA, VacA, HomC, SotB, and MiaA proteins.

Genomic changes during infection have also been studied. The whole-genome sequence of strain HPAG1 was determined by using the whole-genome shotgun method and the data obtained were used to design a custom microarray (Oh et al., 2006). Genotyping of isolates obtained from patients with chronic atrophic gastritis revealed gained and lost genes during progression of the disease, and whole-genome transcriptional profiling identified genes associated with the adaptation of *H. pylori* to chronic atrophic gastritis.

Chronological comparison of the whole genome was also done for 5 sets of *H. pylori* strains obtained from Colombian patients with isolation intervals of 3 to 16 years using the 454 sequencing technology (Kennemann et al., 2011). Comparison of the genomes revealed single-nucleotide polymorphisms and imported clusters that resulted from recombination, which is frequently found in members of the *hop* family.

Not necessarily the whole genome, but sequences of substantial length can provide important evolutionary information. Morelli et al. sequenced an average of 39,300 bp from 97 isolates, including sequential samples that were sampled at intervals of 0.25 to 10.2 years (Morelli et al., 2010). They observed that sequence diversity of the isolates increased in a clock-like manner and estimated rates of mutation, recombination, and mean length of recombination tracts. In this study, the short-term mutation rate was estimated to be 1.4×10^{-6} (serial isolates) and 4.5×10^{-6} (family isolates) per nucleotide per year.

Whole-genome information also revealed relationships between *H. pylori* and other *Helicobacter* species. It was enigmatic that one of the closest relatives of *H. pylori* was *H. acinonychis*, bacteria that colonize the gastric mucosa of large felines like cheetahs, lions, and tigers. Eppinger et al. determined the genomic sequence of the *H. acinonychis* strain Sheeba and compared it with *H. pylori* (Eppinger et al., 2006). The conserved core genes of the species were so similar that they estimated that the host jump from humans to felines

occurred within the last 200,000 years. The whole-genome sequence of *H. felis*, a parasite of dogs and cats, was also determined and compared with that of *H. pylori* (Arnold et al., 2011). Many virulence factors of *H. pylori*, including GGT, NapA, and HtrA, were also found in *H. felis*. The genome of *H. felis* lacks a *cag* PAI and *vacA* but contains the ComB system. These studies provide information on the origin of *H. pylori* and how virulence factors were acquired.

6. Impact of advanced sequencing technology

The rapid advances in sequencing technology enable massive sequence comparison (Suzuki et al., 2011b). One of the prospective applications of the new technology to the study of *H. pylori* is the identification of novel virulence factors (Fischer et al., 2010; Kawai et al., 2011; McClain et al., 2009). Another practical application is the detection of genomic changes related to drug resistance by comparing the genomes of wild type strains and those that survived antibiotic treatments. If a reference genome is available, it is not so difficult to re-sequence the target genome and to map the sequence reads produced by a next-generation sequencer onto the reference genome. In contrast, de novo assembly of complete genomes is more challenging. The current mechanism of massively parallel sequencing produces huge numbers of short sequence reads (tens to hundreds base pairs). It is computationally challenging to assemble short reads to the whole genome without a reference. Bioinformatics analysis is also required for gene prediction and annotation in the next step.

Currently, we took advantage of next-generation sequencers to read genomic sequences of more than 40 *H. pylori* strains mainly from Asian populations and attempted de novo assembly (unpublished observation). Although we cannot determine the whole genomes yet, we could construct a substantial size of contigs and predicted 1,200 to 1,500 genes for each strain. Using these data, we determined orthologous genes among our samples and strains whose whole genomes were released into public databases. A phylogenetic tree constructed by concatenated sequences of the orthologous genes showed more reliable results than a phylogenetic tree constructed by using MLST data (Figure 4). Compared with the tree based on MLST data (Figure 4A), the tree constructed by using concatenated genes (Figure 4B) showed better branching with higher bootstrap values between hpEurope and hpAsia2, as well as between hspEAsia and hspAmerind. Data obtained by using the massively parallel sequencing technology provide valuable information on the genealogy of *H. pylori* strains, as well as on candidates of drug resistance genes and new virulence factors.

The massively parallel sequencing technology is also applicable for the study of the transcriptome. Sharma et al. used a differential approach (dRNA-seq) selective for the 5'-end of transcripts and investigated the primary transcriptome of *H. pylori* strain 26695 (Sharma et al., 2010). They mapped transcriptional start sites and operons onto the genome and identified hundreds of transcriptional start sites within the operons. They also discovered ~60 small RNAs and potential regulators of cis- and trans-coded target messenger RNAs.

The sequencing technology is still advancing. We believe that larger amounts of data will become available at lower costs in the near future. As the amount of genomic data increases, analytic methods that can process them will be required. New sequencing technologies enable us to study organisms of larger genomes, including humans. Currently, genetic susceptibility factors of the human host are studied on the basis of individual genes (Correa, 2005; Kawai et al., 2006; Shanks and El-Omar, 2009), but the new technologies will enhance the identification of host genetic factors. Progress of studies on both human and microbe genomics will unveil conflicts and co-evolution between the host and the parasites.

7. Conclusion

The investigation of virulence factors revealed that specific genes and regions such as *cagA*, *vacA*, and the right end of the *cag* PAI are related to the clinical outcome. Population genetic studies based on MLST analysis help predict pre-historic human migration “accompanied” by *H. pylori*. Advances in sequencing technologies enable us to exert genome-wide analyses on both bacteria and hosts.

Acknowledgments

This report is based on work supported in part by grants from the National Institutes of Health (DK62813), grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan (22390085 and 22659087), Special Coordination Funds for Promoting Science and Technology from MEXT of Japan, and a Research Fund at the Discretion of the President, Oita University.

References

- Schistosomes, liver flukes and *Helicobacter pylori*. IARC Monogr Eval Carcinog Risks Hum; IARC Working Group on the Evaluation of Carcinogenic Risks to Humans; Lyon. 7-14 June 1994; 1994. p. 1-241.
- [access in 2010a] Center for Cancer Control and Information Services. National Cancer Center. Japan [on line], <http://www.ncc.go.jp/>
- [access in 2010b] International Agency for Research on Cancer. The Globocan Project. [online], <http://globocan.iarc.fr/>
- [access in 2010c] Okinawa Prefecture Home Page. [on line] <http://www.pref.okinawa.jp/annai/index.html>
- Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA, van der Ende A, van Doorn LJ. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol*. 1999; 32:459–470. [PubMed: 10320570]
- Argent R, Kidd M, Owen R, Thomas R, Limb M, Atherton J. Determinants and consequences of different levels of CagA phosphorylation for clinical isolates of *Helicobacter pylori*. *Gastroenterology*. 2004; 127:514–523. [PubMed: 15300584]
- Arnold IC, Zigova Z, Holden M, Lawley TD, Rad R, Dougan G, Falkow S, Bentley SD, Muller A. Comparative whole genome sequence analysis of the carcinogenic bacterial model pathogen *Helicobacter felis*. *Genome Biol Evol*. 2011; 3:302–308. [PubMed: 21402865]
- Asahi M, Azuma T, Ito S, Ito Y, Suto H, Nagai Y, Tsubokawa M, Tohyama Y, Maeda S, Omata M, Suzuki T, Sasakawa C. *Helicobacter pylori* CagA protein can be tyrosine phosphorylated in gastric epithelial cells. *J Exp Med*. 2000; 191:593–602. [PubMed: 10684851]
- Atherton J, Cao P, Peek RJ, Tummuru M, Blaser M, Cover T. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. *J Biol Chem*. 1995; 270:17771–17777. [PubMed: 7629077]
- Atherton JC. The pathogenesis of *Helicobacter pylori*-induced gastro-duodenal diseases. *Annu Rev Pathol*. 2006; 1:63–96. [PubMed: 18039108]
- Azuma T, Yamakawa A, Yamazaki S, Fukuta K, Ohtani M, Ito Y, Dojo M, Yamazaki Y, Kuriyama M. Correlation between variation of the 3' region of the *cagA* gene in *Helicobacter pylori* and disease outcome in Japan. *J Infect Dis*. 2002; 186:1621–1630. [PubMed: 12447739]
- Azuma T, Yamakawa A, Yamazaki S, Ohtani M, Ito Y, Muramatsu A, Suto H, Yamazaki Y, Keida Y, Higashi H, Hatakeyama M. Distinct diversity of the *cag* pathogenicity island among *Helicobacter pylori* strains in Japan. *J Clin Microbiol*. 2004; 42:2508–2517. [PubMed: 15184428]
- Backert S, Selbach M. Role of type IV secretion in *Helicobacter pylori* pathogenesis. *Cell Microbiol*. 2008; 10:1573–1581. [PubMed: 18410539]
- Batista SA, Rocha GA, Rocha AM, Saraiva IE, Cabral MM, Oliveira RC, Queiroz DM. Higher number of *Helicobacter pylori* CagA EPIYA C phosphorylation sites increases the risk of gastric cancer, but not duodenal ulcer. *BMC Microbiol*. 2011; 11:61. [PubMed: 21435255]

- Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, Huerre M, Oung C, Raymond J, Sreng Tan T, Thiberge JM, Vong S, Monchy D, Linz B. Evolutionary History of *Helicobacter pylori* Sequences Reflect Past Human Migrations in Southeast Asia. *PLoS One*. 2011; 6:e22058. [PubMed: 21818291]
- Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R, Covacci A. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci U S A*. 1996; 93:14648–14653. [PubMed: 8962108]
- Correa P. New strategies for the prevention of gastric cancer: *Helicobacter pylori* and genetic susceptibility. *J Surg Oncol*. 2005; 90:134–138. discussion 138. [PubMed: 15895454]
- Covacci A, Censini S, Bugnoli M, Petracca R, Burroni D, Macchia G, Massone A, Papini E, Xiang Z, Figura N. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proc Natl Acad Sci U S A*. 1993; 90:5791–5795. [PubMed: 8516329]
- Cover TL, Blanke SR. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. *Nat Rev Microbiol*. 2005; 3:320–332. [PubMed: 15759043]
- Cover TL, Blaser MJ. Purification and characterization of the vacuolating toxin from *Helicobacter pylori*. *J Biol Chem*. 1992; 267:10570–10575. [PubMed: 1587837]
- de Sablet T, Piazzuelo MB, Shaffer CL, Schneider BG, Asim M, Chaturvedi R, Bravo LE, Sicinski LA, Delgado AG, Mera RM, Israel DA, Romero-Gallo J, Peek RM, Cover TL, Correa P, Wilson KT. Phylogeographic origin of *Helicobacter pylori* is a determinant of gastric cancer risk. *Gut*. 2011; 60:1189–1195. [PubMed: 21357593]
- Devi SM, Ahmed I, Francalacci P, Hussain MA, Akhter Y, Alvi A, Sechi LA, Mégraud F, Ahmed N. Ancestral European roots of *Helicobacter pylori* in India. *BMC Genomics*. 2007; 8:184. [PubMed: 17584914]
- Domínguez-Bello MG, Pérez ME, Bortolini MC, Salzano FM, Pericchi LR, Zambrano-Guzmán O, Linz B. Amerindian *Helicobacter pylori* strains go extinct, as european strains expand their host range. *PLoS One*. 2008; 3:e3307. [PubMed: 18830403]
- Eppinger M, Baar C, Linz B, Raddatz G, Lanz C, Keller H, Morelli G, Gressmann H, Achtman M, Schuster SC. Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genet*. 2006; 2:e120. [PubMed: 16789826]
- Falush D, Wirth T, Linz B, Pritchard J, Stephens M, Kidd M, Blaser M, Graham D, Vacher S, Perez-Perez G, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S. Traces of human migrations in *Helicobacter pylori* populations. *Science*. 2003; 299:1582–1585. [PubMed: 12624269]
- Fischer W, Windhager L, Rohrer S, Zeiller M, Karnholz A, Hoffmann R, Zimmer R, Haas R. Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer. *Nucleic Acids Res*. 2010; 38:6089–6101. [PubMed: 20478826]
- Franco AT, Israel DA, Washington MK, Krishna U, Fox JG, Rogers AB, Neish AS, Collier-Hyams L, Perez-Perez GI, Hatakeyama M, Whitehead R, Gaus K, O'Brien DP, Romero-Gallo J, Peek RM. Activation of beta-catenin by carcinogenic *Helicobacter pylori*. *Proc Natl Acad Sci U S A*. 2005; 102:10646–10651. [PubMed: 16027366]
- Franco AT, Johnston E, Krishna U, Yamaoka Y, Israel DA, Nagy TA, Wroblewski LE, Piazzuelo MB, Correa P, Peek RM. Regulation of gastric carcinogenesis by *Helicobacter pylori* virulence factors. *Cancer Res*. 2008; 68:379–387. [PubMed: 18199531]
- Furuta Y, Yahara K, Hatakeyama M, Kobayashi I. Evolution of *cagA* oncogene of *Helicobacter pylori* through recombination. *PLoS One*. 2011a; 6:e23499. [PubMed: 21853141]
- Furuta Y, Kawai M, Yahara K, Takahashi N, Handa N, Tsuru T, Oshima K, Yoshida M, Azuma T, Hattori M, Uchiyama I, Kobayashi I. Birth and death of genes linked to chromosomal inversion. *Proc Natl Acad Sci U S A*. 2011b; 108:1501–1506. [PubMed: 21212362]
- Gangwer KA, Shaffer CL, Suerbaum S, Lacy DB, Cover TL, Bordenstein SR. Molecular evolution of the *Helicobacter pylori* vacuolating toxin gene *vacA*. *J Bacteriol*. 2010; 192:6126–6135. [PubMed: 20870762]

- Ghose C, Perez-Perez GI, Dominguez-Bello MG, Pride DT, Bravi CM, Blaser MJ. East Asian genotypes of *Helicobacter pylori* strains in Amerindians provide evidence for its ancient human carriage. *Proc Natl Acad Sci U S A*. 2002; 99:15107–15111. [PubMed: 12417749]
- Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, Yamaoka Y, Kraft C, Suerbaum S, Meyer TF, Achtman M. Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet*. 2005a; 1:e43. [PubMed: 16217547]
- Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, Yamaoka Y, Kraft C, Suerbaum S, Meyer TF, Achtman M. Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet*. 2005b; 1:e43. [PubMed: 16217547]
- Han YH, Liu WZ, Shi YZ, Lu LQ, Xiao S, Zhang QH, Zhao GP. Comparative genomics profiling of clinical isolates of *Helicobacter pylori* in Chinese populations using DNA microarray. *J Microbiol*. 2007; 45:21–28. [PubMed: 17342051]
- Hatakeyama M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nat Rev Cancer*. 2004; 4:688–694. [PubMed: 15343275]
- Ishida T, Hinuma Y. The origin of Japanese HTLV-I. *Nature*. 1986; 322:504. [PubMed: 3016556]
- Ito S, Azuma T, Murakita H, Hirai M, Miyaji H, Ito Y, Ohtaki Y, Yamazaki Y, Kuriyama M, Keida Y, Kohli Y. Profile of *Helicobacter pylori* cytotoxin derived from two areas of Japan with different prevalence of atrophic gastritis. *Gut*. 1996; 39:800–806. [PubMed: 9038660]
- Jones KR, Jang S, Chang JY, Kim J, Chung IS, Olsen CH, Merrell DS, Cha JH. Polymorphisms in the intermediate region of VacA impact *Helicobacter pylori*-induced disease development. *J Clin Microbiol*. 2011; 49:101–110. [PubMed: 21084502]
- Jones KR, Joo YM, Jang S, Yoo YJ, Lee HS, Chung IS, Olsen CH, Whitmire JM, Merrell DS, Cha JH. Polymorphism in the CagA EPIYA motif impacts development of gastric cancer. *J Clin Microbiol*. 2009; 47:959–968. [PubMed: 19158258]
- Judd LM, Alderman BM, Howlett M, Shulkes A, Dow C, Moverley J, Grail D, Jenkins BJ, Ernst M, Giraud AS. Gastric cancer development in mice lacking the SHP2 binding site on the IL-6 family co-receptor gp130. *Gastroenterology*. 2004; 126:196–207. [PubMed: 14699500]
- Jung SW, Sugimoto M, Graham DY, Yamaoka Y. *homb* status of *Helicobacter pylori* as a novel marker to distinguish gastric cancer from duodenal ulcer. *J Clin Microbiol*. 2009; 47:3241–3245. [PubMed: 19710266]
- Kawai M, Furuta Y, Yahara K, Tsuru T, Oshima K, Handa N, Takahashi N, Yoshida M, Azuma T, Hattori M, Uchiyama I, Kobayashi I. Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes. *BMC Microbiol*. 2011; 11:104. [PubMed: 21575176]
- Kawai S, Goto Y, Ito LS, Oba-Shinjo SM, Uno M, Shinjo SK, Marie SK, Ishida Y, Nishio K, Naito M, Hamajima N. Significant association between PTPN11 polymorphism and gastric atrophy among Japanese Brazilians. *Gastric Cancer*. 2006; 9:277–283. [PubMed: 17235629]
- Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, Reinhardt R, Correa P, Meyer TF, Josenhans C, Falush D, Suerbaum S. *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci U S A*. 2011; 108:5033–5038. [PubMed: 21383187]
- Kersulyte D, Kalia A, Gilman RH, Mendez M, Herrera P, Cabrera L, Velapatino B, Balqui J, Paredes Puente de la Vega F, Rodriguez Ulloa CA, Cok J, Hooper CC, Dailide G, Tamma S, Berg DE. *Helicobacter pylori* from Peruvian amerindians: traces of human migrations in strains from remote Amazon, and genome sequence of an Amerind strain. *PLoS One*. 2010; 5:e15076. [PubMed: 21124785]
- Kersulyte D, Mukhopadhyay AK, Velapatiño B, Su W, Pan Z, Garcia C, Hernandez V, Valdez Y, Mistry RS, Gilman RH, Yuan Y, Gao H, Alarcón T, López-Brea M, Balakrish Nair G, Chowdhury A, Datta S, Shirai M, Nakazawa T, Ally R, Segal I, Wong BC, Lam SK, Olfat FO, Borén T, Engstrand L, Torres O, Schneider R, Thomas JE, Czinn S, Berg DE. Differences in genotypes of *Helicobacter pylori* from different human populations. *J Bacteriol*. 2000; 182:3210–3218. [PubMed: 10809702]
- Kusters JG, van Vliet AH, Kuipers EJ. Pathogenesis of *Helicobacter pylori* infection. *Clin Microbiol Rev*. 2006; 19:449–490. [PubMed: 16847081]

- Lara-Ramirez EE, Segura-Cabrera A, Guo X, Yu G, Garcia-Perez CA, Rodriguez-Perez MA. New implications on genomic adaptation derived from the *Helicobacter pylori* genome comparison. PLoS One. 2011; 6:e17300. [PubMed: 21387011]
- Latifi-Navid S, Ghorashi SA, Siavoshi F, Linz B, Massarrat S, Kheday T, Salmanian AH, Shayesteh AA, Masoodi M, Ghanadi K, Ganji A, Suerbaum S, Achtman M, Malekzadeh R, Falush D. Ethnic and geographic differentiation of *Helicobacter pylori* within Iran. PLoS One. 2010; 5:e9645. [PubMed: 20339588]
- Letley DP, Lastovica A, Louw JA, Hawkey CJ, Atherton JC. Allelic diversity of the *Helicobacter pylori* vacuolating cytotoxin gene in South Africa: rarity of the *vacA* s1a genotype and natural occurrence of an s2/m1 allele. J Clin Microbiol. 1999; 37:1203–1205. [PubMed: 10074554]
- Leunk RD. Production of a cytotoxin by *Helicobacter pylori*. Rev Infect Dis. 1991; 13(Suppl 8):S686–689. [PubMed: 1925309]
- Li WH, Sadler LA. Low nucleotide diversity in man. Genetics. 1991; 129:513–523. [PubMed: 1743489]
- Liao YL, Guo G, Mao XH, Xie QH, Zhang WJ, Liu XF, Zou QM. Core genome haplotype diversity and *vacA* allelic heterogeneity of Chinese *Helicobacter pylori* strains. Curr Microbiol. 2009; 59:123–129. [PubMed: 19452214]
- Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe S, Yamaoka Y, Graham D, Perez-Trallero E, Wadstrom T, Suerbaum S, Achtman M. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 2007; 445:915–918. [PubMed: 17287725]
- Lu H, Hsu PI, Graham DY, Yamaoka Y. Duodenal ulcer promoting gene of *Helicobacter pylori*. Gastroenterology. 2005; 128:833–848. [PubMed: 15825067]
- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 1998; 95:3140–3145. [PubMed: 9501229]
- Malaty H. Epidemiology of *Helicobacter pylori* infection. Best Pract Res Clin Gastroenterol. 2007; 21:205–214. [PubMed: 17382273]
- McClain MS, Shaffer CL, Israel DA, Peek RM Jr, Cover TL. Genome sequence analysis of *Helicobacter pylori* strains associated with gastric ulceration and gastric cancer. BMC Genomics. 2009; 10:3. [PubMed: 19123947]
- Miura M, Ohnishi N, Tanaka S, Yanagiya K, Hatakeyama M. Differential oncogenic potential of geographically distinct *Helicobacter pylori* CagA isoforms in mice. Int J Cancer. 2009; 125:2497–2504. [PubMed: 19588494]
- Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY, Maady A, Bernhöft S, Thiberge JM, Phuanukoonnon S, Jobb G, Siba P, Graham DY, Marshall BJ, Achtman M. The peopling of the Pacific from a bacterial perspective. Science. 2009; 323:527–530. [PubMed: 19164753]
- Morelli G, Didelot X, Kusecek B, Schwarz S, Bahlawane C, Falush D, Suerbaum S, Achtman M. Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. PLoS Genet. 2010; 6:e1001036. [PubMed: 20661309]
- Mukhopadhyay AK, Kersulyte D, Jeong JY, Datta S, Ito Y, Chowdhury A, Chowdhury S, Santra A, Bhattacharya SK, Azuma T, Nair GB, Berg DE. Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. J Bacteriol. 2000; 182:3219–3227. [PubMed: 10809703]
- Nguyen TL, Uchida T, Tsukamoto Y, Trinh DT, Ta L, Mai BH, Le SH, Thai KD, Ho DD, Hoang HH, Matsuhisa T, Okimoto T, Kodama M, Murakami K, Fujioka T, Yamaoka Y, Moriyama M. *Helicobacter pylori* infection and gastroduodenal diseases in Vietnam: a cross-sectional, hospital-based study. BMC Gastroenterol. 2010; 10:114. [PubMed: 20920280]
- Nobuta A, Asaka M, Sugiyama T, Kato M, Hige S, Takeda H, Kato T, Ogoshi K, Keida Y, Shinomura J. *Helicobacter pylori* infection in two areas in Japan with different risks for gastric cancer. Aliment Pharmacol Ther. 2004; 20(Suppl 1):1–6. [PubMed: 15298598]
- Ogiwara H, Graham DY, Yamaoka Y. *vacA* i-region subtyping. Gastroenterology. 2008; 134:1267. author reply 1268. [PubMed: 18395110]

- Ogiwara H, Sugimoto M, Ohno T, Vilaichone RK, Mahachai V, Graham DY, Yamaoka Y. Role of deletion located between the intermediate and middle regions of the *Helicobacter pylori vacA* gene in cases of gastroduodenal diseases. *J Clin Microbiol.* 2009; 47:3493–3500. [PubMed: 19726606]
- Oh JD, Kling-Backhed H, Giannakis M, Xu J, Fulton RS, Fulton LA, Cordum HS, Wang C, Elliott G, Edwards J, Mardis ER, Engstrand LG, Gordon JI. The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc Natl Acad Sci U S A.* 2006; 103:9999–10004. [PubMed: 16788065]
- Ohnishi N, Yuasa H, Tanaka S, Sawa H, Miura M, Matsui A, Higashi H, Musashi M, Iwabuchi K, Suzuki M, Yamada G, Azuma T, Hatakeyama M. Transgenic expression of *Helicobacter pylori* CagA induces gastrointestinal and hematopoietic neoplasms in mouse. *Proc Natl Acad Sci U S A.* 2008; 105:1003–1008. [PubMed: 18192401]
- Ohno T, Sugimoto M, Nagashima A, Ogiwara H, Vilaichone RK, Mahachai V, Graham DY, Yamaoka Y. Relationship between *Helicobacter pylori* hopQ genotype and clinical outcome in Asian and Western populations. *J Gastroenterol Hepatol.* 2009; 24:462–468. [PubMed: 19226380]
- Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, Vauterin M, Suerbaum S, Achtman M, Linz B. A Global Overview Of The Genetic And Functional Diversity In The *Helicobacter Pylori* Cag Pathogenicity Island. *PLoS Genet.* 2010; 6:e1001069. [PubMed: 20808891]
- Rhead JL, Letley DP, Mohammadi M, Hussein N, Mohagheghi MA, Eshagh Hosseini M, Atherton JC. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology.* 2007; 133:926–936. [PubMed: 17854597]
- Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci U S A.* 2000; 97:14668–14673. [PubMed: 11121067]
- Satomi S, Yamakawa A, Matsunaga S, Masaki R, Inagaki T, Okuda T, Suto H, Ito Y, Yamazaki Y, Kuriyama M, Keida Y, Kutsumi H, Azuma T. Relationship between the diversity of the *cagA* gene of *Helicobacter pylori* and gastric cancer in Okinawa, Japan. *J Gastroenterol.* 2006; 41:668–673. [PubMed: 16933004]
- Saunders NJ, Boonmee P, Peden JF, Jarvis SA. Inter-species horizontal transfer resulting in core-genome and niche-adaptive variation within *Helicobacter pylori*. *BMC Genomics.* 2005; 6:9. [PubMed: 15676066]
- Shanks AM, El-Omar EM. *Helicobacter pylori* infection, host genetics and gastric cancer. *J Dig Dis.* 2009; 10:157–164. [PubMed: 19659782]
- Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature.* 2010; 464:250–255. [PubMed: 20164839]
- Shiota S, Matsunari O, Watada M, Yamaoka Y. Virulence factors or ancestral origin of *Helicobacter pylori*: which is a better predictor of gastric cancer risk? *Gut.* 2011 Epub ahead of print.
- Shiota S, Matsunari O, Yamaoka Y. Relationship between J-Western CagA subtype and the *vacA* m2 region of *Helicobacter pylori*. *J Clin Microbiol.* 2010; 48:3033–3034. [PubMed: 20519467]
- Suerbaum S, Josenhans C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat Rev Microbiol.* 2007; 5:441–452. [PubMed: 17505524]
- Suerbaum S, Michetti P. *Helicobacter pylori* infection. *N Engl J Med.* 2002; 347:1175–1186. [PubMed: 12374879]
- Sugimoto M, Yamaoka Y. The association of *vacA* genotype and *Helicobacter pylori*-related disease in Latin American and African populations. *Clin Microbiol Infect.* 2009; 15:835–842. [PubMed: 19392900]
- Sugimoto M, Zali MR, Yamaoka Y. The association of *vacA* genotypes and *Helicobacter pylori*-related gastroduodenal diseases in the Middle East. *Eur J Clin Microbiol Infect Dis.* 2009a; 28:1227–1236. [PubMed: 19551413]
- Sugimoto M, Zali MR, Yamaoka Y. The association of *vacA* genotypes and *Helicobacter pylori*-related gastroduodenal diseases in the Middle East. *Eur J Clin Microbiol Infect Dis.* 2009b; 28:1227–1236. [PubMed: 19551413]
- Suzuki M, Kiga K, Kersulyte D, Cok J, Hooper CC, Mimuro H, Sanada T, Suzuki S, Oyama M, Kozuka-Hata H, Kamiya S, Zou QM, Gilman RH, Berg DE, Sasakawa C. Attenuated CagA

oncoprotein in *Helicobacter pylori* from Amerindians in Peruvian Amazon. *J Biol Chem*. 2011a; 34:29964–29972.

- Suzuki S, Ono N, Furusawa C, Ying BW, Yomo T. Comparison of sequence reads obtained from three next-generation sequencing platforms. *PLoS One*. 2011b; 6:e19534. [PubMed: 21611185]
- Tay CY, Mitchell H, Dong Q, Goh KL, Dawes IW, Lan R. Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. *BMC Microbiol*. 2009; 9:126. [PubMed: 19538757]
- Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzgerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*. 1997; 388:539–547. [PubMed: 9252185]
- Truong B, Mai V, Tanaka H, Ly I.T, Thong T, Hai H, Van Long D, Furumatsu K, Yoshida M, Kutsumi H, Azuma T. Diverse characteristics of the CagA gene of *Helicobacter pylori* strains collected from patients from southern vietnam with gastric cancer and peptic ulcer. *J Clin Microbiol*. 2009; 47:4021–4028. [PubMed: 19846630]
- Tummuru MK, Cover TL, Blaser MJ. Cloning and expression of a high-molecular-mass major antigen of *Helicobacter pylori*: evidence of linkage to cytotoxin production. *Infect Immun*. 1993; 61:1799–1809. [PubMed: 8478069]
- Uchida T, Nguyen LT, Takayama A, Okimoto T, Kodama M, Murakami K, Matsuhisa T, Trinh TD, Ta L, Ho DQ, Hoang HH, Kishida T, Fujioka T, Moriyama M, Yamaoka Y. Analysis of virulence factors of *Helicobacter pylori* isolated from a Vietnamese population. *BMC Microbiol*. 2009; 9:175. [PubMed: 19698173]
- van Doorn L, Figueiredo C, Sanna R, Plaisier A, Schneeberger P, de Boer W, Quint W. Clinical relevance of the *cagA*, *vacA*, and *iceA* status of *Helicobacter pylori*. *Gastroenterology*. 1998; 115:58–66. [PubMed: 9649459]
- Van Doorn LJ, Figueiredo C, Mégraud F, Pena S, Midolo P, Queiroz DM, Carneiro F, Vanderborght B, Pegado MD, Sanna R, De Boer W, Schneeberger PM, Correa P, Ng EK, Atherton J, Blaser MJ, Quint WG. Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. *Gastroenterology*. 1999; 116:823–830. [PubMed: 10092304]
- Vilaichone RK, Mahachai V, Tumwasorn S, Wu JY, Graham DY, Yamaoka Y. Molecular epidemiology and outcome of *Helicobacter pylori* infection in Thailand: a cultural cross roads. *Helicobacter*. 2004a; 9:453–459. [PubMed: 15361085]
- Vilaichone RK, Mahachai V, Tumwasorn S, Wu JY, Graham DY, Yamaoka Y. Molecular epidemiology and outcome of *Helicobacter pylori* infection in Thailand: a cultural cross roads. *Helicobacter*. 2004b; 9:453–459. [PubMed: 15361085]
- Wirth T, Wang X, Linz B, Novick RP, Lum JK, Blaser M, Morelli G, Falush D, Achtman M. Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proc Natl Acad Sci U S A*. 2004; 101:4746–4751. [PubMed: 15051885]
- Xia Y, Yamaoka Y, Zhu Q, Matha I, Gao X. A comprehensive sequence and disease correlation analyses for the C-terminal region of CagA protein of *Helicobacter pylori*. *PLoS One*. 2009; 4:e7736. [PubMed: 19893742]
- Yamaoka Y. *Helicobacter pylori* typing as a tool for tracking human migration. *Clin Microbiol Infect*. 2009; 15:829–834. [PubMed: 19702588]
- Yamaoka Y. Mechanisms of disease: *Helicobacter pylori* virulence factors. *Nat Rev Gastroenterol Hepatol*. 2010; 7:629–641. [PubMed: 20938460]
- Yamaoka Y, El-Zimaity H, Gutierrez O, Figura N, Kim J, Kodama T, Kashima K, Graham D, Kim J. Relationship between the *cagA* 3' repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. *Gastroenterology*. 1999a; 117:342–349. [PubMed: 10419915]
- Yamaoka Y, El-Zimaity HM, Gutierrez O, Figura N, Kim JG, Kodama T, Kashima K, Graham DY. Relationship between the *cagA* 3' repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. *Gastroenterology*. 1999b; 117:342–349. [PubMed: 10419915]

- Yamaoka Y, Kikuchi S, el-Zimaity HM, Gutierrez O, Osato MS, Graham DY. Importance of *Helicobacter pylori oipA* in clinical presentation, gastric inflammation, and mucosal interleukin 8 production. *Gastroenterology*. 2002a; 123:414–424. [PubMed: 12145793]
- Yamaoka Y, Kodama T, Gutierrez O, Kim JG, Kashima K, Graham DY. Relationship between *Helicobacter pylori iceA*, *cagA*, and *vacA* status and clinical outcome: studies in four different countries. *J Clin Microbiol*. 1999c; 37:2274–2279. [PubMed: 10364597]
- Yamaoka Y, Kodama T, Kashima K, Graham D, Sepulveda A. Variants of the 3' region of the *cagA* gene in *Helicobacter pylori* isolates from patients with different H. pylori-associated diseases. *J Clin Microbiol*. 1998a; 36:2258–2263. [PubMed: 9666002]
- Yamaoka Y, Kodama T, Kashima K, Graham DY, Sepulveda AR. Variants of the 3' region of the *cagA* gene in *Helicobacter pylori* isolates from patients with different H. pylori-associated diseases. *J Clin Microbiol*. 1998b; 36:2258–2263. [PubMed: 9666002]
- Yamaoka Y, Kwon DH, Graham DY. A M(r) 34,000 proinflammatory outer membrane protein (*oipA*) of *Helicobacter pylori*. *Proc Natl Acad Sci U S A*. 2000a; 97:7533–7538. [PubMed: 10852959]
- Yamaoka Y, Orito E, Mizokami M, Gutierrez O, Saitou N, Kodama T, Osato MS, Kim JG, Ramirez FC, Mahachai V, Graham DY. *Helicobacter pylori* in North and South America before Columbus. *FEBS Lett*. 2002b; 517:180–184. [PubMed: 12062433]
- Yamaoka Y, Osato M, Sepulveda A, Gutierrez O, Figura N, Kim J, Kodama T, Kashima K, Graham D. Molecular epidemiology of *Helicobacter pylori*: separation of H. pylori from East Asian and non-Asian countries. *Epidemiol Infect*. 2000b; 124:91–96. [PubMed: 10722135]
- Yamazaki S, Yamakawa A, Okuda T, Ohtani M, Suto H, Ito Y, Yamazaki Y, Keida Y, Higashi H, Hatakeyama M, Azuma T. Distinct diversity of *vacA*, *cagA*, and *cagE* genes of *Helicobacter pylori* associated with peptic ulcer in Japan. *J Clin Microbiol*. 2005; 43:3906–3916. [PubMed: 16081930]

Highlights

- Geographic differences of gastric cancer incidence can be explained by *H. pylori*.
- *H. pylori* genotype is useful as a tool to track human migration.
- Next-generation sequencers enabled us to investigate evolution of *H. pylori*.
- Next-generation sequencers enabled us to investigate novel virulence factors.

Figure 1. *Helicobacter pylori* populations and phylogenetic tree

Phylogenetic tree constructed by using multilocus sequence typing (MLST) data (neighbor-joining [NJ]-tree, Kimura 2 parameter). Branch colors represent *H. pylori* populations determined by using the Bayesian clustering method. hspSAfrica and hspWAfrica are subpopulations of hpAfrica1 and hspAmerind, hspMaori, and hspEAsia are subpopulations of hpEastAsia.

Figure 2. Schematic representation of the *vacA* alleles

The *vacA* s1 genotype has a 27-bp deletion as compared to the s2 genotype. The *vacA* i region can be classified into 2 types (i1 and i2, shaded in orange and red, respectively) according to the amino acid sequences denoted as clusters A, B, and C (from left to right). The *vacA* d2 genotype has a 69- to 81-bp deletion as compared to the d1 genotype. The *vacA* m1 genotype has a 73-bp deletion as compared to the m2 genotype.

Figure 3. Geographic distribution of *Helicobacter pylori* populations and predicted traces of prehistoric human migration

Colored circles illustrate the putative distribution of *H. pylori* populations before the “Age of Exploration.” Black arrows and numbers represent predicted paths and times of migration.

Figure 4. Phylogenetic trees by using multilocus sequence typing (MLST) data and concatenated genes

Phylogenetic trees (neighbor-joining [NJ]-tree, Kimura 2 parameter, bootstrap 1000) (A) by using MLST data and (B) by using concatenated genes (~1 Mb). Red numbers indicate bootstrap values greater than 60%.

Table 1
Virulence genotypes on *cagA* and *vacA* in association with geographical area

Geographical area	virulence genes genotype	<i>cagA</i>	<i>cag</i> right end junction	<i>vacA</i> s1	<i>vacA</i> m1
East Asia	East Asia	East Asian	II	s1c	m1b
South/Central Asia	South/Central Asia		III (IV)		m1c
Europe(except for Iberian peninsula)	Europe	Western		s1a	
Europe(Iberian peninsula)	Iberian/Africa		I	s1b	m1a
Africa					