J. Microbiol. Biotechnol. (2005), 15(4), 844–854

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

© The Korean Society for Microbiology and Biotechnology

# *Helicobacter pylori* Strain 51 (Korean Isolate): Ordered Overlapping BAC Library, Combined Physical and Genetic Map, and Comparative Analysis with *H. pylori* Strain 26695 and Strain J99

KANG, HYUNG-LYUN[1,4], WOO-KON LEE[1,5], JAE-YOUNG SONG[1,4], SANG-HAENG CHOI[6], SEONG-GYU PARK[1], BOK-DEOK RYU[1], EUN-JOO LEE[1], JI-SUN KIM[6], JEONG-UCK PARK[1,4], SEUNG-CHUL BAIK[1,4], MYOUNG-BUM CHOI[3], HEE-SHANG YOUN[3], GYUNG-HYUCK KO[2], DONG-WON BAE[5], YONG-SUNG KIM[6], MYUNG-JE CHO[1,5], AND KWANG-HO RHEE[1]*

*Department of [1]Microbiology, [2]Pathology, and [3]Pediatrics, and [4]Gyeongsang Institute of Health Sciences, Gyeongsang National University College of Medicine, 90 Chiram-dong, Jinju, Gyeongsangnam-do 660-751, Korea*
*[5]Research Institute of Life Science and Central Laboratory, Gyeongsang National University, 900 Gajwa-dong, Jinju, Gyeongsangnam-do 660-751, Korea*
*[6]Korea Research Institute for Bioscience and Biotechnology, Taejon 305-333, Korea*

Received: December 15, 2004
Accepted: February 20, 2005

**Abstract** We constructed a defined physical and genetic map of *H. pylori* strain 51, previously isolated from a Korean patient with a duodenal ulcer, by combining a restriction analysis by pulse-field gel electrophoresis with the construction of a BAC library. A *Not*1-digest of *H. pylori* strain 51 genome yielded seven fragments, from which the genomic size was estimated to be 1,698±24 kb. The BAC library was constructed from 50 to 200 kb fragments of *Hin*dIII-digested genomic DNA. From 700 BAC clones, an ordered overlapping maxi-set of 82 BAC clones was assembled that covered the entire genome. The positions of 15 genes were localized in the strain 51 genome with 4–22 kb of resolution and were compared with their orthologues in strain 26695 and strain J99. The arrangement of the 15 genes was identical in strain 51 and strain J99, except for *flaA* and *hpaA*. The plasticity zone of strain 51, like that of strain J99, was located in the single region, and was shorter than those of strain 26695 and strain J99. The strain 51 plasticity zone consisted of ORFs common only to strain 51 and J99 or to strain 51 and 26695, as well as strain 51-specific ORFs. Three genetic translocations and/or inversions were found between orthologue ORFs in strain 51 and strain J99. These results show that the chromosomal organization of strain 51 differs from Western strains such as strain 26695 and strain J99.

**Key words:** *Helicobacter pylori*, genomics, physical map

*Corresponding author
Phone: 82-55-751-8746; Fax: 82-55-759-1588;
E-mail: khrhee@gaechuk.gsnu.ac.kr

*Helicobacter pylori* is a slow-growing, microaerophilic, Gram-negative bacterium that colonizes the gastric mucosa of humans [45]. *H. pylori* causes a longstanding and possibly lifelong infection in the gastric mucosa [10], resulting in various gastric maladies, including active chronic gastritis [8, 34], gastric atrophy [9], and gastroduodenal ulcers [8], as well as gastric cancers [7, 19]. About half of the world population and more than 90% of Korean adults are suffering from gastric maladies caused by *H. pylori* [34]. The strong association of *H. pylori* infection with gastric cancer and mucosa-associated lymphoid tissue lymphoma led to the classification of this organism as a class I carcinogen by the World Health Organization [5]. This classification was validated by the experimental production of adenocarcinoma in the Mongolian gerbil animal model [46].

The severity of gastric diseases caused by *H. pylori* ranges from asymptomatic to ulcerative or malignant [6, 11, 39], although most persons infected with *H. pylori* do not progress to serious diseases [31, 32]. An enormous effort has been made to develop either typing methods to differentiate strains from clinical isolates and to identify the virulent strains that cause serious symptoms [4, 29], or drugs to kill *H. pylori* [15, 26, 27]. A number of data obtained with genetic techniques have documented the extensive genetic heterogeneity of *H. pylori* isolates both at specific genetic loci and at the level of genomic organization [2, 23]. The mechanisms by which this genetic diversity is generated within *H. pylori* species have been

investigated, but remain poorly understood [16, 20, 25, 30, 38, 44].

Colonization of individuals by multiple strains is apparently not rare [21, 41]. Competition between these multiple strains within the gastric mucosa could theoretically lead to DNA mutation, transfer, and rearrangement between the different strains, potentially generating new genotypes that are more adapted to the host environment during chronic infection [25]. In addition to the genetic heterogeneity of *H. pylori*, intrafamilial infections [13] and a longstanding infection [10] could increase the probability of clonal grouping in different geographical regions. In fact, the allelic population differs between strains isolated from East Asians and from Westerners [1]. Allelic analysis of the *vacA* and *cagA* genes in the bacterial population revealed that recombination is rare between bacteria from different continents, and that particular alleles are selected in certain populations [22, 43, 15]. Genomic comparisons among clonal strains are needed for a more comprehensive understanding of the genomics of *H. pylori*.

Since the genomes of two isolates of *H. pylori*, strain 26695 and strain J99, were sequenced, focus on comparative analysis of the genome sequences, proteomic analysis, and DNA microarray technology have been providing insights into the bacterial virulence, functional genomics, and proteomics [24, 35]. Proteomics, in particular, enables us to find ways to develop of promising vaccines [14].

Here, we constructed an ordered BAC library and a genomic map of *H. pylori* strain 51, which was isolated from a Korean patient with a duodenal ulcer. We then compared specific genetic loci and overall genomic organization between strain 51, strain 26695, and strain J99.

## MATERIALS AND METHODS

### Bacterial Strain, Plasmids, and Chemicals

*H. pylori* strain 51 was recovered from a patient with a duodenal ulcer at Gyeongsang National University Hospital, and was identified in a previous report [33]. *E. coli* DH10B/r and XL1-blue MR were used as hosts for BAC and pTZ19U vector, respectively. pBeloBAC11 [37] was kindly provided by M. I. Simon (Division of Biology, California Institute of Technology, Pasadena, CA, U.S.A.). pTZ19U and proteinase K were purchased from United States Biochemical and Boehringer Mannheim, respectively. Chemicals were from Sigma-Aldrich, and United States Biochemical. The restriction enzymes and gelase were from Takara, Roche Molecular Biochemicals, and Epicentre Technologies.

### Library Construction

For construction of the BAC library, high-molecular-weight genomic DNA was obtained from *H. pylori* by the preparation of cell-embedded low melting point (LMP) agarose blocks, as described previously with slight modification [28, 40]. Agarose-embedded genomic DNA of *H. pylori* strain 51 was partially digested with *Hin*dIII (Roche Molecular Biochemicals). Pulsed-field gel electrophoresis (PFGE) of digested DNA was performed in a 1% LMP-agarose gel using a 2015 Pulsaphor (Pharmacia LKB) with a pulse of 35 s for 20 h at 10°C and 7.5 V/cm. A zone of 50–200 kb was excised from the gel and digested with gelase at 45°C for 1 h. Ligation with pBeloBAC11 and transformation into *E. coli* DH10B were performed as described previously [37]. Plasmid library of *H. pylori* strain 51 was prepared into the pTZ19U plasmid vector as described previously [12, 18].

### Preparation of BAC DNA from Recombinant

BAC DNA was prepared by the following alkaline lysis method [36]. One to two hundred ng of BAC DNA were digested with *Sal*I or *Not*I, and separated by pulsed-field gel electrophoresis (PFGE) using a Pulsaphor with a pulse of 20 s for 10 h at 10°C and 5 V/cm. The insert size of each clone was determined by direct comparison with a lambda concatamer PFGE marker.

### Preparation of Hybridization Filters for BAC and Plasmid DNA

BAC DNAs were denatured by mixing with the same volume of 0.4 N NaOH solution and were blotted on a Hybond-$N^+$ nylon filter (Amersham Pharmacia Biotech Korea) with a dot-blot array of 175 clones per filter. The filters were washed with 2× SSC for 30 min and stored in 2× SSC/0.1% SDS solution at 4°C until used.

For hybridization of plasmid DNAs, *E. coli* transformants were blotted in a highly packed array on nylon filters (384 clones per filter) using a 384-Pin Replicator (Nalge Nunc International). The filters were processed as described above [36]. DNA fragments from known genes or gene segments were generated by polymerase chain reaction (PCR) amplification from *H. pylori* strain 26695 DNA. Fifteen sets of primers (Table 1) were designed using sequence data of *H. pylori* genes from a public database (GenBank).

### Physical Mapping

Agarose-embedded genomic DNA of *H. pylori* strain 51 was digested with *Apa*I, *Not*I, or *Sfi*I, and the restriction fragments were separated by PFGE using a 1% agarose gel with a pulse of 35 s at 10°C and 7.5 V/cm for an appropriate period, depending on the fragment size. The DNA was transferred to nylon filters by Southern blotting. Overlapping restriction fragments were isolated from PFGE gels using a QIAEX II Gel Extraction Kit (Qiagen), labeled by random-hexamer priming [17] with [$\alpha$-$^{32}$P]dATP, and hybridized with the blots. The hybridization reactions were performed

**Table 1.** Oligonucleotide primers used for PCR amplification of 15 genes.

| Locus | Genes | Orientation | Sequences* |
|---|---|---|---|
| *cagA* | Cytotoxin-associated gene | Forward | 5'-TCATGCGAGCGGCGATGTG-3' |
| | | Reverse | 5'-GTGCCTGCTAGTTTGTCAGCG-3' |
| *flaA* | Major flagellin | Forward | 5'-AATCGGTCAGGTTCGTATCG-3' |
| | | Reverse | 5'-AAGCACTAGGCCCATTACTG-3' |
| *flaB* | Minor flagellin | Forward | 5'-AAGACATTCTGTTGCACCGC-3' |
| | | Reverse | 5'-AACACTTTAGGCGTTAGGGC-3' |
| *frdC* | Fumarate reductase cytochrome b subunit | Forward | 5'-AGCATACAAGCTCCCAATGG-3' |
| | | Reverse | 5'-TCCATGACGCTCTTTATCGC-3' |
| *fucT* | Fucosyltransferase | Forward | 5'-CTGAATTCATGTTCCAACCCC-3' |
| | | Reverse | 5'-GCGAATTCAAATCTTTCGCCACGC-3' |
| *ggt* | γ-Glutamyltranspeptidase | Forward | 5'-GAAAACGATTGGCTTGGG-3' |
| | | Reverse | 5'-TCTTTCCTTGGATCCG-3' |
| *hpaA* | Flagellar sheath adhesin | Forward | 5'-TTCAAAGCGCTCTTGATCGC-3' |
| | | Reverse | 5'-ATTACCATCCAGCTAGCGAG-3' |
| *hpn* | Histidine-rich metal-binding polypeptide | Forward | 5'-AGCCACACCGATATTACTCG-3' |
| | | Reverse | 5'-TATTCAACCCGCATGAAGGC-3' |
| *metG* | Aminoacyl tRNA synthetase | Forward | 5'-TGGCCATGCTTATACGAC-3' |
| | | Reverse | 5'-TACCCCATTCAAAGCTCG-3' |
| *sodB* | Superoxide dismutase | Forward | 5'-ATCCACTGATCCTAAGCC-3' |
| | | Reverse | 5'-TGCTAAAGACAGCATGGG-3' |
| *tsaA* | Alkyl hydroperoxide reductase | Forward | 5'-CCACGCCCAATAACGATG-3' |
| | | Reverse | 5'-TCCAAAGAGCCGTCATTG-3' |
| *ureB* | Urease structural protein (subunit B) | Forward | 5'-TGGGACTGATGGCGTGAGGG-3' |
| | | Reverse | 5'-ATCATGACATCAGCGAAGTTAAAAATGG-3' |
| *vacA* | Vacuolating cytotoxin | Forward | 5'-ATGGAAATACAACAAACACAC-3' |
| | | Reverse | 5'-CTGCTTGAATGCGCCAAAC-3' |
| 16S rRNA | 16S ribosomal RNA | Forward | 5'-ACGTATTCACCGCAACATGG-3' |
| | | Reverse | 5'-GAGGTAGGTGGAATTCTTGG-3' |
| 23S rRNA | 23S ribosomal RNA | Forward | 5'-GAACAAGTCAGATGCTGCAG-3' |
| | | Reverse | 5'-TAGAGGCTTTCTTGGCACG-3' |

*Nucleotide sequences were obtained from the whole genome sequence of *H. pylori* strain 26695 [42].

at 48°C for 8 h. The blots were washed 3 times for 30 min each with 2× SSC/0.1% SDS solution and 1× SSC/0.1% SDS solution. The filters were wrapped in plastic wrap and autoradiographed using Kodak X-OMAT AR film (Eastman Kodak Company) with a double intensifying screen, at -70°C for 24–48 h.

**BAC Contig Mapping**

Overlapping BAC clones were identified by chromosome walking by dot-blot hybridizations with BAC array filters. Restriction fragments of genomic DNA and insert DNAs of stepwise-selected BAC clones were used as probes. In order to secure a sufficient amount of DNA for chromosome walking and confirmation of by, BAC DNAs within a contig were prepared using a plasmid large-scale preparation procedure. Large-scale preparations of BAC DNA were digested with *Sal*I, and the inserts were separated from the vectors by PFGE and extracted from the gel with a QIAEX II Gel Extraction Kit (Qiagen). For hybridization, insert DNAs were radiochemically labeled as described above. The contig map was assembled by assigning positive BAC clones identified by dot-blot hybridization.

## RESULTS AND DISCUSSION

**Construction of Physical Map**

Most of the enzymes cut *H. pylori* DNA either into many fragments that were too small and too numerous for genome sizing and mapping or did not cut the genomic DNA at all. However, three enzymes, *Apa*I, *Not*I, and *Sfi*I, cleaved the genome into a convenient number of fragments with suitable sizes for PFGE analysis (Figs. 1 and 2). *Apa*I yielded seven fragments; 316±5, 308±1, 288±2, 234±3, 199±2, 177±4, and 163±7 kb in size. *Not*I also yielded seven fragments; 516±7, 356±8, 288±5, 242±5, 210±3, 82±4, and 5 kb in size. *Sfi*I yielded only two fragments; one bigger than 1,000 kb and one 328±3 kb in size. The genome size of *H. pylori* 51 was estimated to be 1,698± 24 kb by summing the size of the *Not*I fragments.

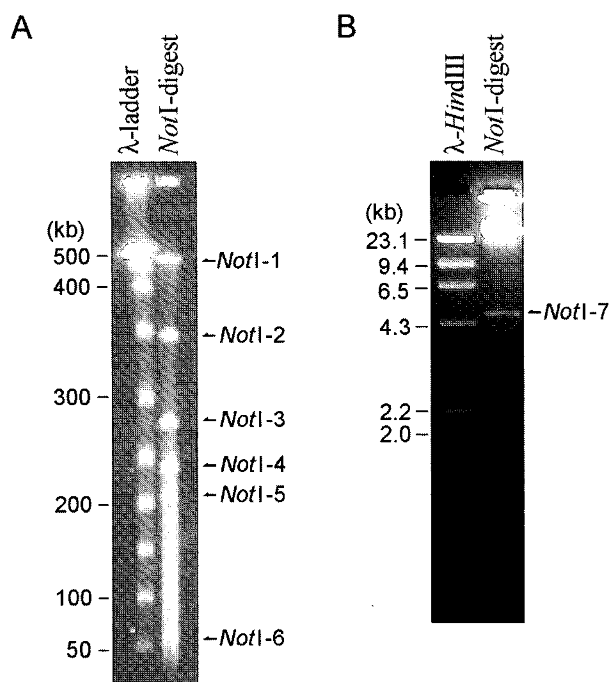


**Fig. 1.** PFGE separation (A) and standard agarose gel electrophoresis (B) of *H. pylori* strain 51 genomic DNA digested with *Not*I.

For PFGE, agarose-embedded genomic DNA was digested with *Not*I and then separated in a 1% agarose gel with a pulse of 35 s at 10°C, 7.5 V/cm, for 20 h. The gels were stained with ethidium bromide. A λ phage DNA concatamer (5-1,000 kb) and *Hin*dIII-digested λ phage DNA were used as molecular markers for PFGE and standard agarose gel electrophoresis, respectively.

A physical map of *H. pylori* strain 51 was constructed by a cross-hybridization strategy described previously [23]. Fragments of *Apa*I-digested genomic DNA were used as probes to hybridize to Southern blots obtained from PFGE of genomic DNA digested with either *Not*I or *Sfi*I (Fig. 2). Similarly, the *Not*I-digested genomic DNA fragments were hybridized to Southern blots obtained from PFGE of genomic DNA digested with either *Apa*I or *Sfi*I (Fig. 2). As shown in Fig. 3, the map was then constructed by analyzing these hybridization data. Information obtained from hybridizing *Apa*I-*Not*I, *Apa*I-*Sfi*I, or *Not*I-*Sfi*I double-digest fragments (Fig. 2) helped to establish the physical map.

## Ordering BAC Clones by Chromosome Walking

To determine the structure of the strain 51 genome, we constructed a genomic library of *H. pylori* strain 51 with BAC vector, because of its advantage of not only high stability of the clones containing DNA inserts of bigger than 100 kb [37], but also constructing the physical and genetic map of the *H. pylori* genome. BAC library was a good choice to overcome the difficulty of the defined physical mapping of the *H. pylori* genome, due to DNA modification systems and genetic variability.

The constructed BAC library produced approximately $6\times10^3$ transformants. PFGE profiles of randomly selected *Sal*I-cleaved BACs showed that the insert sizes of most of the clones ranged between 50 and 200 kb with an average size of 80 kb (data not shown).

Given a genome size of about 1,700 kb, 700 random members of the BAC library should nearly be 32-fold

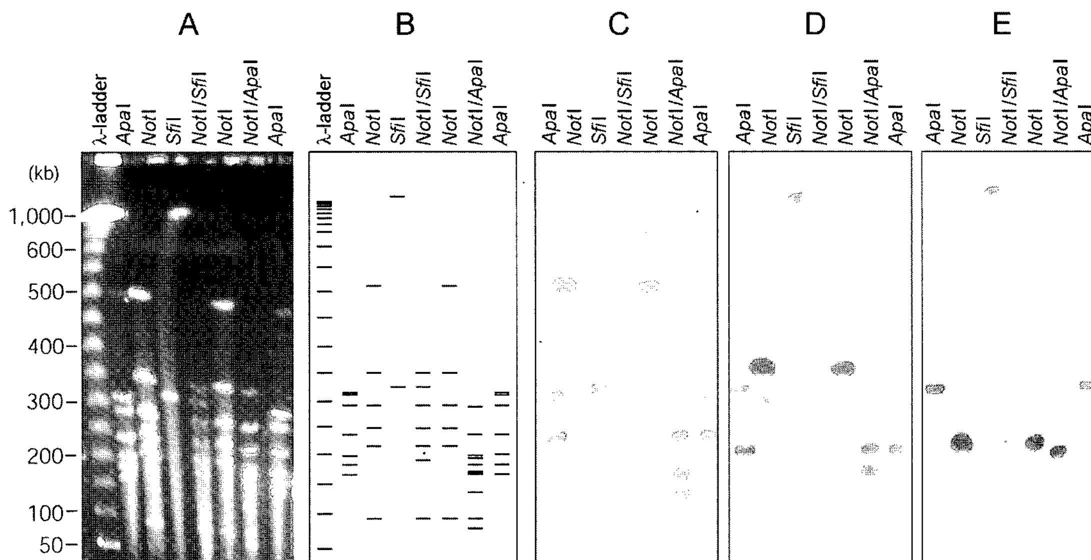


**Fig. 2.** PFGE separation and Southern hybridization analysis of *H. pylori* strain 51 genomic DNAs digested with restriction endonucleases, *Apa*I, *Not*I, *Sfi*I, *Not*I/*Sfi*I, and *Not*I/*Apa*I.

A, PFGE separation of *H. pylori* strain 51 genomic DNAs digested with *Apa*I, *Not*I, *Sfi*I, *Not*I/*Sfi*I, and *Not*I/*Apa*I; B, diagrams of PFGE separation of *H. pylori* strain 51 genomic DNAs digested with restriction endonucleases; C, D, and E, representative Southern hybridization analyses of the gel shown in panel A, using $^{32}$P-labeled genomic *Not*I fragments 1, 2, and 5, respectively.

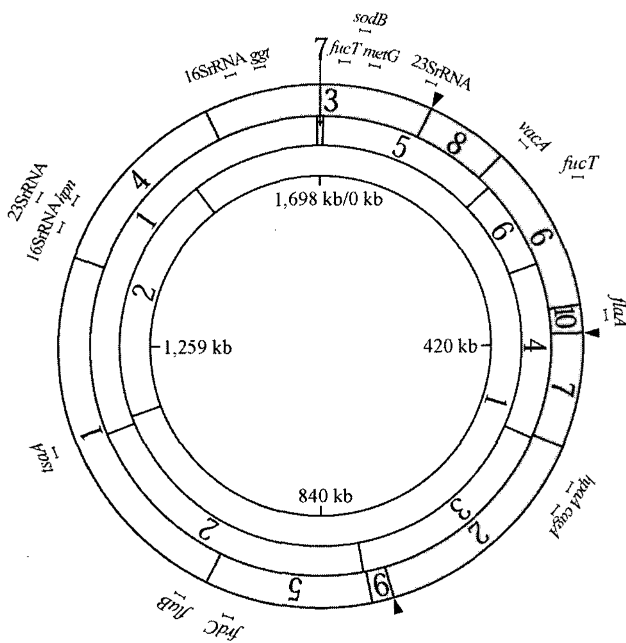


**Fig. 3.** Physical and genetic map of *H. pylori* strain 51 generated with *Apa*I, *Not*I, and *Sfi*I.
The outer circle represents *Apa*I fragments, the second circle *Not*I fragments, and the third circle *Sfi*I fragments. The arc lines outside the circle indicate the locations of the genes. Restriction fragments were numbered in order of their sizes. The orders of the restriction fragments in the genome are 2, 8, 7, 10, 6, 3, 9, 5, 1, and, 4 for the *Apa*I fragments and 5, 6, 4, 3, 2, 1, and 7 for the *Not*I fragments. Dark arrows represent additional *Apa*I sites confirmed by the restriction analysis of BAC clones.

redundant. The BAC clones were ordered with the chromosome walking method, in which the 700 members of the BAC library were hybridized with the insert DNAs of BAC clones to identify linked clones. For the initial probes for chromosome walking, we randomly selected 6 clones from the BAC library that had reacted exclusively with 6 *Not*I-digested genomic DNA fragments of more than 80 kb. The BAC DNAs were digested with *Sal*I and separated in an agarose gel to prepare the insert DNA probes. After the insert DNA probes of the 6 selected BAC clones were hybridized to the 700 BAC clones, a second round of hybridizations was performed with the insert DNAs of the BAC clones to which the initial 6 probes had bound. All subsequent hybridizations were carried out using the probes prepared with insert DNAs of BAC clones that had reacted to previous probes, until a total of 80 kd of probes were obtained. The order and orientation of the BACs were then confirmed, and the extent of the overlapping regions of neighbor clones was estimated by restriction analysis. The DNA of each BAC clone was fingerprinted by digestion with the restriction enzyme *Bfr*I to estimate the lengths of segments shared by overlapping neighbor BACs [12]. Finally, the closely overlapping set of the 82 selected BACs was assembled to cover all regions of the strain 51 genome, as shown in Fig. 4. In addition, the insert DNAs of the 82 BAC clones were digested with *Not*I, separated in an agarose gel by PFGE, and subjected to Southern hybridization using chromosomal *Not*I fragments

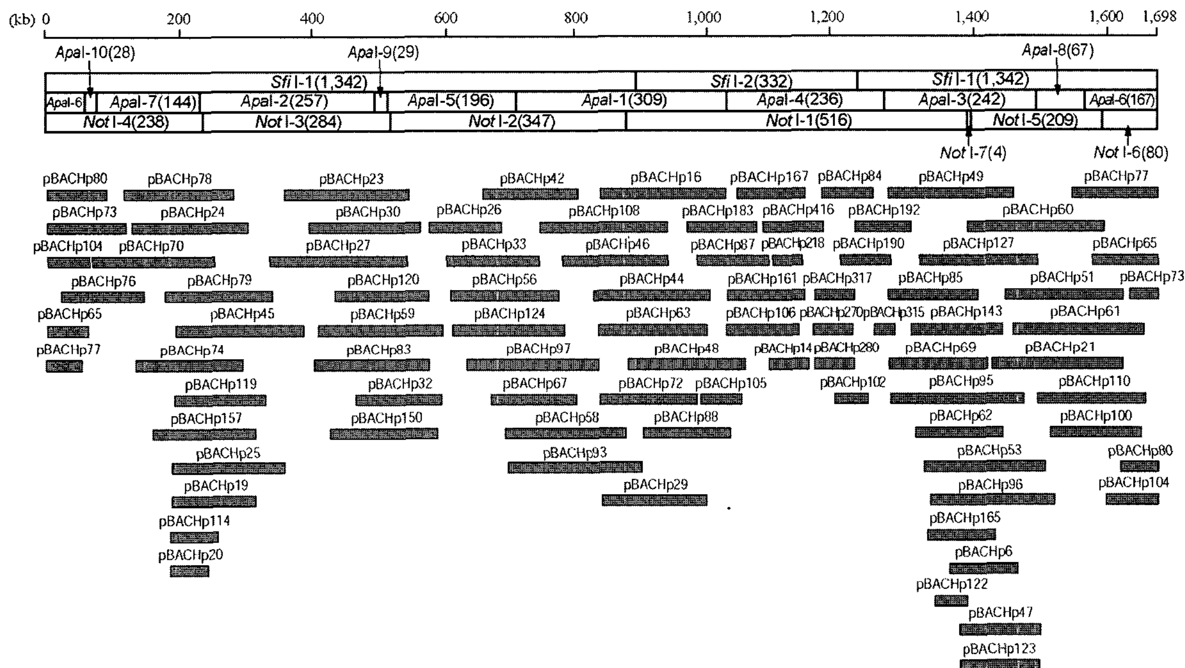


**Fig. 4.** Linear restriction maps of *H. pylori* strain 51 and map of an ordered overlapping maxi-set of 82 BAC clones that covered the entire genomic region of *H. pylori* strain 51.
Parentheses indicate the sizes (kb) of restriction DNA fragments.

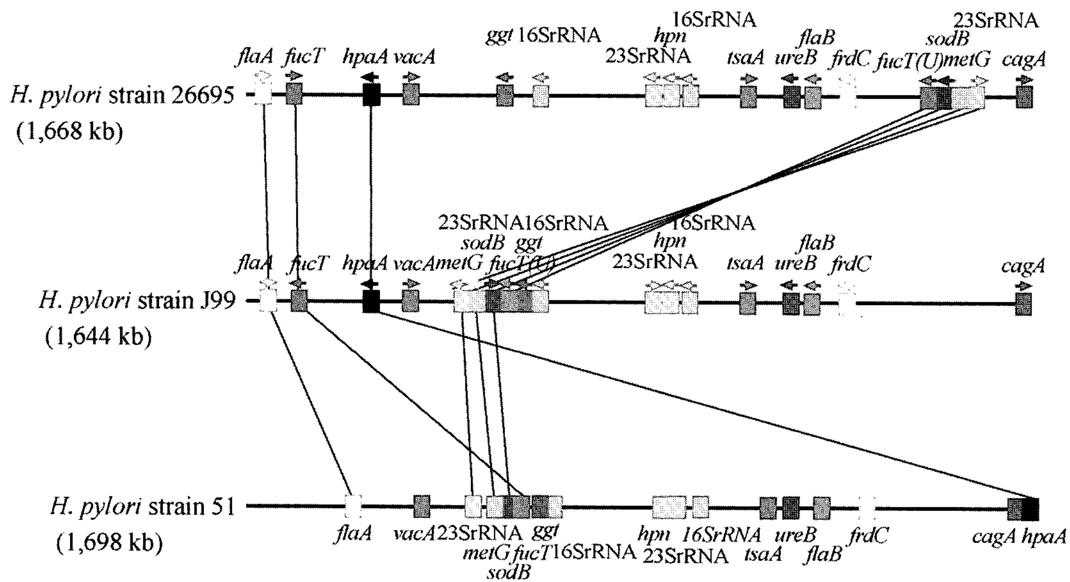


**Fig. 5.** Comparison of the genetic maps among *H. pylori* strain 26695, strain J99, and strain 51.
The genetic maps included *flaA, fucT, hpaA, vacA, metG, sodB, ggt, hpn, tsaA, ureB, flaB, frdC, cagA*, 23S rRNA genes, and 16S rRNA genes. The genetic maps of strain 26695 and strain J99 are based on the whole genome sequences, and those of strain 51 were constructed by Southern hybridization using an ordered overlapping maxi-set of 82 BAC clones. Arrow markers represent the orientation of the ORFs.

as probes in order to verify whether the physical map, based on the chromosome walking of BAC clones, was consistent with the map constructed with *Not*I fragments of the strain 51 genome (Fig. 4). A similar analysis was carried out with *Apa*I digests. The map based on the ordered overlapping set of 82 BAC clones matched well with the physical map constructed by restriction analysis, except for 3 additional *Apa*I sites identified in the BAC set.

**Construction of Genetic Maps of *H. pylori***

Fifteen known *H. pylori* genes (*cagA, flaA, flaB, frdC, fucT, ggt, hpaA, hpn, metG, sodB, tsaA, ureB, vacA*, and genes for 16S and 23S ribosomal RNAs) were mapped to the ordered overlapping set of 82 BAC clones by dot-blot hybridization and by Southern hybridization to PFGE-separated *Apa*I or *Not*I fragments of the strain 51 genome (Fig. 5). The gene probes used in this study were generated by PCR with the primers listed in Table 1. The overlapping set of 82 BAC clones made it possible to localize the genes with resolution of 4–22 kb. All of the genes except for the rRNA genes were mapped as single-copy genes. Both the 16S rRNA and 23S rRNA genes were mapped to BAC clones, corresponding to two chromosomal loci.

The genetic map of *H. pylori* strain 51 was compared with those of strain 26695 [42] and strain J99 [3] (Fig. 5). The urease subunit B gene was used as the reference point for comparing the three genetic maps. In all three *H. pylori* strains, most genes were present in a single copy, and the 16S and 23S rRNA genes were present in two copies. The exception was the fucosyltransferase gene (*fucT*); strain 51 has only one copy, whereas strain 26695 and strain J99 each has two copies. The chromosomal loci of eight genes (*cagA, flaB, frdC, ggt, hpn, tsaA, ureB*, and *vacA*) and one copy each of the 16S and 23S rRNA genes were generally conserved among the three strains. In contrast, there were differences in the loci of *flaA* and *hpaA* between strain 51 and strain J99 and in the loci of the *flaA, fucT, hpaA, metG, sodB* genes, and one copy of the 23S rRNA gene between strain 51 and strain 26695 (Fig. 5).

**Analysis of the Plasticity Zone and Strain-Specific Sequences**

A comparison of the genomic sequences of *H. pylori* in strains 26695 and J99 identified a 55-kb region, named plasticity zone, which contains many strain-specific genes and a low percentage of G+C nucleotides. It needs to be clarified whether the strain 51 genome also contains a plasticity zone containing a large percentage of strain-specific genes. Conserved ORFs (JHP913 and JHP962) located in the 5'- and 3'-boundary regions of the plasticity zone in the strain J99 genome were used to screen the maxi-set of 82 ordered overlapping clones to identify a BAC clone corresponding to the plasticity zone of the strain 51 genome. pBACHp123 was selected for analyzing the plasticity zone of strain 51, because it had the smallest insert among the clones that reacted with both ORF probes (Fig. 6). The insert DNA fragment of pBACHp123, whose size was estimated to be 132 kb by PFGE, contains ORFs corresponding to JHP880 and JHP1003 of strain J99 in its

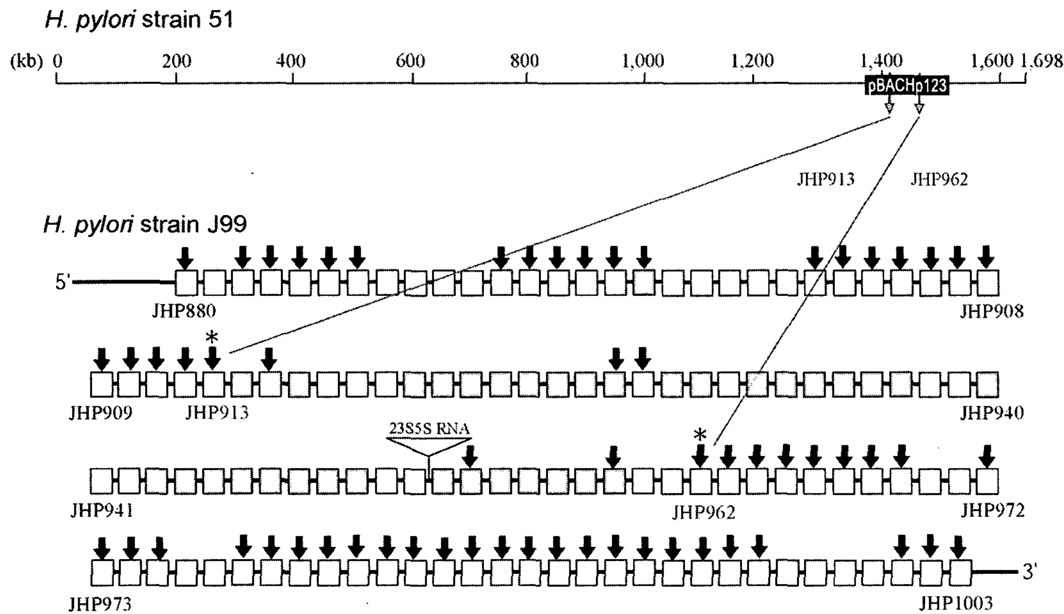


**Fig. 6.** A schematic diagram of the ORF organization (from JHP880 to JHP1003) in the genomic regions of *H. pylori* strain J99 corresponding to the 132-kb insert DNA of pBACHp123 that contained the plasticity zone.

The plasticity zone is from JHP914 to JHP961 (ORFs marked by an asterisk are just outside the plasticity zone). Open rectangles, conserved ORFs; shaded rectangles, strain J99-specific ORFs; arrow markers above rectangles, ORFs containing nucleotide sequences showing homology to the read sequences of plasmid clones that were selected from the plasmid library of *H. pylori* strain 51 genome by hybridization analysis using the 132 kb-insert DNA of pBACHp123.

5'- and 3'-terminus, respectively (Fig. 6). In strain J99, the plasticity zone is a 53.4-kb region from JHP914 to JHP961, located in the larger genomic region from JHP880 to JHP1003, which is 137.3 kb and contains 123 ORFs [3]. To partially compare the structure of the plasticity zone between strain J99 and strain 51, 266 clones were randomly selected from a plasmid library of the strain 51 genome by colony hybridization using the insert DNA of pBACHp123

**Table 2.** Analysis of the nucleotide sequences of 266 plasmid clones that were randomly selected from the plasmid library of *H. pylori* strain 51 by dot-blot hybridization with pBACHp123.

<table>
<tr><th colspan="4">Number, location, and specificity of ORFs to which the sequence reads showed homology</th><th>G+C%</th><th>Remark</th></tr>
<tr><td>No. of clones containing sequence reads matching ORFs scattered outside of the JHP0880-1003 region of strain J99</td><td>24</td><td></td><td></td><td rowspan="2">39.5</td><td>Pertinent to 20 ORFs outside of the JHP0880-1003 region*</td></tr>
<tr><td rowspan="3">No. of clones containing sequence reads matching ORFs within the JHP0880-1003 region of strain J99 or in the corresponding region of strain 26695</td><td rowspan="3">236</td><td>No. of sequence reads matching ORFs surrounding the plasticity zone</td><td>231</td><td>Pertinent to 63 ORFs located within the JHP0880-1003 region</td></tr>
<tr><td>No. of sequence reads matching J99-specific ORFs within the plasticity zone</td><td>4</td><td rowspan="2">35.4</td><td>Pertinent to 4 ORFs located within the JHP0914-0961 region</td></tr>
<tr><td>No. of sequence reads matching 26695-specific ORFs within the plasticity zone</td><td>1</td><td>Pertinent to 1 ORF located within the plasticity zone of strain 26695**</td></tr>
<tr><td>No. of clones containing sequence reads that did not show homology to any ORFs of strain 26695 and J99</td><td>6</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td>266</td><td></td><td></td><td></td><td></td></tr>
</table>

*JHP0130, JHP0131, JHP0344, JHP0345, JHP0348, JHP0678, JHP0731, JHP0806, JHP0807, JHP0808, JHP1084, JHP1085, JHP1086, JHP1153, JHP1155, JHP1306, JHP1316, JHP1346, JHP1411, and JHP1412.

**HP0423.

as a probe; the clones were then subjected to sequencing of a single terminus of their inserts. Here, the clones containing the 5S-23S rRNA loci were excluded because of the multiple copies. An average of 0.5 kb could be read in the sequencing reaction, revealing that total size of the reading sequences are approximately one coverage of the insert DNA of pBACHp123. The sequences were analyzed with the BLAST network service of the NCBI and are summarized in Table 2 and Fig. 6. Of 266 sequence reads, 24 contained sequences matching the ORFs scattered outside of the JHP880-1003 region of strain J99, and 236 contained sequences matching ORFs within the JHP880-1003 region of strain J99 or in the corresponding region of strain 26695. Six reads did not show homology to any ORFs of the strain 26695 or strain J99 genomes. Of the 236 reads showing sequences homologous to ORFs within the JHP880-1003 region, 231 contained sequences of 63 ORFs located on the outside of the plasticity zone (JHP914-961) within JHP880-1003 regions of strain J99 and in the corresponding regions of strain 26695. Four reads matched 4 strain-specific ORFs in the plasticity zone of strain J99, and 1 read contained sequences matching a strain-specific ORFs in the plasticity zone of strain 26695. The 24 reads and 231 reads located outside of the plasticity zone contained 39.5% G+C. The remaining reads, which were presumably located in the plasticity zone, were 35.4% G+C. These results demonstrate that strain 51 carries a plasticity zone that is shorter than those of strain 26695 and strain J99, and consists of ORFs common only to strain 51 and strain J99 or to strain 51 and strain 26695, in addition to its own specific ORFs.

**Analysis of Genomic Organization Between Strain 51 and Strain J99**

Of the 82 BAC clones, the insert DNAs of 13 clones covering nearly all regions of the strain 51 genome were isolated and used to select clones from the plasmid library of the strain 51 genome. Each probe DNA was hybridized with 1,152 clones that were randomly selected from the plasmid library of the strain 51 genome, resulting in screening of a total of 14,976 clones for the insert DNAs of 13 clones. Clones showing definite hybridization to the probes were randomly selected, and the sequences of both ends of the inserts were determined. We then searched for probable orthologue ORFs of strain J99 that were homologous to these insert DNAs. A total of 1,627 clones, 93 to 266 for each of the probes, contained sequences homologous to orthologue ORFs of strain J99. The locations of these ORFs were marked in the orthologue list of the strain J99 genome, and their arrangement within the genome were compared with those of strain 51, as shown in Fig. 7. The insert DNA of pBACHp190 was used as the reference region for comparing the genomic arrangement between the two strains, because its positive clones were clustered in a single region of strain J99. The clones that were positive for the 131-kb insert DNA of pBACHp104 were clustered separately in 2 regions of the strain J99 genome; this location was estimated to correspond to the pBACHp45 regions of the strain 51 genome. The clones positive for the 196-kb insert DNA of pBACHp45 were clustered separately in 2 regions of the strain J99 genome; of these, one is translocated to the region corresponding to the location of pBACHp104 in the strain 51 genome (200-kb upstream

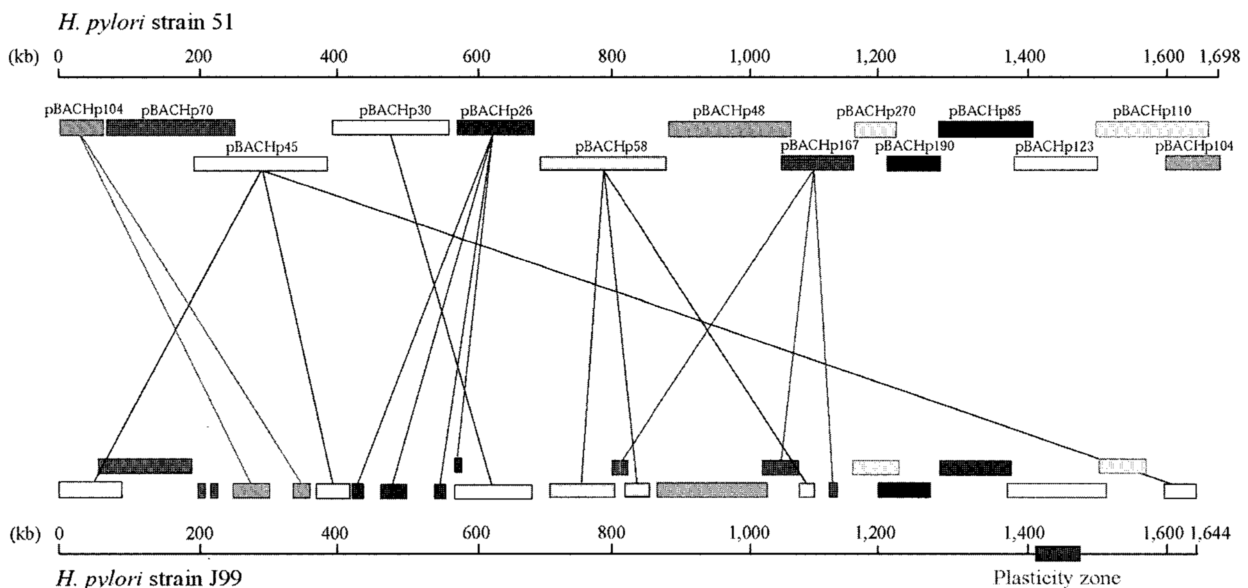


**Fig. 7.** The overall genomic organization between *H. pylori* strain 51 and strain J99.
The locations of homologous orthologue ORFs were compared between strains 51 and J99. Plasmid clones showing definite positivity to each of 13 insert DNA probes were randomly selected; both ends were sequenced and the sequences were used to search for homology to the orthologue ORFs of strain J99.

region of pBACHp45). The clones positive for the 115-kb insert DNA of pBACHp26 were clustered separately in 4 regions of the strain J99 genome, corresponding to the pBACHp30 region of the strain 51 genome (150-kb upstream of pBACHp26). The clones positive for the 163-kb DNA of pBACHp30 were located in the region corresponding to pBACHp26 of the strain 51 genome (150-kb downstream of pBACHp30). Thus, the order of the genomic regions corresponding to pBACHp30 and pBACHp26 were reversed between strain 51 and strain J99. The clones positive for the 171-kb insert DNA of pBACHp58 were divided into 3 clusters in the strain J99 genome. Of these, 2 clusters were located in the strain J99 genomic regions identical to the pBACHp58 region of strain 51 and separated by insertion of parts of clones positive for the insert DNA of the pBACHp167. The third cluster was located in the strain J99 genomic regions corresponding to the pBACHp167 region of strain 51. The clones positive for the 66-kb insert DNA of pBACHp270, the 132-kb insert DNA of pBACHp85, and the 133-kb insert DNA of pBACHp123 were located in the identical regions in strain J99 and strain 51. Taken together, these results reveal that there are 3 genetic translocations and/or inversions between strain J99 and strain 51.

*H. pylori* has notoriously high chromosomal variations among isolates at the levels of micro- and macrodiversity, and carries a variety of DNA restriction and modification systems. This study demonstrates that a genomic pBAC library covering the entire genome of *H. pylori* strain 51 is a powerful resource for constructing a defined and accurate genomic map, despite the difficulties presented by the large amount of diversity in the genome. Furthermore, BAC clones might easily and stably maintain the whole *H. pylori* genome without allowing genomic variation, thereby creating a reference genetic resource for the investigation of postgenomic applications.

Since *H. pylori* genomes were sequenced, research fields including functional genomics and proteomics have been launched with the genome sequences. Proteomics, in particular, will provide insights into the pharmaceutical research for the development of drug targets, including vaccines to cure the gastric pathogens. Furthermore, the DNA microarray technology will also be greatly helpful to not only search for the genes or gene products for the drug target, but also find the virulence factors of *H. pylori*.

## Acknowledgments

This work was partially supported by a grant from the 21C Frontier Program (Functional Analysis of Human Genome) and the Korea Science and Engineer Foundation (R01-2000-000-00075-0). Jae-Young Song, Sang-Haeng Choi, Seong-Gyu Park, Eun-Joo Lee, Ji-Sun Kim, and Jeong-Uck Park were supported by the Brain Korea 21 program from the Ministry of Education, Korea. Jeoung-Uck Park was supported by a Korea Research Foundation Grant (KRF-2002-050-E00002). Hyung-Lyun Kang and Woo-Kon Lee contributed equally to this study.

## REFERENCES

1. Achtman, M., T, Azuma, D. E. Berg, Y. Ito, G. Morelli, Z. J. Pan, S. Suerbaum, S. A. Thompson, A. van der Ende, and L. J. van Doorn. 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol. Microbiol.* **32:** 459–470.
2. Akopyanz, N., N. O. Bukanov, T. U. Westblom, S. Kresovich, and D. E. Berg. 1992. DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. *Nucleic Acids Res.* **20:** 5137–5142.
3. Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori. Nature* **397:** 176–180.
4. Alm, R. A. and T. J. Trust. 1999. Analysis of the genetic diversity of *Helicobacter pylori*: The tale of two genomes. *J. Mol. Med.* **77:** 834–846.
5. Anonymous. 1994. Schistosomes, liver flukes and *Helicobacter pylori. IARC Monogr. Eval. Carcinog. Risks Hum.* **61:** 1–241.
6. Baik, S.-C., J.-B. Kim, M.-J. Cho, Y.-C. Kim, C.-G. Park, H.-H. Rhou, H.-J. Choi, and K.-H. Rhee. 1990. Prevalence of *Helicobacter pylori* infection among normal Korean adults. *J. Korean Soc. Microbiol.* **25:** 455–462.
7. Baik, S.-C., H.-S. Youn, M.-H. Chung, W.-K. Lee, M.-J. Cho, G.-H. Ko, C.-G. Park, H. Kasai, and K.-H. Rhee. 1996. Increased oxidative DNA damage in *Helicobacter pylori*-infected human gastric mucosa. *Cancer Res.* **56:** 1279–1282.
8. Blaser, M. J. 1987. Gastric Campylobacter-like organisms, gastritis and peptic ulcer disease. *Gastroenterology* **93:** 371–383.
9. Blaser, M. J. 1990. *Helicobacter pylori* and the pathogenesis of gastroduodenal inflammation. *J. Infect. Dis.* **161:** 626–633.
10. Blaser, M. J. 1992. Hypotheses on the pathogenesis and natural history of *Helicobacter pylori*-induced inflammation. *Gastroenterology* **102:** 720–727.
11. Blaser, M. J. 1998. *Helicobacter pylori* and gastric diseases. *Br. Med. J.* **316:** 1507–1510.
12. Bukanov, N. O. and D. E. Berg. 1994. Ordered cosmid library and high-resolution physical-genetic map of *Helicobacter pylori* strain NCTC11638. *Mol. Microbiol.* **11:** 509–523.
13. Cave, D. R. 1997. Section I: Epidemiology and transmission of *Helicobacter pylori* infection - how is *Helicobacter pylori* transmitted? *Gastroenterology* **113:** S9–14.

14. Chakravarti, D. N., M. J. Fiske, L. D. Fletcher, and R. J. Zagursky. 2000. Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. *Vaccine* **19:** 601–612.
15. Choi, K. M., J. K. Park, and S. Y. Hwang. 2003. Pathogenetic impact of vacuolar degeneration by accelerated transport of *Helicobacter pylori* VacA. *J. Microbiol. Biotechnol.* **13:** 666–672
16. Falush, D., C. Kraft, N. S. Taylor, P. Correa, J. G. Fox, M. Achtman, and S. Suerbaum. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size, and minimal age. *Proc. Natl. Acad. Sci. USA* **98:** 15056–15061.
17. Feinberg, A. and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.
18. Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. Mckenny, G. Sutton, W. Fitzhugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L. Liu, A. Glodek, J. M. Kelly, J. F. Weidman, C. A. Philips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, C. M. Fraser, H. O. Smith, and J. C. Venter. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.
19. Forman, D., D. G. Newell, F. Fullerton, J. W. Yarnell, A. R. Stacey, N. Wald, and F. Sitas. 1991. Association between infection with *Helicobacter pylori* and risk of gastric cancer: Evidence from a prospective investigation. *Br. Med. J.* **302:** 1302–1305.
20. Hook-Nikanne, J., D. E. Berg, R. M. Peek Jr., D. Kersulyte, M. K. Tummuru, and M. J. Blaser. 1998. DNA sequence conservation and diversity in transposable element IS605 of *Helicobacter pylori. Helicobacter* **3:** 79–85.
21. Hua, J., H. C. Ng, K. G. Yeoh, and B. Ho. 1999. Predominance of a single strain of *Helicobacter pylori* in gastric antrum. *Helicobacter* **4:** 28–32.
22. Ito, Y., T. Azuma, S. Ito, H. Suto, H. Miyaji, Y. Yamazaki, Y. Kohli, and M. Kuriyama. 1998. Full-length sequence analysis of the *vacA* gene from cytotoxic and noncytotoxic *Helicobacter pylori. J. Infect. Dis.* **178:** 1391–1398.
23. Jiang, Q., K. Hiratsuka, and D. E. Taylor. 1996. Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. *Mol. Microbiol.* **20:** 833–842.
24. Jungblut, P. R., D. Bumann, G. Haas, U. Zimny-Arndt, P. Holland, S. Lamer, F. Siejak, A. Aebischer, and T. F. Meyer. 2000. Comparative proteome analysis of *Helicobacter pylori. Mol. Microbiol.* **36:** 710–725.
25. Kersulyte, D., H. Chalkauskas, and D. E. Berg. 1999. Emergence of recombinant strains of *Helicobacter pylori* during human infection. *Mol. Microbiol.* **31:** 31–43.
26. Kim, J. M., J. E. Shin, M. J. Han, S. H. Park, and D. H. Kim. 2003. Inhibitory effect of Ginseng saponins and polysaccharides on infection and vacuolation of *Helicobacter pylori. J. Microbiol. Biotechnol.* **13:** 706–709.
27. Kim, S. M., K. H. Lee, N. K. Lee, C. J. Kim, C. H. Kim, and H. D. Paik. 2004. Antagonistic activity of polyfermenticin SCD against *Helicobacter pylori* KCTC 2948. *J. Microbiol. Biotechnol.* **14:** 148–152.
28. Majumder, R., S. Sengupta, G. Khetawat, R. K. Bhadra, S. Roychoudhury, and J. Das. 1996. Physical map of the genome of *Vibrio cholerae* 569B and localization of genetic markers. *J. Bacteriol.* **178:** 1105–1112.
29. Marshall, D. G., W. G. Dundon, S. M. Beesley, and C. J. Smyth. 1998. *Helicobacter pylori* - a conundrum of genetic diversity. *Microbiology* **144:** 2925–2939.
30. Nedenskov-Sorensen, P., G. Bukholm, and K. Bovre. 1990. Natural competence for genetic transformation in *Campylobacter pylori. J. Infect. Dis.* **161:** 365–366.
31. Parsonnet, J., S. Hansen, L. Rodriguez, A. B. Gelb, R. A. Warnke, E. Jellum, N. Orentreich, J. H. Vogelman, and G. D. Friedman. 1994. *Helicobacter pylori* infection and lymphoma. *N. Engl. J. Med.* **330:** 1132–1136.
32. Peterson, W. 1991. *Helicobacter pylori* and peptic ulcer disease. *N. Engl. J. Med.* **324:** 1043–1048.
33. Rhee, K.-H., M.-J. Cho, J.-B. Kim, S.-K. Choi, C.-K. Park, Y.-C. Kim, J.-H. Choi, and K. J. Choe. 1988. A prospective study on the *Campylobacter pylori* isolated from patients of gastroduodenal inflammatory conditions. *J. Korean Soc. Microbiol.* **23:** 9–16.
34. Rhee, K.-H., H.-S. Youn, S.-C. Baik, W.-K. Lee, M.-J. Cho, H.-J. Choi, K.-Y. Maeng, and K.-W. Ko. 1990. Prevalence of *Helicobacter pylori* infection in Korea. *J. Korean Soc. Microbiol.* **25:** 475–490.
35. Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow. 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl. Acad. Sci. USA* **97:** 14668–14673.
36. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor, NY : Cold Spring Harbor Laboratory Press.
37. Shizuya, H., B. W. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89:** 8794–8797.
38. Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman. 1998. Free recombination within *Helicobacter pylori. Proc. Natl. Acad. Sci. USA* **95:** 12619–12624.
39. Taylor, D. N. and M. J. Blaser. 1991. The epidemiology of *Helicobacter pylori* infection. *Epidemiol. Rev.* **13:** 42–59.
40. Taylor, D. E., M. Eaton, N. Chang, and S. M. Salama. 1992. Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. *J. Bacteriol.* **174:** 6800–6806.
41. Taylor, N. S., J. G. Fox, N. S. Akopyants, D. E. Berg, N. Thompson, B. Shames, L. Yan, E. Fontham, F. Janney, and F. M. Hunter. 1995. Long-term colonization with single

and multiple strains of *Helicobacter pylori* assessed by DNA fngerprinting. *J. Clin. Microbiol.* **33:** 918–923.

42. Tomb, J.-F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, E. K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. Utterback, I. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori. Nature* **388:** 539–547.
43. Van Der Ende, A., Z.-J. Pan, A. Bart, R. W. M. van Der Hulst, M. Feller, S.-D. Xiao, G. N. J. Tytgat, and J. Dankert. 1998. *cagA*-positive *Helicobacter pylori* populations in China and The Netherlands are distinct. *Infect. Immun.* **66:** 1822–1826.
44. Wang, G., M. Z. Humayun, and D. E. Taylor. 1999. Mutation as an origin of genetic variability in *Helicobacter pylori. Trends Microbiol.* **7:** 488–493.
45. Warren, J. R. and B. Marshall. 1983. Unified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* **i:** 1273–1275.
46. Watanabe, T., M. Tada, H. Nagai, S. Sasaki, and M. Nakao. 1998. *Helicobacter pylori* infection induces gastric cancer in Mongolian gerbils. *Gastroenterology* **115:** 642–648.
47. Woo, J. S., B. H. Ha, T. G. Kim, Y. Lim, and K. H. Kim. 2003. Inhibition of *Helicobacter pylori* adhesion by acidic polysaccharide isolated from *Artemisia capillaries. J. Microbiol. Biotechnol.* **13:** 853–858.