J. Korean Soc. Microbiol., Vol. 34, No. 6, 1999

Genomic Diversity of *Helicobacter pylori*

Woo-Kon Lee, Sang-Haeng Choi, Seong-Gyu Park, Yeo-Jeong Choi, Mi-Young Choe, Jeong-Won Park, Sun-Ae Jung, Eun-Young Byun, Jae-Young Song, Tae-Sung Jung, Byung-Sang Lee, Seung-Chul Baik, Myung-Je Cho, Hee-Shang Youn*, Gyung-Hyuck Ko, Yong-Sung Kim§, Jong-Hoon Park§, Dae-Sil Lee†, Hyang-Sook Yoo§, Sa-Youl Ghim¶, Kwang-Ho Rhee**

Department of Microbiology, Pediatrics and Pathology**, Gyeongsang National University College of Medicine, Chinju, Genome Research Center§ and Laboratory of Glycobiology†, KRIBB, Taejon, Department of Microbiology¶, College of Natural Sciences, Kyungpook National University, Taegu, Korea*

Helicobacter pylori is a causative agent of type B gastritis and plays a central role in the pathogenesis of gastroduodenal ulcer and gastric cancer. To elucidate the host-parasite relationship of the *H. pylori* infection on the basis of molecular biology, we tried to evaluate the genomic diversity of *H. pylori*. An ordered overlapping bacterial artificial chromosome (BAC) library of a Korean isolate, *H. pylori* 51 was constructed to set up a genomic map. A circular physical map was constructed by aligning *Apa*I, *Not*I and *Sfi*I-digested chromosomal DNA. When the physical map of *H. pylori* 51 was compared to that of unrelated strain, *H. pylori* 26695, completely different restriction patterns were shown. Fifteen known genes were mapped on the chromosome of *H. pylori* 51 and the genetic map was compared with those of strain 26695 and J99, of which the entire genomic sequences were reported. There were some variability in the gene location as well as gene order among three strains. For further analysis on the genomic diversity of *H. pylori*, when comparing the genomic structure of 150 *H. pylori* Korean isolates with one another, genomic macrodiversity of *H. pylori* was characterized by several features: whether or not susceptible to restriction digestion of the chromsome, variation in chromosomal restriction fingerprint and/or high frequency of gene rearrangement. We also examined the extent of allelic variation in nucleotide or deduced amino acid sequences at the individual gene level. *fucT*, *cagA* and *vacA* were confirmed to carry regions of high variation in nucleotide sequence among strains. The plasticity zone and strain-specific genes of *H. pylori* 51 were analyzed and compared with the former two genomic sequences. It should be noted that the *H. pylori* 51-specific sequences were dispersed on the chromosome, not congregated in the plasticity zone unlike J99- or 26695-specific genes, suggesting the high frequency of gene rearrangement in *H. pylori* genome. The genome of *H. pylori* 51 shows differences in the overall genomic organization, gene order, and even in the nucleotide sequences among the *H. pylori* strains, which are far greater than the differences reported on the genomic diversity of *H. pylori*.

Key Words: *Helicobacter pylori*, Gastritis, Genome analysis, Plasticity zone

Received for publication: 2000. 2. 12, Accepted for publication: 2000. 2. 12

Corresponding author: Kwang-Ho Rhee, Department of Microbiology, Gyeongsang National University College of Medicine, Chinju, Kyung-Nam 660-751, Korea, 0591-751-8746 (phone) / 0591-759-1588 (fax), E-mail address: khrhee@gshp.gsnu.ac.kr

INTRODUCTION

In Korea, *Helicobacter pylori* infection begins at infancy. The infection rate reaches up to 50% at 5～6 years of age and 80～90% at 8 years of age and then maintains the percentage throughout the whole ages of adults (4,19). *H. pylori* colonizes the gastric mucosa of humans and causes gastritis and peptic ulcer diseases. *H. pylori* infection also causes gastric adenocarcinoma and MALT lymphoma, which has been verified by epidemiological studies as well as by an experimental infection leading to gastric adenocarcinoma in Mongolian gerbils (5,15, 20,21,24).

B type chronic gastritis, peptic ulcer diseases and gastric cancer could be annihilated by the direct countermeasures targeting *H. pylori*. Unfortunately, however, strategies based on conventional principles and practice of microbiology for the control of *H. pylori*-induced diseases are of limited utility, because of the following four reasons (25). First, antimicrobial drugs are not effective for the eradication of *H. pylori* on the basis of asymptomatic infection. Second, diagnostic utilities convenient for the *H. pylori* eradication are not available at this moment. Third, it seems theoretically impossible to develop vaccine against *H. pylori*. Fourth, it does not seem to be easy to block the transmission of *H. pylori* infection at the level of public health.

These barriers make it practically impossible to decrease rapidly the prevalence rate of *H. pylori* infection. Nevertheless, the control of the *H. pylori*-induced gastroduodenal diseases is urgently required in a country with a high prevalence rate of *H. pylori* infection, like Korea. Molecular biology of *H. pylori* on the basis of genome analysis could provide a breakthrough for the above-mentioned obstacles and for the understanding of *H. pylori*-human relationship.

Recently, the complete genomic sequences of two unrelated isolates of *H. pylori*, strain 26695 (23) and J99 (2), were reported. The sequence comparison revealed structural and allelic differences between the two genomes. It is remarkable that 23% of the genes in both strains are *H. pylori*-specific, which show no sequence similarity with genes availabe in public databases and that about 6 to 7% of the genes are specific to each strain, with almost half of these genes being clustered in a single hypervariable region, namely plasticity zone (2). This region is postulated to play a significant role in the pathogenesis of *H. pylori* infection.

Here, we tried to evaluate the genomic diversity of *H. pylori*.

Construction of ordered overlapping BAC library. A bacterial artificial chromosome (BAC) library of *H. pylori* was constructed to set up a genomic map. Chromosomal DNA of *H. pylori* strain 51, a clinical isolate recovered from patients with peptic ulcer at Gyeongsang National University Hospital, was prepared in low-melting point agarose block, partially digested with *Hin*dIII and then ligated with pBeloBAC11 (22), which was kindly provided by Dr. Melvin Simon at the Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

Seven hundred BAC DNAs which contained inserts with a range of 50～200 kb were selected and separately dot-blotted on 4 gridded nylon filters for hybridization. The set of filters was used as a random BAC library of *H. pylori*. Ordering of BAC DNAs was done by chromosome walking method involving hybridization of probes made from the inserts of representative BAC clones to the set of filters. Eighty-four BACs were mapped to a maxiset contig covering the entire chromosome, and are represented by a miniset contig consisting of 16 ordered BACs (data not shown). The order of BACs was confirmed and extents of overlap among them were estimated by restriction analysis.

Analysis of physical map. Chromosomal DNA from *H. pylori* 51 was digested with 42 restriction endonucleases to identify restriction patterns suitable for mapping the genome. We identified three enzymes, *Apa*I, *Not*I, and *Sfi*I, which gave a small number of DNA fragments of higher molecular weight that were well resolved after pulsed-field gel electrophoresis (PFGE). The *H. pylori* chromosome contained 10 *Apa*I fragments ranging from 28 to 309 kb, 7 *Not*I fragments ranging from 4.5 to 516 kb, and 2 *Sfi*I fragments of 332 and 1,342 kb. The genome size of *H. pylori* 51 estimated by summing sizes of 7 *Not*I fragments was 1,679 kb. A circular physical map of the *H. pylori* was constructed using a cross-hybridization strategy (17). Briefly, fragments of *Not*I-digested chromosomal DNA were used as probes to hybridize to Southern blots obtained from PFGE of *Apa*I- and *Sfi*I-digested chromosomal DNA. Similarly, the *Apa*I-digested DNA fragments were hybridized to Southern blots from PFGE of *Not*I- and *Sfi*I-digested chromosomal DNA. The map was then constructed by analysing these hybridization data. Additional hybridizations of *Apa*I-*Not*I, *Apa*I-*Sfi*I, or *Not*I-*Sfi*I double-digest fragments were performed to establish the physical map.

When the physical map of *H. pylori* 51 was compared to that of unrelated strain, *H. pylori* 26695 (23), of which the entire genomic sequence was reported, completely different restriction patterns were shown (17).

Putting the previous reports together, several factors may contribute to the observed variation: (i) considerable diversity in DNA restriction sites, (ii) high mutational frequency of nucleotide within genes, and (iii) well-developed restriction and modification system of *H. pylori.*

In early studies including ours, the evidence that *H. pylori* would be a species characterized by a large degree of genome variability has been suggested by a variety of genomic analysis using restriction fragment length polymorphism (RFLP) analysis of genomic DNA or specific genes (1,8,9,12,16), PFGE fingerprinting (16), or randomly amplified polymorphic DNA (RAPD) fingerprinting (7,8).

Genomic diversity in *H. pylori* is much greater than that observed in other bacterial species whose DNA gives a characteristic fingerprint after PFGE. Although heterogeneity of genome sizes among natural isolates of Escherichia coli has recently been reported (11), the genomic maps of *E. coli* K-12 and *Salmonella typhimurium* can be aligned showing signifi-cant conservation (18).

Analysis of genetic map. 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 on the chromosome of *H. pylori* 51 by hybridization to the maxiset BAC library and/or to PFGE-separated chromosomal restriction fragments. The gene probes used in this study were generated by PCR using public data or data from GenBank (Table 1). With each gene probe but rRNA genes, all BAC DNAs to which it hybridized were mapped to just one location. 16S rRNA and 23S rRNA probe hybridized to BACs from two chromosomal loci, respectively.

The genetic map of *H. pylori* 51 was compared with those of strain 26695 (23) and J99 (2). The urease gene was used as the reference point for comparing three genetic maps. In all three *H. pylori* strains, most genes were present in a single copy and 16S and 23S rRNA genes were present in two copies. In the case of fucosyltransferase gene (*fucT*), however, strain 51 was confirmed to have only one copy unlike strain 26695 and J99 of which the gene copy numbers were two. The chromosomal loci of eight genes (*cagA, flaB, frdC, ggt, hpn, tsaA, ureB*, and *vacA*) and one copy of 16S and one copy of 23S rRNA genes were generally conserved among three strains. In contrast, there

Table 1. Oligonucleotide primers used for PCR amplification

Locus	Gene	Orientation	Sequence
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	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'-TAGAGGCTTTTCTTGGCACG-3'

were differences in the loci of *flaA, fucT,* and *hpaA* between strain 51 and J99 and in those of f*laA, fucT, hpaA metG, sodB*, and one copy of 23S rRNA genes between 51 and 26695 (Fig. 1). In addition to gene location, the genetic maps of these three strains showed some

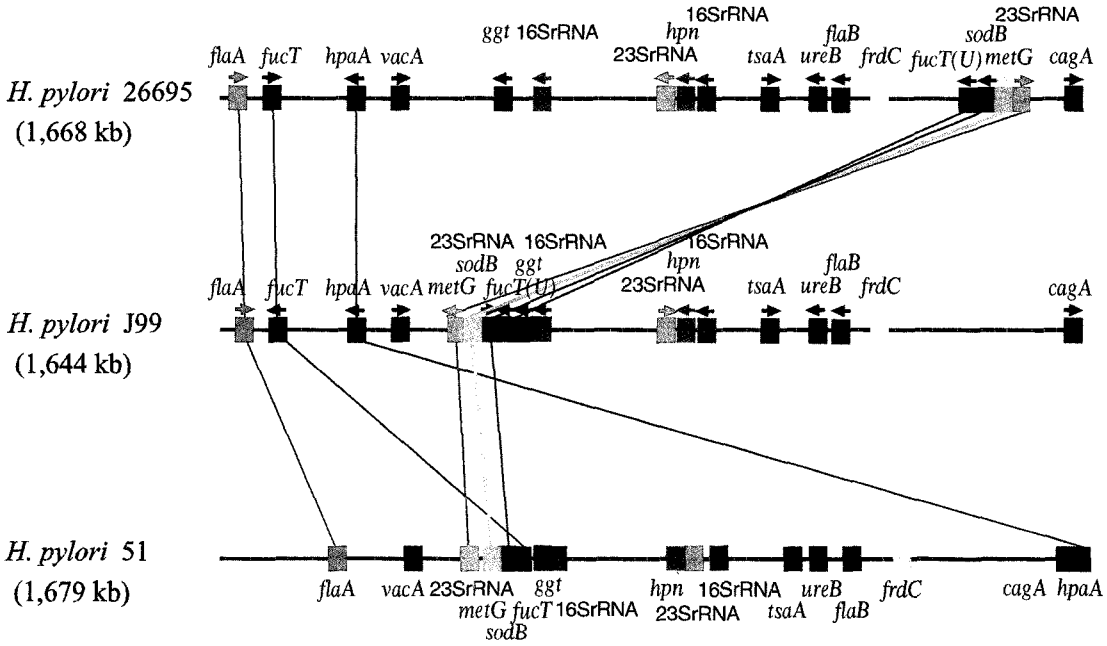
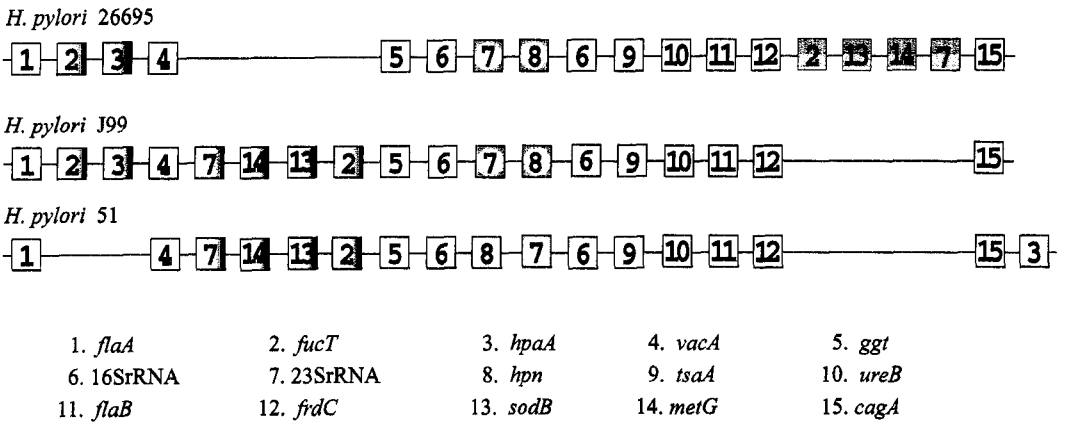


Figure 1. Comparison of the genetic maps of *H. pylori* strain 51, 26695, and J99.



1. *flaA* 2. *fucT* 3. *hpaA* 4. *vacA* 5. *ggt*
6. 16SrRNA 7. 23SrRNA 8. *hpn* 9. *tsaA* 10. *ureB*
11. *flaB* 12. *frdC* 13. *sodB* 14. *metG* 15. *cagA*

Figure 2. Schematic comparison of the gene order on the chromosomes of *H. pylori* strain 51, 26695, and J99.

variability in gene order (Fig. 2).

For further analysis on the genomic diversity of *H. pylori*, 150 *H. pylori* strains were isolated from patients at the Gyeongsang National University Hospital. Digestion of chromosomal DNA from *H. pylori* isolates with *Not*I and PFGE analysis gave widely different restriction patterns. Eighty-five isolates (57%), unexpectedly, were confirmed not to be cut with *Not*I. The remaining isolates showed 4 to 9 well-resolved DNA fragments ranging from 50 to 700 kb. When comparing the isolates with one another, PFGE gave no characteristic fingerprints for *H. pylori* DNA and almost all these isolates gave different patterns. Among them, fifteen isolates of which fingerprint patterns were estimated to be roughly similar to that of *H. pylori* 51 were selected and their

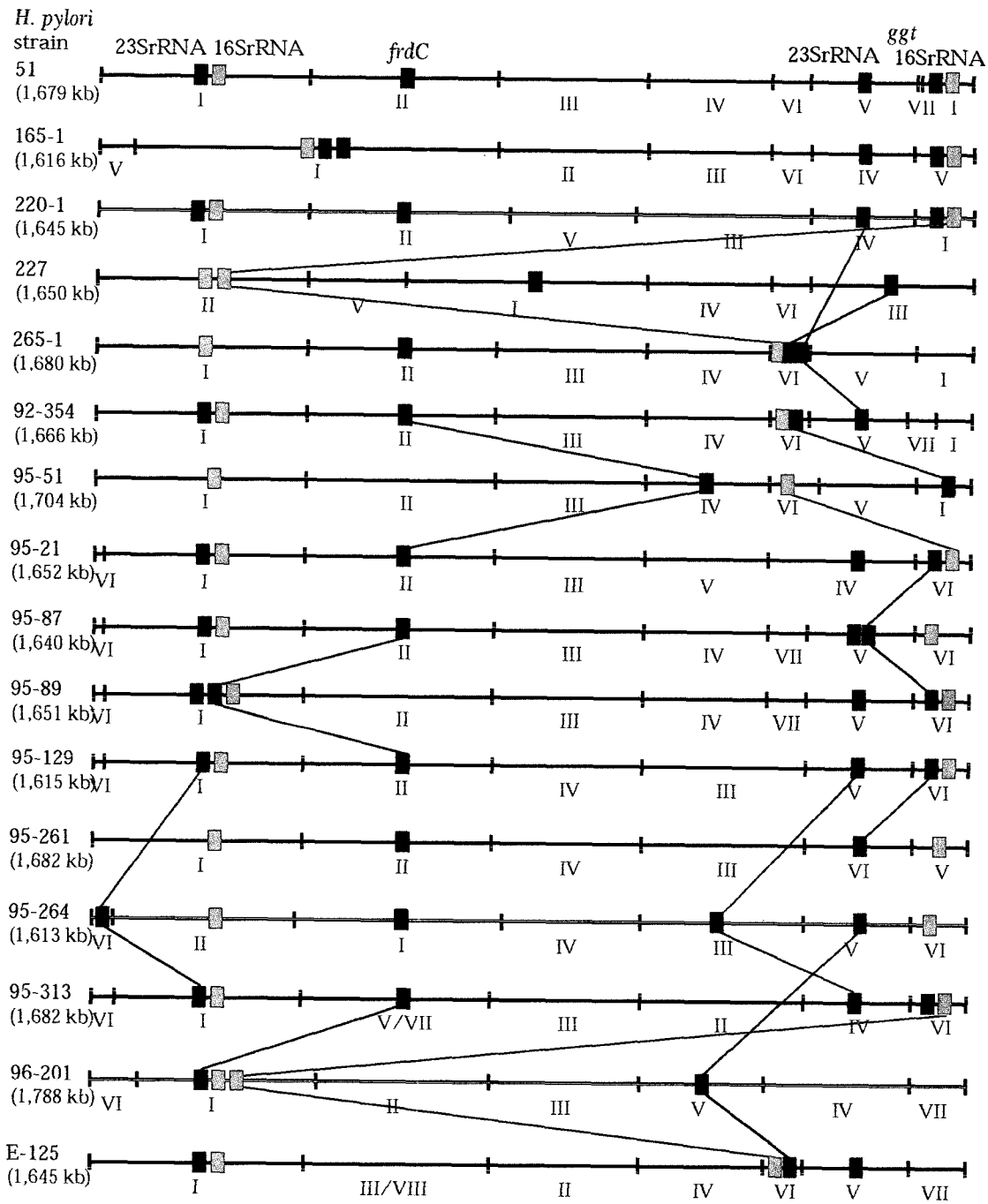


Figure 3. Comparison of physical and genetic maps among *H. pylori* strains isolated from patients at the Gyeongsang National University Hospital.

physical maps were constructed by hybridization of radiolabelled chromosomal *Not*I-fragments of *H. pylori* 51 to Southern blots prepared from PFGE of the isolates. The sizes of the genomes varied between 1.61 and 1.79 Mb.

Four genes (*frdC, ggt*, and 16S and 23S rRNA genes) were assigned to the maps by the same method as above. When the fifteen genome

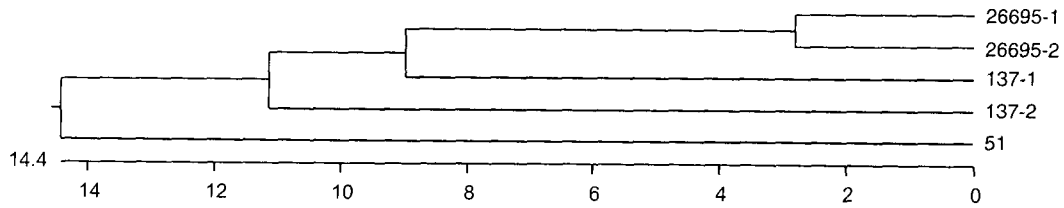


Figure 4. Phylogenetic tree of *fucT* amino acid sequences of 2 *H. pylori* Korean strains (*H. pylori* 51 and 137), and a reference strain (*H. pylori* 26695). There are just two copies of each *fucT* gene in the strain 137, and 26695.

maps were compared, significant diversity in gene location as well as gene order was evident among them, suggesting considerable flexibility in *H. pylori* genome organization (Fig. 3).

Our study demonstrates that genomic macrodiversity of *H. pylori* was characterized by several features: whether or not susceptible to restriction digestion of the chromsome, variation in chromosomal restriction fingerprint (i.e. variation in physical map) and/or high frequency of gene rearrangement.

Analysis of genetic diversity. It has been suggested that genomic diversity in *H. pylori* also would occur at the individual gene level (microdiversity), e.g. in the allelic variation seen in PCR-RFLP analysis of several genes such as urease genes, *ureA, ureB, ureC*, and *ureD* (1, 13). Similar variability was also reported in *vacA* gene (3,8), *flaA* and *flaB* genes (12), 16S ribosomal genes (9), and *copA* and *copP* (14), which encode a P-type ATPase, etc. In this study, we examined the extent of allelic variation at the level of nucleotide or deduced amino acid sequences by analyzing four genes, *ggt, cagA, vacA*, and *fucT*, which have been supposed to be important virulence factors of *H. pylori*.

Hybridization of gene probes, the same PCR products as previously used in the genetic mapping, was performed to dot-blot array filters containing a random plasmid library of *H. pylori* 51, of which details were described below. Plasmid DNAs hybridized strongly to each probe were randomly selected and used as templates for forward- and reverse-sequencing reactions. Nucleotide sequences were determined by the dideoxy chain termination method with an automated DNA sequencer (ABI 373 DNA Sequencer, Perkin-Elmer Applied Biosystems) and ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems). Sequence fragments were assembled to complete open reading frame (ORF) of each gene with the SeqMan II software (LASERGENE System, DNASTAR Inc.). A search of databases for homologous sequences was accomplished by using the BLAST network service of the National Center for Bio-htechnology Information (NCBI).

Sequence comparisons of the entire *ggt* ORF (1,701 bp) from *H. pylori* strain 51 with those of strain 26695 or J99 indicated that there was about 97% nucleotide identity among the three homologs, reflecting a good conservation of this gene (data not shown).

When the α (1,3)-fucosyltransferase gene (*fucT*) of *H. pylori* 51 was sequenced, a 1,368-bp ORF encoded 455 amino acids was present. A PCR product from another Korean isolate, *H. pylori* 137 was generated with primers that flanked the *fucT* region in *H. pylori* 51. Sequence analysis of the subcloned PCR product revealed that strain 137, in contrast to strain 51, carried two *fucT* homologs which encoded 439 and 442 amino acids, respectively. A comparison between the two *fucT* homologs of strain 137 by deduced amino acid sequence showed an unexpected low identity (82%). In further comparison among two Korean strains (*H. pylori* 51 and 137) and a foreign strain (*H. pylori* 26695)

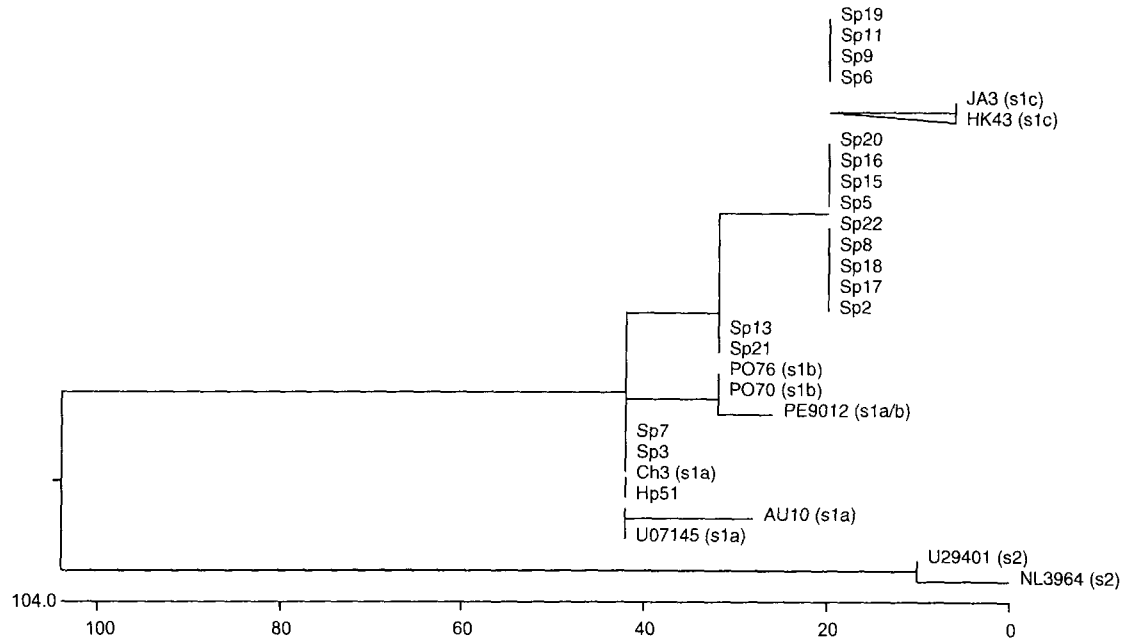


Figure 5. Phylogenetic tree of *vacA* s-region nucleotide sequences of 18 *H. pylori* Korean strains (*H. pylori* 2, 3, 5, 6, 7, 8, 9, 11, 13, 15, 16, 17, 18, 19, 20, 21, 22 and 51), and 10 reference strains (*H. pylori* Ch3, AU10, U07145, PE9012, PO70, PO76, HK43, JA3, NL3964, and U29401). Letters in the names of the isolates indicate the country or location of origin: Australia (AU), China (Ch), Hong Kong (HK), Japan (JA), Peru (PE), Portugal (PO), and Netherlands (NL).

(23), there was higher degree of sequence variation (data not shown): about 66% identity of deduced amino acid sequence between 51 and 26695; 72~81% identity between 137 and 26695; and 73% identity between 51 and 137, respectively. The greatest genetic diversity was observed in the C-terminal region of the predicted protein (41~64% identity). Phylogenetic analysis by Clustal method with MEGALIGN software (LASERGENE System, DNASTAR Inc.) showed obvious association between sequence variants and the geographic origins of the isolates (Fig. 4).

Analysis of the genetic diversity of the vacuolating cytotoxin gene (*vacA*) focused on variation in the previously defined s and m regions, as determined by PCR and direct sequencing. It was reported that the *vacA* contained at least two variable parts. The s region (encoding the signal peptide) existed as s1 or s2 allelic types. Among type s1 strains, subtype s1a, s1b, and s1c have been identified. The m (middle) region occurred as m1 or m2 allelic types (10).

It is commonly believed that virulent *H. pylori* could produce severe gastroduodenal symptoms like peptic ulcers. Particular *vacA* s genotype and the presence of *cagA* locus has been used as a marker of the pathogenicity of an individual strain in the Western World (3,10). The virulent *H. pylori* is designated as type 1 strain which carries both *cagA* and *vacA* genotype s1. Our previous studies demonstrated that most of *H. pylori* Korean isolates were type 1 strains regardless of the clinical consequences (8). However, a lot of Korean strains have not yet been classified into subtype systems established for strains from other countries.

For analysis of the s region, PCR amplicons with the expected size (323 bp) were generated from 18 Korean isolates including strain

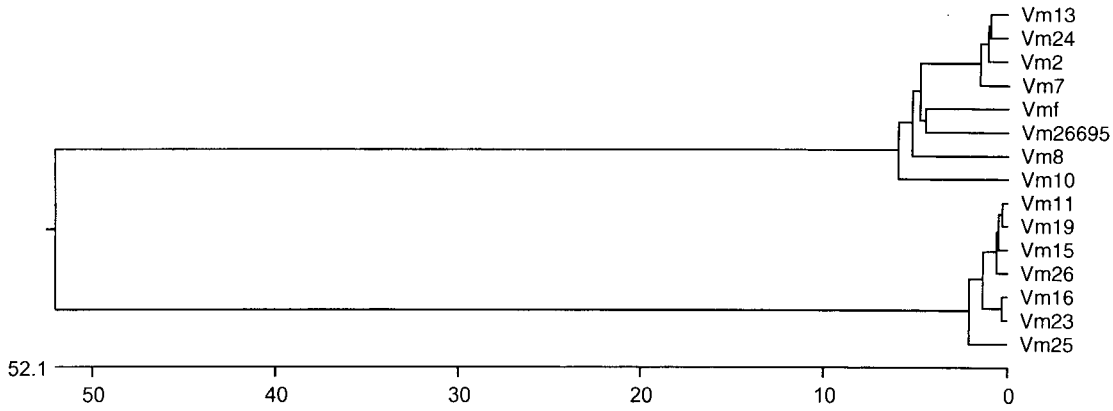


Figure 6. Phylogenetic tree of *vacA* m-region nucleotide sequences of 14 *H. pylori* Korean strains (*H. pylori* 2, 7, 8, 10, 11, 13, 15, 16, 23, 24, 25, and 26) and a reference strain (*H. pylori* 26695).

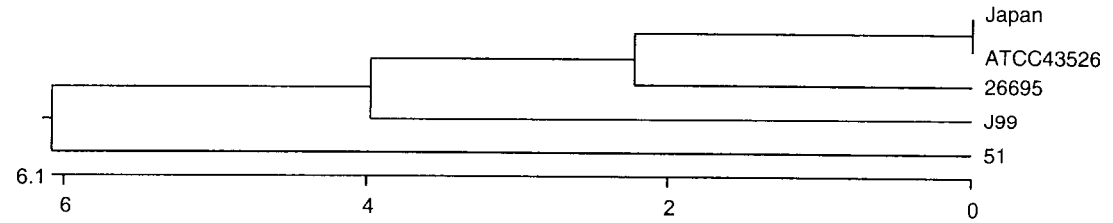


Figure 7. Phylogenetic tree of *cagA* nucleotide sequences of 5 *H. pylori* strains, including 4 reference strains of different geographic origins, Japan strain, ATCC43526, 26696, and J99.

51 and were sequenced. All sequences were compared with one another and with the existing sequence data obtained from 10 *H. pylori* isolates of different geographic origins as a reference (10). Phylogenetic analysis revealed the existence of three distinct types of s-region alleles in the Korean isolates. Three (Sp3, Sp7 and Hp51) and two (Sp13 and Sp21) among 18 isolates were categorized as s1a and s1b subtypes, respectively. However, all the remaining 13 isolates (72%) formed a new homogeneous cluster in the phylogenetic tree (>93% nucleotide identity in all comparisons), implying the possibility of a new subtype aside from the existing s1a, s1b, s1c, and s2 allelic subtypes and the existence of geographic segregation of *H. pylori* isolates within s1 subtypes (Fig. 5).

In order to analyze m region, 480-bp PCR amplicons generated from 14 Korean isolates were examined by sequence analysis. Phylogenetic tree showed a clear separation between m1 and m2 allelic types (Fig. 6). Although nucleotide sequence homology between m1 and m2 alleles was very low (ranged from 47~52% nucleotide identity), both m1 and m2 sequences were well conserved among the same allelic sequences (>92% and >94% nucleotide identity in m1 and m2, respectively).

Our results confirm and extend the finding of previous studies suggesting marked variation of nucleotide sequences of *vacA* gene. Particularly, the fact that a novel subtype is observed exclusively in isolates from Korea and appears to be the major s1 allele indicates the substantial s-region heterogeneity.

The *cagA* ORF (3,573 bp) of *H. pylori* 51

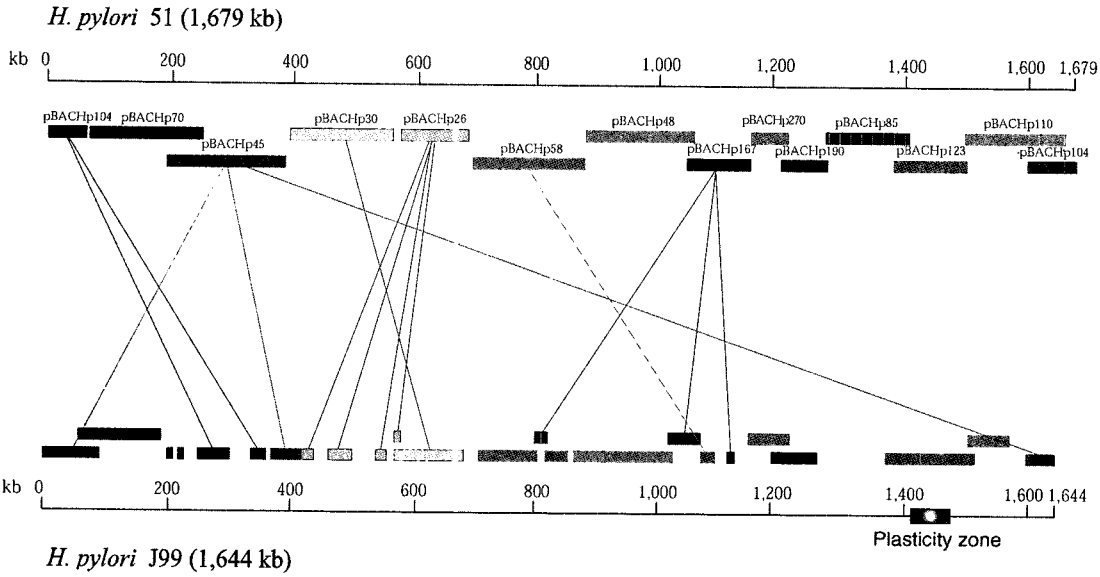


Figure 8. Comparison of the overall genomic organization between *H. pylori* strain 51, and J99.

was sequenced and compared with those of 4 foreign strains, 26695, J99, ATCC43526, and Japan strain. There was 66~79% similarity in nucleotide sequence among the 5 homologs (data not shown). Phylogenetic analysis of the entire *cagA* sequences of them revealed that the strain 51 formed a separate branch with the most far distances from the other strains (Fig. 7).

It may be considered that allelic variation (microdiversity) could arise by two mechanisms, either by a progressive accumulation of point mutations in different strains over time, or by horizontal gene transfer of sequences between strains allowing numerous reassortment of a few variant sequence.

Analysis of plasticity zone and strain-specific sequence. The plasticity zone is a region of lower (G+C) percentage and contains 46% and 48% of the genes that are unique to *H. pylori* 26695 and J99, respectively. Although this region is continous in J99, it is split in 26695 into two domains that are separated by ~600 kb (2).

We analyzed the plasticity zone and strain-specific genes of *H. pylori* 51 and compared it with the previously reported two genomic sequences.

A plasmid library of *H. pylori* 51 was constructed to facilitate molecular genetic analysis of the plasticity zone and to prepare plasmid DNA templates for random shotgun sequencing of the genome. DNA fragments with an average size of 1.6~2.0 kb were isolated from mechanically sheared *H. pylori* chromosome and cloned into the pTZ19U vector by a two-step ligation procedure (11) to minimize contamination of double-insert chimeras or free vector. Aliquots of 18,000 transformants culture from forty-eight 384-well microtiter plates were arragnged in a highly packed dot-blot array on 12 nylon filters (1,536 dots/filter) and their DNAs were fixed to filters for hybridization. The set of filters was used as a random plasmid library of *H. pylori.*

Plasmid DNAs were assorted into 13 groups by hybridization of BAC DNAs comprising a miniset contig to filters carrying the random plasmid library. And then, plasmid templates radomly selected from each group were sequenced and analyzed by the BLAST network service of the NCBI. Of a total 987 distinct sequence fragments, 59 turned to be as strain-

Table 2. Distribution of strain-specific genes on the chromosome of *H. pylori* #51

Order in BAC clone contig	BAC clone	Insert size (kb)	No. of sequence fragment	Proportion of sequence-matching ORFs to total ORFs in corresponding region of J99 genome map (%)	Relative ratio**	No. of *H. pylori* 51-specific sequence fragment
1	pBACHp123	130	137	82	1.00	16
2	pBACHP110	168	64	63	0.77	5
3	pBACHP104	131	52	47	0.57	1
4	pBACHP70	191	100	56	0.68	5
5	pBACHP45	196	109	62	0.76	6
6	pBACHP30	163	86	71	0.86	3
7	pBACHP26	115	49	39	0.48	1
8	pBACHP58	171	81	44	0.54	6
9	pBACHP48	184	85	54	0.66	4
10	pBACHP167	86	57	58	0.71	4
11	pBACHP270	66	40	47	0.57	3
12	pBACHP190	84	48	58	0.71	0
13	pBACHP85	132	79	N.D.*	–	5
	Total		987			59

*Not determined. **Relative ratio is defined as each proportion/proportion of ORFs matched to sequence fragments from pBACHp123.

specific DNA sequences which were unique to *H. pylori* 51. It was identified that pBACHp 123 would correspond to the chromosomal region encompassing the plasticity zone of J99 by sequence matches of conserved ORFs. However, of the 59 *H. pylori* 51-specific sequence fragments, only 16 (27%) were detected from the pBACHp123, in contrast to the data that there was 42 J99-specific ORFs (48%) in the plasticity zone. The remaining strain-specific sequence fragments were allocated to another 11 BAC DNAs with variation in number (from 1 to 6 fragments, Table 2). Moreover, considering the proportion of sequence-matching ORFs to total ORFs in corresponding regions of J99 chromosome, there is a possibility that additional specific sequences can be found out of all of BAC DNAs. Therefore, it could be concluded that the *H. pylori* 51-specific sequences are dispersed on the chromosome, not congregated in the plasticity zone, unlike J99- or 26695-specific genes.

The remaining 928 sequence fragments were confirmed to possess the counterpart ORFs in 26695 or J99. When comparing their relative loci on the chromosome between 51 and J99, 197 sequence fragments (21%) were found to be located beyond the loci of the corresponding J99 ORFs (Fig. 8). This finding demonstrates that there is an extensive gene shuffling between the two strains, and that differences in genomic organization and gene order between Korean and foreign isolates are much greater than those observed between the two

foreign isolates, J99 and 26695 (2).

CONCLUSION

We compared the genomic structure of *H. pylori* 51 with other Korean isolates as well as with the reported genomic sequences of *H. pylori* 26695 and J99. The genome of *H. pylori* 51 shows differences in the overall genomic organization, gene order, and even in the nucleotide sequences among the *H. pylori* strains, which are far greater than the differences reported on the genomic diversity of *H. pylori* (2).

The genomic diversity noted among *H. pylori* strains could reflect a long evolutionary history or genomic rearrangements after infection, perhaps in response to and adaptation to a new environment. This genomic diversity could enable the organism to cause chronic infection and diverse gastroduodenal diseases.

Further genomic analysis of *H. pylori* strains would provide a discriminatory tool for typing the clinical isolates and an insight into the host-parasite relationship for the breakthrough in the control of *H. pylori* infection.

Acknowledgements

This study was supported by research Grants (96-, 97- and 98-NF-04-03-A-01) from National R&D Program of the Ministry of Science and Technology and in part by Grant from Brain Korea 21 Project of the Ministry of Education (Jae-Young Song, Sang-Haeng Choi, Seong-Gyu Park, Mi-Young Choe, and Byung-Sang Lee), Korea. Yeo-Jeong Choi is a recipient of the Internship Program from KOSEF, Korea.

REFERENCES

1) Akopyanz N, Bukanov NO, Westblom TU, Berg DE: PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobactrer pylori. Nucl Acids Res* **20:** 6221-6225, 1992.

2) Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ: Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori. Nature* **397:** 176-180, 1999.

3) Atherton JC, Cao P, Peek Jr. RM, Tummuru MKR, Blaser MJ, Cover TL: Mosaicism in vacuolating cytotoxin alleles of *Helicobactrer pylori. J Biol Chem* **270:** 17771-17777, 1995.

4) Baik SC, Kim JB, Cho MJ, Kim YC, Park CK, Ryou HH, Choi HJ, Rhee KH: Prevalence of *Helicobacter pylori* infection among normal Korean adults. *J Kor Soc Microbiol* **25:** 455-462, 1990.

5) Baik SC, Youn HS, Chung MH, Lee WK, Cho MJ, Ko GH, Park CK, Kasai H, Rhee KH: Increased oxidative DNA damage in *Helicobacter pylori*-infected human gastric mucosa. *Cancer Res* **56:** 1279-1282, 1996.

6) Bergthorsson U, Ochman H: Heterogeneity of genome sizes among natural isolates of *Escherichia coli. J Bacteriol* **177:** 5784-5789, 1995.

7) Cho MJ, Lee WK, Jeon YS, Kim KH, Kim SH, Baik SC, Rhee KH, Kim YO, Youn HS, Kim NS: Intrafamilial transmission of *Helicobactrer pylori* detected by random amplified polymorphic DNA fingerprinting. *Mol Cells* **5:** 508-513, 1995.

8) Cho MJ, Song JY, Lee WK, Baik SC, Choi YJ, Choi SH, Jeon YS, Lee IG, Lee BS, Youn HS, Rhee KH: Genotyping and molecular epidemiology of *Helicobacter pylori* strains isolated from Korea. *J Kor Soc Microbiol* **34:** 189-199, 1999.

9) Desai M, Linton D, Owen RJ, Cameron H, Stanley J: Genetic diversity of *Helicobacter*

pylori indexed with respect to clinical symptomatology, using a 16S rRNA and a species-specific DNA probe. *J Appl Bacteriol* **75:** 574-582, 1993.

10) Doorn LJ, Figueiredo C, Sanna R, Pena S, Midolo P, Ng EKW, Atherton JC, Blaser MJ, Quint WGV: Expanding allelic diversity of *Helicobacter pylori vacA*. *J Clin Microbiol* **36:** 2597-2603, 1988.
11) Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C, Gocayne JD, Scott J, Shirley R, Liu LI, Glodek A, Kelley JM, Weidman JM, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Branson RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496-512, 1995.
12) Forbes KJ, Fang Z, Pennington TH: Allelic variation in the *Helicobactrer pylori* flagellingenes *flaA* and *flaB*: its consequences for strain typing schemes and population structure. *Epidemiol Infect* **114:** 257-266, 1995.
13) Foxall PA, Hu LT, Mobley HLT: Use of polymerase chain reaction-amplified *Helicobactrer pylori* urease structural genes for differentiation of isolates. *J Clin Microbiol* **30:** 739-741, 1992.
14) Ge Z, Hiratsuka K, Taylor DE: Nucleotide sequence and mutational analysis indicate that two *Helicobactrer pylori* genes encode a P-type ATPase and a cation-binding protein associated with copper transport. *Mol Microbiol* **15:** 97-106, 1995
15) Honda S, Fujioka F, Tokieda M, Satoh R, Nishizono A, Nasu M: Development of *Helicobacter pylori*-induced gastric carcinoma in Mongolian gerbils. *Cancer Res* **58:** 4255-4259, 1988.
16) Lee WK, Cho MJ, Kim KH, Kim SH, Park PS, Lee SY, Baik SC, Rhee KH: Identification of *Helicobactrer pylori* strains by pulse-field gel electrophoresis and restriction fragment length polymorphism. *J Kor Soc. Microbiol* **30:** 641-650, 1995.
17) Lee WK, Ryu BD, Baik SC, Yoon YH, Choi YJ, Song JY, Choi SH, Lee IG, Jeon YS, Cho MJ, Rhee KH: Physical map of the *Helicobacter pylori* chromosome. *J Kor Soc Microbiol* **33:** 55-68, 1998.
18) Liu SL, Hessel A, Sanderson KE: Genomic mapping with I-*Ceu*I, and intron-encoded endonuclease specified for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. *Proc Natl Acad Sci USA* **90:** 6874-6878, 1993.
19) Rhee KH, Youn HS, Baik SC, Lee WK, Cho MJ, Choi HJ, Maeng KY, Ko KW: Prevalence of *Helicobacter pylori* infection in Korea. *J Kor Soc. Microbiol.* **25:** 475-490, 1990.
20) Rhee KH: *Campylobacter pylori* as a major determinant in hypothetical carcinogenesis of stomach cancer. *J Kor Soc. Microbiol.* **23:** 550-552, 1988.
21) Rhee KH: Carcinogenesis of stomach cancer by *Helicobacter pylori* infection. *J Kor Soc. Microbiol.* **25:** 583-589, 1990.
22) Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M: 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, 1992.
23) Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quachenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B,

Rhchardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzegerald 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* **388:** 539-545, 1997.

24) World Health Organization. Schistosomes, liver flukes and *Helicobacter pylori.* IARC Monographs on the Evaluation of Carcinogenic Risks to Human. **61:** 177-240, 1994.

25) Youn HS, Ko GH, Chung MH, Lee WK, Cho MJ, Rhee KH: Pathogenesis and prevention of stomach cancer. *J Kor Med Sci* **11:** 373-385, 1996.