[SVII-2]

Whole Genome Sequence of a Korean Isolate (strain 51) of *Helicobacter pylori*

Woo Kon Lee*, Myung Je Cho, Seung Chul Baik, Jae Young Song, Jeong Uck Park, Hyung Lyun Kang, Hee Shang Youn[1], Gyung Hyuck Ko[2], Kwang Ho Rhee

Department of Microbiology, Pediatrics[1] and Pathology[2], Gyeongsang National University College of Medicine, Jinju, Gyeongsangnam-do 660-751, South Korea

ABSTRACT

Substantial genomic diversity has been expected among clinical isolates of *H. pylori*. We have suggested that the two complete *H. pylori* genomes already sequenced may be insufficient for providing a discriminatory tool for typing clinical isolates as well as an insight into the genomic diversity, which enable to establish strategy for control of *H. pylori* infection. In this study, we determine the nucleotide sequence of the entire genome of Korean strain 51 and compare it with two reported genomic sequences to suggest validity for extensive genomic sequencing of *H. pylori*. The genome of *H. pylori* 51 consists of a circular chromosome with a size of 1,591,297 bp, which is corresponding to 95.4% and 96.8% of the 26695 and J99 chromosome length, respectively. We predict that there are 1,454 open reading frames (ORFs) in 51, representing 91.4% and 97.2% of the reported numbers of ORF of 26695 and J99, respectively. In contrast to 26695 and J99 that have 123 and 65 strain-specific genes, respectively, of the 1,454 genes, only 39 genes are unique to 51. Differences in genomic organization between 51 and each foreign strain were greater than between 2 foreign strains in pair wise entire sequence alignments by BLASTN. Particularly, the extent of genomic rearrangement observed between 51 and 26695 is higher than between 51 and J99. Multiple sequence alignment of orthologous genes among 3 strains showed that 51 is genetically closer to 26695 rather than J99. Phylogenetic analysis of nonsynonymous and synonymous mutation indicated J99 has the longest branch length in the unrooted phylogenetic tree, suggesting that J99 has higher mutation rate than the other 2 strains.

PURPOSE

H. pylori is the most prevalent chronic bacterial infection of human being, especially in Korea. The high prevalence rate of *H. pylori* infection clearly explains the high incidence of chronic gastritis, peptic ulcer diseases and gastric cancer in Korea. Due to its importance as a human pathogen, its whole genome sequence has been determined for two strains: 26695 by TIGR in 1997 and J99 by Astra-Zeneca and Genome Therapeutics in 1999. The genomic sequence information demonstrated that 6-7% of total predicted coding sequences is strain-specific.

Microdiversity and macrodiversity of *H. pylori* genome cause difficulty for the establishment of strain classification system by DNA technology. A taxonomic principle for strain classification requires genomic sequence information of at least ten *H. pylori* strains. Accomplishment of classification system for *H. pylori* strains will provide the turning point to investigate pathogenesis of gastroduodenal diseases and carcinogenesis of gastric cancers.

With this goal in mind, the whole-genome sequencing of Korean strain 51 has been undertaken as part of the Korean Functional Genome Project.

Table 1. Number of orthologous or strain-specific genes in each strain

Specificity	51	26695	J99
All strains	1,346	1,370	1,352
Shared only between 51 & 26695	46	46	–
Shared only between 51 & J99	23	–	26
Shared only between J99 & 26695	–	52	53
Each strain only	39	123	65
Total	1,454	1,591	1,496

Table 2. Weighted average of amino-acid and nucleotide identities between *H.pylori* orthologues

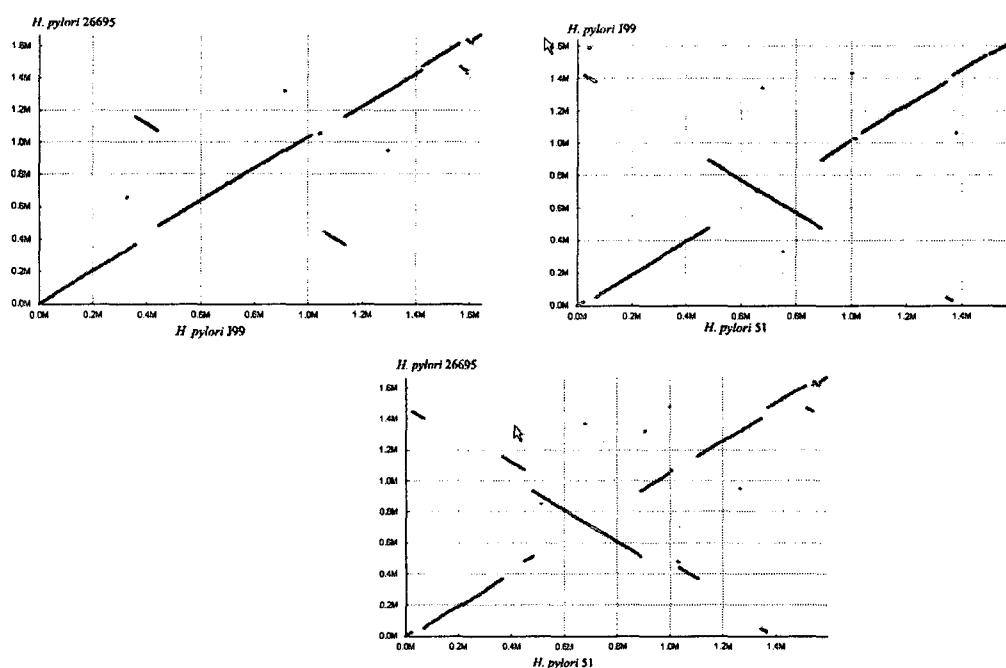|	51 vs 26695	51 vs 26695	26695 vs J99
Number of compared orthologous genes	1,304	1,304	1,304
Total number of compared amino-acid residues including gaps	427,080	428,428	428,884
Total number of compared nucleotide residues including gaps	1,220,142	1,222,843	1,224,472
Nucleotide identity	95.10 %	93.80 %	94.53 %
Amino-acid identity	94.78 %	93.83 %	94.61 %

Figure 1. Alignment of orthologous genes in genomic scale

Table 3. Number of orthologues in each amino-acid identity range

Amino-acid identity range (%)	51-26695 (%, cumulative)	51-J99 (%, cumulative)	26695-J99 (%, cumulative)
96 - 100	773 (59.3)	570 (43.7)	699 (53.6)
92 - 95	397 (89.7)	492 (81.4)	446 (87.8)
88 - 91	76 (95.6)	167 (94.2)	100 (95.5)
83 - 87	23 (97.3)	38 (97.2)	34 (98.1)
79 - 82	7 (97.9)	14 (98.2)	7 (98.6)
70 - 78	19 (99.3)	11 (99.1)	12 (99.5)
0 - 69	9 (100)	12 (100)	6 (100)
Total compared orthologues	1,304	1,304	1,304

Table 4. Comparison of IS elements among 3 strains

Insertion element	H. pylori 26695	H. pylori J99	H. pylori 51
Complete IS605 copies	5	0	0
Partial IS605 copies	8	5	7
Complete IS606 copies	2	1	0
Partial IS606 copies	2	4	10
Complete IS607 copies	0	0	1
Complete IS10 copies	0	0	1

CONCLUSION

The taxonomic system for classification of strain should be useful in estimating pathogenicity or infectivity of the bacterial isolates and should be applicable for epidemiological approaches such as the transmission route of bacterial infection. No classification system for *H. pylori* strain could be established on the basis of serological, biological and molecular biological features up to date. Because the complete genomic sequence contains all biological parameters, its high through-put analysis will throw the light on the establishment of the classification system for *H. pylori* strains. In this study, the features of genomic diversity among *H. pylori* isolates begin to be obvious. The ultimate objective of this project is to validate the fact that genomic comparisons among minimal ten strains would be a prerequisite for establishing the classification system for clinical isolates and understanding the host-parasite relationship of *H. pylori* infection.

REFERENCES

1. Alm RA, Ling LS, Moir DT, King BL, Brown ED, et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176-180, 1999.

2. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496-512, 1995.

3. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, et al.The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388:539-545, 1997.