The Geographic Origin of Helicobacter pylori Influences the Association of the homB gene with Gastric Cancer

Jieun Kang1#, Kathleen R. Jones2#, Sungil Jang1, Cara H. Olsen3, Yun-Jung Yoo1, D. Scott Merrell2*, and Jeong-Heon Cha1,4*

1 Department of Oral Biology, Oral Science Research Center, BK21 Project, Research Center for Orofacial Hard Tissue Regeneration, Yonsei University College of Dentistry, Seoul, South Korea

2 Department of Microbiology and Immunology, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814, USA

3 Department of Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814, USA

4 Oral Cancer Research Institute, Yonsei University College of Dentistry, Seoul, South Korea

# These authors contributed equally to this work

* Corresponding authors, Mailing addresses:

Jeong-Heon Cha, Department of Oral Biology, Yonsei University College of Dentistry, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, Korea, Phone:+82-2-2228-3061, Fax:+82-2-2227-7903, E-mail:Jcha@yuhs.ac

D. Scott Merrell, Department of Microbiology and Immunology, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814, USA, Phone: (301) 295-1584, Fax: (301) 295-3773, E-mail:dmerrell@usuhs.mil
Abstract
We found that South Korean *H. pylori* isolates predominantly carry homB at locus B, and that there is no association between the homB allele and the cagA allele or the development of gastric cancer within this population. Uniquely, several East Asian strains carried multiple copies of the hom genes.
Body

*Helicobacter pylori* colonizes the gastric mucosa of over 50% of the world’s population (12, 22), and is the etiological agent of gastritis, duodenal ulcers, gastric ulcers, gastric adenocarcinoma, and mucosa-associated lymphoid tissue lymphoma (4, 6, 18, 19, 21). Due to this bacterium’s association with gastric cancer, which is the second most common cause of cancer-associated death (13), the World Health Organization has classified *H. pylori* as a class I carcinogen (5). Gastric cancer mortality rates vary geographically; the highest rates are in East Asian countries like China, Japan, and South Korea, which also display high rates of *H. pylori* infection (5, 7, 20, 23). Clearly gastric diseases are due, at least in part, to infection by *H. pylori*, and the ultimate disease developed appears to be affected by variability in *H. pylori* virulence factors. Recently, we presented detailed epidemiological studies of *cagA* and *vacA* from a collection of 260 isolates from South Korea (8, 9, 10). Our studies showed that there is a significant association between infection with *H. pylori* strains carrying the EPIYA-ABD *cagA* genotype and the development of gastric cancer (10). Moreover, the majority of *H. pylori* isolates encoded for the most virulent CagA (EPIYA-ABD) and VacA (s1/i1/m1) (8, 10). The polymorphisms in *cagA* and *vacA*, alone and in concert, impact the progression to severe gastric disease, but the impact of these two virulence factors alone is not sufficient to explain the vast discrepancy in gastric cancer rates in East Asian as compared to Western populations. Thus, it is important to examine the impact of different virulence factors among both Western and East Asian populations (8, 9, 10, 15).

In *vitro, Helicobacter* outer-membrane (Hom) B promotes the secretion of the pro-inflammatory cytokine IL-8 and increases *H. pylori*’s ability to adhere to host cells (14).
More importantly, *homB* presence is significantly associated with development of peptic ulcer disease in Portuguese children and young adults (14, 17), and with gastric cancer development and the presence of *cagA* in US and Colombian populations (11). These findings suggest that the outer membrane protein HomB is a novel virulence factor. Thus, it and other members of the small paralogous family of *hom* adhesion molecules are currently being investigated. The two currently best-studied *hom* genes, *homA* and *homB* are 90% identical at the nucleotide level (2, 3, 17). These *homA* and *homB* genes can be present at two different loci within the *H. pylori* genome: locus A and locus B. Strains can carry a single copy of one of the *hom* genes, a double copy of a single gene, or a single copy of each gene (14, 15). Previous studies suggest geographic differences, either in distribution, location, or copy number of the *hom* genes in the genome, and suggest that these differences influence any association with disease outcome (14, 15, 16).

In the present study, our collection of 260 South Korean isolates was assessed for any associations between the distribution of the *homA* or *homB* genes and disease state, as well as any associations between the *hom* genes and the different *cagA* and *vacA* alleles. The South Korean isolates include 115 isolates from patients diagnosed with gastritis, 60 isolates from patients diagnosed with gastric ulcers, 55 isolates from patients diagnosed with duodenal ulcers, and 30 isolates from patients diagnosed with gastric cancer (8, 9, 10). A complete description of all strains can be found in Supplementary Table 1.

To analyze the *hom* genotype of the South Korean *H. pylori* strains, the presence of the *homA* and/or the *homB* gene(s) was identified by a single PCR with the *hom* primers hf and hr (Table 1 and Fig. 1). We successfully genotyped 225 samples for which
we had complete epidemiological data for the hom genes (Fig. 2 and Supplementary Table 2). Of note, two strains showed an intermediate length amplicon (approximately 146 bp) compared to what was expected for either the homA PCR product (~128 bp) or homB PCR product (~161 bp). This intermediate length hom or ihom genotype has been previously described and shown to be due to random deletions and/or insertions within the hom genes (11).

Once the strains were genotyped for the presence of the two hom genes, we next sought to define the copy number and location of the gene(s). This was accomplished through two additional PCR reactions (Fig. 1). The distribution of homA/B is shown in Fig. 2. Within this population, 212 isolates carried a single hom gene at locus B (35 -homA, 175 -homB, and 2 -ihom). This is in contrast to Western strains that carry a single hom gene at locus A 100% of the time (14, 15). Also, this distribution is different from what has been reported for Western strains, which show a much more evenly distributed population of isolates carrying homA or homB (14, 15). In our population, three isolates were indeterminate for the presence of the homB gene at locus A but were positive for occupation at locus B (homB+/homB). These 3 homB+/homB were included as homB positive strains for the statistical analysis, but were eliminated from the data set when assessing the impact of multiple copies of the hom genes. Six strains carried multiple hom genes (4 homA/homA, 1 homB/homB, and 1 homA/homB), which is again in contrast to previous studies that suggested that Western strains carry multiple copies of homA and/or homB, but East Asian strains do not (14, 15). Our finding is perhaps not surprising since our collection of East Asian isolates is much larger than the collection previously examined.
Finally, four strains failed to amplify either *hom* gene at either locus A or locus B and therefore were considered *hom* negative. For these four +/- strains, PCR amplification of locus A and locus B yielded products that were indicative of an empty locus. However, each of the four +/- strains indicated the presence of a *hom* gene through the *hom* PCR amplification; two *homA*, one *ihom*, and one *homA* and *homB*. This suggests that for these strains, *homA* or *homB* are presumably located at an alternate unknown location within the genome. It is interesting to speculate that perhaps *homA* or *homB* may be carried in the normal locus for the virtually unstudied *homC* or *homD* gene. Further study is clearly necessary to elucidate whether the location of the *homA/B* genes corresponds to a functional difference and whether *homA/B* can be located at other location besides locus A and locus B. Of note, no strains carrying *homA/-, homB/-*, or *homB/homA* alleles were found.

We next assessed whether there was an association between the distribution of the individual *hom* gene(s) and disease state. A complete breakdown of the *hom* allele and disease state is provided in Supplementary Table 1 and Supplementary Table 2. The Fisher’s exact test was used to analyze two way associations using SAS version 9.1 software (SAS Institute Inc., Cary, NC). There was no association between the distribution of the *hom* gene and disease state among this population of East Asian strains (P=0.9978), which is in direct contrast to Western populations (11, 14). In fact, there was no association no matter which disease states were assessed (Table 2).

Since there is a statistical association with the presence of *homB* and gastric cancer in Western populations, which carry the single *hom* gene at locus A, and not in East Asian populations, which carry the single *hom* gene at locus B, these data perhaps...
suggest that the location of the hom gene within the genome is important. Genes carried
at one particular locus could be expressed at greater levels; the promoter of homA or
homB may differ enough to influence transcriptional levels of each gene, or the different
loci within the genome may provide the ability for different enhancers/inhibitors of each
hom gene to bind and influence overall levels of the hom transcript.

Recently, it has become clear that individual virulence factors interact in order to
impact H. pylori pathogenesis (1, 8, 9, 24). Since homB is associated with cagA within
Western populations (11, 14), we next assessed the distribution of hom genes, in
combination with the cagA alleles, the vacA alleles, and disease state. A complete
breakdown of the strains based on these factors is provided in Supplementary Table 3.
We first assessed if there was any association between the distribution of the hom alleles
among the different cagA alleles (a canonical EPIYA-ABD versus all other EPIYA
motifs). We found that there was no association between the distribution of hom alleles
among the different cagA alleles (P=0.0872). Furthermore, there was no association
when each gene was considered separately: distribution of homA (P=0.6139) or homB
(P=0.2217) across the different cagA alleles. We next analyzed the association between
the distribution of the hom alleles among the different vacA alleles (Supplementary Table
3) and found that the distribution of vacA alleles among the two hom genes was
statistically significant (P=0.0142). This association was dependent only on the homB
allele (P=0.0275), since there was no association between the distribution of the homA
allele and the vacA allele (P=0.3955). The overall association between the vacA alleles
and the hom alleles was also influenced by the distribution of cagA alleles; the
association was present in the non-EPIYA-ABD population (P=0.0319), but not in the
EPIYA-ABD population (P=0.1014). Due to this difference, we used log linear modeling to determine if there was a three way association between the cagA, vacA, and hom alleles. However, no association between these three virulence factors was identified (P=0.681). Another interesting aspect of this vacA/hom association, was that it appeared to be influenced by the age of the patient; the association only became evident in the population above 60 years of age (P=0.0076). A higher order association between the cagA, vacA, hom alleles and disease states was also assessed, but no significant associations existed between the virulence factors and disease state.

In conclusion, this is the first study to assess the association between the presence of the homB gene and gastric cancer in a population of predominantly East Asian strains; we found that the impact of the homB allele on disease is geographically dependent. In Western strains, there is a more even distribution of the homA and homB genes, while in East Asian strains, homB is more common (Fig. 2) (14, 15). Moreover, Western strains carry a single hom gene at locus A, whereas East Asian strains carry a single hom gene at locus B (Fig. 2) (14, 15). This study was the first to identify the presence of any East Asian isolates that carry multiple copies of the hom genes. Interestingly, in this population, no association between the presence of homB and the progression to gastric cancer was found (Table 2), suggesting that a hierarchy of virulence factors exists, and that virulence factors have different impacts on disease based on the presence of other virulence factor polymorphisms. Within East Asian strains, EPIYA-ABD CagA appears to be the “master” virulence factor. En masse, these data exemplify the need for information about the presence and function of different virulence factors within different populations, and the need to develop geographically tailored treatment regimens.
We would like to thank Jeannette Whitmire for critical reading of the manuscript.

This work was supported by the Basic Science Research Program through the National Research Foundation of South Korea (NRF) funded by the Ministry of Education, Science, and Technology (R13- 2003-013-04002-0) (to J.-H.C.). The contents of this article are the sole responsibility of the authors and do not necessarily represent the official views of the Department of Defense or the funding agency.
References:


Figure 1: Genotyping of the hom genes at the respective loci. (Top) Schematic representation of the two loci where the hom genes are traditionally found; locus A and locus B. The annealing positions (arrows) and names of the primers used in this study are shown. The presence of a hom gene in a particular locus is depicted by the presence of a dashed box. (Bottom) The strains were genotyped for the hom gene by a single PCR with the hf and hr primers. A PCR amplicon of 128 bp indicates the presence of the homA gene, and an amplicon of 161 bp denotes the presence of the homB gene. In order to determine the location (locus A or B) of the hom gene, two additional PCR reactions were performed. To amplify locus A, primers Af and Ar were used, and to amplify locus B, the Bf and Br primers were used, as previously described (17). If an indeterminate result from the PCR reaction using the Af and Ar primers or Bf and Br primers, respectively, was obtained, another PCR reaction with the K-Af and K-Ar primers or K-Bf and K-Br primers (Table 1) was performed. These K-Af/K-Ar and K-Bf/K-Br primers were designed according to the genome sequences of the Korean H. pylori strains HP51 and HP52 (Genebank accession numbers CP000012 and CP001680, respectively). For locus A a resulting amplicon of 300 to 900-bp indicates that locus A is empty whereas the presence of a 2,000 to 2,500-bp amplicon confirms that locus A is occupied by a hom gene. In the case of locus B, a 1,300 to 1,800-bp amplicon denotes that B is empty and the presence of a 2,500 to 4,000-bp amplicon indicates that locus B is occupied by a hom gene. (Left) For K3-CA the PCR reaction with the hf and hr primers yielded a single amplicon of 161 bp (second lane), denoting the presence of the homB gene. The PCR reaction with the Af and Ar primers (amplifying locus A) yielded a 600-bp product indicating that locus A is empty (third lane), whereas a 3,000-bp amplicon produced from...
the PCR reaction using the Bf and Br primers indicates that locus B is occupied (fourth lane). These results indicate that K3-CA has a genotype of /homB. (Right) For K57-G and any strains that carried both homA and homB, an additional set of nested PCR reactions was also performed. First, PCR reactions with the hf and hr primers yields two different sized amplicons, 128 bp indicating the presence of a homA allele and 161 bp indicating the presence of a homB allele (first lane). Next, a PCR reaction using the Af and Ar primers, or the K-Af and K-Ar primers yielded a 2,000-bp amplicon, which denotes an occupied locus. In the case of K57-G (right, third lane), the K-Af and K-Ar primers were used to amplify locus A because it showed no band for PCR using Af/Ar primers. This PCR product was then purified and used as the template in a PCR reaction using the hf and hr primers (fourth lane). This PCR reaction yielded a 128-bp amplicon, which indicates that homA is located at locus A. Next, a PCR reaction was performed using the Bf and Br primers, which yielded a 3,000-bp product (fifth lane). This again, indicates an occupied locus B, and this PCR product was purified and used as the template in a PCR reaction using the hf and hr primers. This PCR reaction yielded a 161-bp amplicon (right, sixth lane), which indicates that homB is located at locus B. Thereby, these results indicate that K57-G has a genotype of homA/homB.

Figure 2: Schematic representation of the distribution of hom genes at the respective loci. A schematic of the distribution of hom genes at the respective loci within this South Korean population is shown. A homB+/- indicates that the amplification of locus A was unsuccessful. Therefore, the strains are homB positive and there is at least a single copy of homB found at locus B, but whether or not they have two copies of homB cannot be
determined. These strains were included as homB positive strains for the statistical analysis, but were eliminated from the data set when assessing the impact of multiple copies of the hom genes. To the right of the schematic, percentage of the overall population for each individual genotype is indicated.
### Table 1. Primer sequences

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>hf (F1-jhp0870/jhp0649)</td>
<td>AGAGGGTGTTTGAAACGCTCAATA GGTGAATTCTTCTGCGGTTTG</td>
<td>(17)</td>
</tr>
<tr>
<td>hr (R1-jhp0870/jhp0649)</td>
<td>TAATTTCGCCGAAAAAACATC ATTCCAGCGCCTAATGGAC</td>
<td>(17)</td>
</tr>
<tr>
<td>Af (F1-jhp0648/HP0709)</td>
<td>AAGAGGATTGCGTGGTGGAGTTG GGGTTGCCTTGGGGCTTGGA</td>
<td>(17)</td>
</tr>
<tr>
<td>Ar (R1-jhp0650/HP0711)</td>
<td>TGGAATATTGATATAAGAAGTG GGGTTAATAGGATGAGCCGC</td>
<td>This study</td>
</tr>
<tr>
<td>K-Af</td>
<td>GATTTTCCCCACTCTTTTATGG GGTGTTGTCCATGAACATGC</td>
<td>This study</td>
</tr>
<tr>
<td>K-Ar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-Bf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-Br</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 2. *P* values of the Distribution of homA/B and Disease

<table>
<thead>
<tr>
<th>Comparison of the distribution of homA/B in patients with different diseases</th>
<th>P value homA</th>
<th>homB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Across all diseases</td>
<td>0.9978</td>
<td>0.9802</td>
</tr>
<tr>
<td>Gastritis vs all other diseases</td>
<td>0.8953</td>
<td>0.7213</td>
</tr>
<tr>
<td>Duodenal ulcers vs all other diseases</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Gastric ulcers vs all other diseases</td>
<td>0.9219</td>
<td>1.0000</td>
</tr>
<tr>
<td>Gastric cancer vs all other gastric diseases</td>
<td>0.8805</td>
<td>0.7931</td>
</tr>
<tr>
<td>Peptic ulcers (both duodenal and gastric ulcers) vs gastritis and gastric cancer</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>More severe disease (gastric ulcer and gastric cancer) vs less severe diseases (gastritis and duodenal ulcer)</td>
<td>0.9379</td>
<td>0.8481</td>
</tr>
</tbody>
</table>

*hom* represents the distribution of the homA, homB, and hom negative strains among the different disease states listed.
K3-CA

locus A empty

K57-G

locus A homA

locus B homB

For K3-CA:
- Locus A is empty.
- Locus B has the homB marker.

For K57-G:
- Locus A has the homA marker.
- Locus B has the homB marker.

PCR Results:
- **Forward Primer**:
  - K3-CA: hf
  - K57-G: hf

- **Reverse Primer**:
  - K3-CA: hr
  - K57-G: hr K-Af

**Nested PCR**:
- K3-CA: K-Af
- K57-G: K-Af

**Genotype Analysis**:
- **-homB**
  - K3-CA: 161 bp
  - K57-G: 128 & 161 bp

- **homA/homB**
  - K3-CA: 600 bp
  - K57-G: 2000 & 128 bp

(bp: base pairs)
<table>
<thead>
<tr>
<th><strong>H. pylori genome</strong></th>
<th>locus A</th>
<th>locus B</th>
<th>Single-copy genotype</th>
<th>35 / 225 = 15.6 %</th>
<th>94.2 %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>homA</td>
<td>ihom</td>
<td>homB</td>
<td>175 / 225 = 77.8 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 / 225 = 0.9 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Double-copy genotype</strong></td>
<td>homA</td>
<td>homA</td>
<td>homB</td>
<td>4 / 225 = 1.8 %</td>
<td>2.7 %</td>
</tr>
<tr>
<td></td>
<td>homB</td>
<td></td>
<td>homB</td>
<td>1 / 225 = 0.4 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>homA</td>
<td>homB</td>
<td></td>
<td>1 / 225 = 0.4 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>homB</td>
<td>homA</td>
<td></td>
<td>0 / 225 = 0 %</td>
<td></td>
</tr>
<tr>
<td><strong>None</strong></td>
<td></td>
<td></td>
<td></td>
<td>4 / 225 = 1.8 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>homB*</td>
<td>homB</td>
<td></td>
<td>3 / 225 = 1.3 %</td>
<td></td>
</tr>
</tbody>
</table>