Multiple Chromosomal Loci for the \textit{babA} Gene in \textit{Helicobacter pylori}

Ewa E. Hennig, Johnna M. Allen and Timothy L. Cover


Updated information and services can be found at:
http://iai.asm.org/content/74/5/3046

These include:

**REFERENCES**

This article cites 15 articles, 11 of which can be accessed free at: http://iai.asm.org/content/74/5/3046#ref-list-1

**CONTENT ALERTS**

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml
To subscribe to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
Multiple Chromosomal Loci for the babA Gene in Helicobacter pylori

Ewa E. Hennig,1 Johnna M. Allen,2 and Timothy L. Cover2,3,4,*

Department of Gastroenterology, Medical Center for Postgraduate Education, Cancer Center, 02-781 Warsaw, Poland1; Departments of Medicine2 and Microbiology and Immunology,3 Vanderbilt University School of Medicine, Nashville, Tennessee 37232; and VA Medical Center, Nashville, Tennessee 372124

Received 4 October 2005/Returned for modification 16 November 2005/Accepted 7 February 2006

Helicobacter pylori babA encodes an outer membrane protein that binds to fucosylated Lewis b blood group antigen. We analyzed a panel of 35 H. pylori strains and identified three possible chromosomal loci for babA. There was a significant association between the presence of babA and the presence of cagA (P = 0.0001). Phylogenetic analysis of babA alleles revealed two divergent families of signal sequences. Among 17 strains in which an intact in-frame babA allele was identified, 10 expressed a detectable BabA protein. Expression of a BabA protein and the Lewis b-binding phenotype were not dependent on the chromosomal locus of babA. These data indicate that there is marked heterogeneity among H. pylori strains in babA genetic content and BabA expression.

Helicobacter pylori is a gram-negative bacterium that colonizes the human stomach. H. pylori genomes contain about 30 related hop genes, which are predicted to encode outer membrane proteins (1). Several Hop proteins, including BabA, SabA, AlpA, AlpB, and HopZ, can mediate adherence of H. pylori to gastric epithelial cells. The BabA adhesin binds to the fucosylated Lewis b blood group antigen on the surfaces of these cells (3, 8).

Some H. pylori strains bind Lewis b antigen, whereas other strains do not (3, 5, 8). Differences among H. pylori strains in the Lewis b binding phenotype can potentially be explained by the observation that only some of them express a BabA protein (7, 15). Peptic ulcer disease and gastric adenocarcinoma occur more commonly in persons infected with BabA-expressing H. pylori strains than in persons infected with strains that do not express BabA (5, 15). The molecular basis for variation among strains in expression of the BabA protein has not yet been investigated in any detail. In a previous study, we analyzed BabA expression by immunoblotting with two anti-BabA ScFv antibodies and detected expression of BabA in only about half of the H. pylori strains tested (7). In the current study, we studied the same panel of strains in order to analyze babA genetic content and investigate the molecular basis for variation among strains in BabA expression.

Genome sequence analyses of H. pylori strains J99 and 26695 indicated that babA can be present in one of two possible chromosomal loci (2, 14), which we designate locus A and locus B, respectively. Strain J99 expresses BabA and binds to Lewis b antigen, whereas strain 26695 does not express a detectable BabA protein and does not bind to Lewis b antigen (7, 8, 11). The mechanistic basis for the absence of BabA expression in strain 26695 is not known. As a first step in analyzing babA, we designed PCR primers that could amplify locus A and locus B (Table 1), and we analyzed a panel of 35 previously described H. pylori strains (7). PCR products of the expected sizes were successfully amplified from locus A (>3.5 kb) and from locus B (>3 kb) in 32 and 33 of the H. pylori strains tested, respectively (Table 2). These amplicons were analyzed initially by nucleotide sequencing using primers A1-S, A3-AS, and HopZ, from locus B (exemplified by HP0317 in strain 26695) were de-
of babA from these strains, using primers OPE1243, OP9647, OPE1244, and OPE1245 in all four possible combinations (Table 1). These primers were designed, based on aligned sequences from many different H. pylori strains, to amplify babA but not related H. pylori hop gene sequences. PCR products of the expected size were amplified from 2 strains (87-91 and 87-203) but not from the remaining 11 strains. Sequence analysis confirmed that these two amplicons were fragments of babA. We hypothesize that there is an unidentified chromosomal locus for babA in these two strains. Thus, among the 35 strains analyzed in the current study, 24 contained babA sequences and 11 did not. Our conclusion that some strains lack babA (P = 0.0001) (Table 3).

To gain further insight into babA genetic diversity, we analyzed the complete babA nucleotide sequences from 18 different H. pylori strains, including 3 strains (87-199, 60190, and J195) that contained two copies of babA (Table 3). This was accomplished by sequencing the PCR products described above on both strands with appropriate primers. In strain 87-199 (one of the strains that contained multiple copies of babA), the babA sequences found in loci B and C were identical and encoded full-length BabA proteins. In strain 60190, the babA sequences from loci A and C were almost identical except for

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A90-S</td>
<td>5′GTAATTTTGTGATCTTGTGTTGGA</td>
<td>jhp0832 (HP0893); locus A</td>
</tr>
<tr>
<td>A90-AS</td>
<td>5′GCAGTTGACATGTCAGCTTGGAG</td>
<td>jhp0835 (HP0898); locus A</td>
</tr>
<tr>
<td>AHP-S</td>
<td>5′CCACAGCATGTACATAGATATAGGAG</td>
<td>jhp1163 (HP1242); locus B</td>
</tr>
<tr>
<td>AHP-S</td>
<td>5′GGCTTTAAATCCCTACATTGGGA</td>
<td>jhp1165 (HP1244); locus B</td>
</tr>
<tr>
<td>H1P1-S</td>
<td>5′GAAGCTTGTGGTGGTGTGGTGA</td>
<td>jhp0298 (HP0313); locus C</td>
</tr>
<tr>
<td>H1P1-AS</td>
<td>5′ACCCTAATGGGATCGATTTGGA</td>
<td>jhp0301 (HP0318); locus C</td>
</tr>
<tr>
<td>OPE1243</td>
<td>5′ACCATCCTCAACAAAGGAGCAGGAG</td>
<td>jhp0833 (babA), bp 418–441 (forward)</td>
</tr>
<tr>
<td>OPE9647</td>
<td>5′GTTATGTCATCCACTTCCATCA</td>
<td>jhp0833 (babA), bp 440–461 (forward)</td>
</tr>
<tr>
<td>OPE1244</td>
<td>5′AGCCTCTGGCGCTCCGTGATC</td>
<td>jhp0833 (babA), bp 1002–1022 (reverse)</td>
</tr>
<tr>
<td>OPE1245</td>
<td>5′GTGAAAGGGTTGAAAGGCTTgc</td>
<td>jhp0833 (babA), bp 1085–1106 (reverse)</td>
</tr>
<tr>
<td>A1-S</td>
<td>5′CCCCGGGAGAGGCGGCTTTACACAb</td>
<td>jhp0833 (babA), bp 63–81 (forward)</td>
</tr>
<tr>
<td>A3-AS</td>
<td>5′TCGAGGAAGTGCAGTTTACACATTTca</td>
<td>jhp0833 (babA), bp 1380–1362 (reverse)</td>
</tr>
<tr>
<td>A3b-AS</td>
<td>5′GCTGTATCTGCTGGCTCTGGAGTC</td>
<td>jhp0833 (babA), bp 1414–1390 (reverse)</td>
</tr>
</tbody>
</table>

a Primers were derived from the indicated H. pylori chromosomal loci in H. pylori strain 26695 or J99. Locus A corresponds to the site of babA in strain J99 and the site of babB (HP0896) in strain 26695. Locus B corresponds to the site of babA (HP1243) in strain 26695 and the site of babB (HP1164) in strain J99. Locus C corresponds to the site of omp9 (HP0317) in strain 26695.

b Underlined sequences represent restriction sites for SmaI and XhoI.

deduced lengths (in amino acids) of the encoded BabA proteins. Five previously determined babA sequences correspond to GenBank accession numbers AY549174 to AY549178 (7), and new babA sequences have been assigned GenBank accession numbers DQ225153 to DQ225165. ND, not determined; NA, not applicable.

TABLE 1. Oligonucleotide primers used for analysis of H. pylori babA

<table>
<thead>
<tr>
<th>Locus</th>
<th>No. of strains</th>
<th>babA</th>
<th>babB</th>
<th>omp9</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>32</td>
<td>19</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>33</td>
<td>4</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

a Chromosomal loci A, B, and C are defined as described in the text.

b Total number of strains containing a detectable hop gene in locus A, B, or C.

c Presence or absence of cagA nucleotide sequences (7).

d Immunoreactivity with anti-BabA ScFv antibodies (7).
three substitutions near the 5' ends of the genes; both copies of \( \text{babA} \) in this strain encoded full-length BabA proteins. In strain J195, the \( \text{babA} \) sequences found in loci A and B were identical, and each contained a frameshift mutation that prevented expression of a full-length BabA protein. The multiple copies of \( \text{babA} \) in these strains presumably resulted from gene conversion (intragenomic nonreciprocal recombination) events (4, 10, 13). With the exception of the two copies of \( \text{babA} \) in strain J195, all of the \( \text{babA} \) alleles analyzed were predicted to encode full-length BabA proteins, ranging from 721 to 745 amino acids (Table 3).

Analysis of a consensus alignment of the 5' portion of \( \text{babA} \) alleles revealed two divergent families of sequences (Fig. 1). Four sequences (from strains J116, 86-338, J166, and 92-25) were markedly different from those in the other 18 strains. Shading indicates nucleotides different from the consensus.

To analyze the conservation and diversity of BabA amino acid sequences, we examined a consensus alignment of deduced BabA sequences. Consistent with the nucleotide sequence data shown in Fig. 1, two divergent families of amino acid sequences were detected near the amino terminus of BabA (Fig. 2). The region of divergence (amino acids 4 to 19) is predicted to constitute part of an amino-terminal signal peptide. The signal peptides encoded by four strains (J116, 86-338, J166, and 92-25-B) were closely related to signal peptides of paralogous BabB proteins (exemplified by HP0896 in strain 26695 and jhp1164 in strain J99) (Fig. 2). Notably, other parts of these four proteins were most closely related to BabA sequences (Fig. 2). It is striking that all four proteins had BabB-type signal peptides but sequences further downstream were typical BabA sequences.

We next sought to investigate potential links between \( \text{babA} \) genetic features, BabA protein expression, and binding to...
FIG. 2. Divergence among BabA signal peptides. Deduced N-terminal amino acid sequences of BabA from 21 \textit{H. pylori} strains (including 3 strains with multiple copies of \textit{babA}) were aligned. Deduced amino acid sequences of BabB (from strains J99 and 26695) and Omp9 (from strain 26695) are included in the alignment. Putative signal peptides are shaded and boxed. Two distinct families of BabA signal peptides can be distinguished. The signal peptides encoded by four strains are identical to BabB signal peptides. Black highlighting indicates amino acids different from the consensus.
As expected, all seven strains that bound to Lewis b antigen (Table 3). Nearly all of these 16 strains contained babA in locus A, but I contained babA in locus C and I contained two copies of babA in loci B and C. Among 17 strains in which an intact in-frame babA allele (predicted to encode a BabA protein of >700 amino acids) was identified, only 10 expressed a detectable BabA protein (Table 3). Potentially, BabA expression was not detected in some strains due to BabA amino sequence diversity and failure of anti-BabA antibodies to recognize certain epitopes. However, in a phylogenetic analysis, the babA alleles from BabA-expressing strains did not cluster separately from strains that failed to express BabA (data not shown).

We then tested all 35 strains for the ability to bind Lewis b antigen, using an in vitro adhesion assay (9). Among strains that expressed BabA, there was considerable variation in the level of binding to Lewis b antigen (Table 3). In previous studies, Lewis b/control binding ratios of >1.5 have been considered indicative of binding to Lewis b antigen (5, 9, 13). Based on this criterion, 7 strains exhibited detectable binding to Lewis b antigen and the other 28 strains did not (Table 3). As expected, all seven strains that bound to Lewis b antigen expressed a detectable BabA protein. Five of these seven strains contained babA in locus A. Among the four strains with divergent sequences near the 5’ end of babA (J166, 86-338, J116, and 92-25) (Fig. 1 and 2), two strains produced a detectable BabA protein and two did not. Strain J116 expressed a detectable BabA protein but did not exhibit detectable binding to Lewis b antigen. Strain 92-25 expressed a BabA protein and bound to Lewis b antigen, but this strain contained two non-identical copies of babA. Therefore, it is not possible to reach any conclusions about the functional role of the divergent copy of babA in this strain.

It has been reported previously that the babAI gene in strain CCUG17875 lacks a translation initiation codon (8). To determine whether the lack of BabA expression in some strains was due to this phenomenon, we analyzed the 5’ end of babA (J166, 86-338, J116, and 92-25) (Fig. 1 and 2). Two strains produced a detectable BabA protein and two did not. Strain J116 expressed a detectable BabA protein but did not exhibit detectable binding to Lewis b antigen. Strain 92-25 expressed a BabA protein and bound to Lewis b antigen, but this strain contained two non-identical copies of babA. Therefore, it is not possible to reach any conclusions about the functional role of the divergent copy of babA in this strain.

There are multiple possible chromosomal loci for babA. Individual strains can contain anywhere from 0 to 2 copies of babA. Frameshift mutations can result in the absence of BabA expression. Chimeric babB/babA alleles are commonly detected, and polymorphisms are present throughout babA alleles. In addition to this extensive variation among strains in babA content, a recent study showed that there can be metastability in babA sequences within the same strain (4). Based on the observation that many strains do not contain babA and do not bind to Lewis b antigen, it may be presumed that BabA expression is not required for persistent H. pylori colonization of the human stomach. However, adherence of H. pylori to gastric epithelial cells via the BabA adhesin may be a factor that contributes to the development of gastrodudodenal disease (5, 6).

**Nucleotide sequence accession numbers.** The babA sequences newly determined in this study have been assigned GenBank accession numbers DQ225153 to DQ225165.

This work was supported by NIH ROI DK53623, the Medical Research Service of the Department of Veterans Affairs, and CMKP grant 501-1-109-16/05, Poland.

**REFERENCES**


Editor: J. B. Bliska