An in vivo expression technology screen for Vibrio cholerae genes expressed in human volunteers

Mary-Jane Lombardo*, Jane Michalski*, Hector Martinez-Wilson1, Cara Morin*, Tamara Hilton*, Carlos G. Osorio1, James P. Nataro*, Carol O. Tacket*, Andrew Camilli1, and James B. Kaper‡§

*Center for Vaccine Development and Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201; †Howard Hughes Medical Institute and Department of Molecular Biology and Microbiology, Tufts University Medical School, Boston, MA 02111; and ¶Programa de Microbiología y Micología, Facultad de Medicina, Universidad de Chile, Santiago, Chile

Edited by Roy Curtiss III, Arizona State University, Tempe, AZ, and approved October 1, 2007 (received for review June 15 2007)

In vivo expression technology (IVET) has been widely used to study gene expression of human bacterial pathogens in animal models, but has heretofore not been used in humans to our knowledge. As part of ongoing efforts to understand Vibrio cholerae pathogenesis and develop improved V. cholerae vaccines, we have performed an IVET screen in humans for genes that are preferentially expressed by V. cholerae during infection. A library of 8,734 nontoxicogenic V. cholerae strains carrying transcriptional fusions of genomic DNA to a resolvase gene was ingested by five healthy adult volunteers. Transcription of the fusion leads to resolvase-dependent excision of a sacB-containing cassette and thus the selectable phenotype of sucrose resistance (SucR). A total of ~20,000 SucR isolates, those carrying putative in vivo-induced fusions, were recovered from volunteer stool samples. Analysis of the fusion junctions from >7,000 SucR isolates from multiple samples from multiple volunteers identified 217 candidate genes for preferential expression during human infection. Of genes or operons induced in three or more volunteers, the majority of those tested (65%) were induced in an infant mouse model. VC0201 (fhuC), which encodes the ATPase of a ferrichrome ABC transporter, is one of the identified in vivo-induced genes and is required for virulence in the mouse model.

gene expression | genetics | vaccinology | virulence

Our understanding of the complex interactions between bacterial pathogens and humans relies heavily on the use of animal and tissue culture model systems that serve as surrogates of human infection. One useful genetic tool for discovering genes and pathways involved in virulence is in vivo expression technology (IVET), which was designed to identify genes of pathogens that are preferentially expressed during infection and have been extensively used in model systems (reviewed by refs. 1 and 2). IVET is a promoter-trapping strategy in which cells carrying a library of transcriptional fusions of genomic DNA to a reporter gene are used to infect a model host and those carrying fusions that are expressed in vivo can be identified by either a genetic screen or selection. This technique allows the identification of genes that may be expressed only under in vivo conditions [in vivo-induced genes (ivi genes)]. Such genes may be very difficult to identify during growth under laboratory conditions, but are likely to be important in the host for survival and virulence. One form of IVET, used in this work, is recombination-based IVET (3, 4) in which the reporter gene encodes a resolvase that effects a permanent genetic change, allowing a direct selection for cells that expressed the resolvase even transiently during infection.

Here, we describe the use of recombination-based IVET to identify genes of Vibrio cholerae, the causative agent of the diarrheal disease cholera, that are expressed during human infection. This approach has the potential to identify important virulence genes that may not be expressed in vitro or during infection of an animal host. This experiment was possible because of the well described attenuated nature of V. cholerae strain CVD110, which lacks cholera toxin, and can be safely ingested by volunteers. However, it is still capable of causing infection (a mild version of that caused by ctx+ V. cholerae) as manifested by shedding of high numbers of V. cholerae in diarrheal stools and immunological responses typical of cholera (5). This reactogenicity (diarrhea and other symptoms) makes CVD110 unsuitable for vaccine purposes, but excellent for our goal of identifying genes induced in vivo in a human infection. This same IVET library was used previously in an infant mouse model (3), allowing comparison of the ivi gene sets. In this human IVET study we identified 217 putative human ivi genes of V. cholerae. We show that some of these are mouse ivi genes, whereas others may represent the interesting class of genes that are specifically induced in human infection.

Results

The V. cholerae IVET Library. The recombination-based IVET system used in this work and the construction of this exact library have been described in detail (3). The library consists of clones that contain a chromosomal transcriptional fusion to a promoterless resolvase gene (tpnR) and an unlinked neo-sacB cassette flanked by recognition sequences (res) for the resolvase [supporting information (SI) Fig. 1]. Transcription of the fusion will lead to expression of resolvase, which will recognize the res sites and excise the cassette, making the cells resistant to sucrose (SucR). Thus, selection for SucR cells from this library provides a positive selection for promoters that were active, even transiently, under the growth conditions tested. As described (3), 11,000 clones were prepared for the library, but those with good in vitro expression during growth in LB (leading to a high frequency of SucR cells) were excluded, leaving a library of 8,734 clones. This step, which enriches the library for genes that are poorly expressed in vitro, was intended to improve the recovery of genes that are induced specifically in vivo during infection.

Certain complexities of the library proved useful in this human study. First, the library was designed to have expanded promoter detection capabilities by using three vectors, each having a different tpnR ribosome binding site (RBS). This design allows for the possibility of detecting very efficient in vivo promoters that might otherwise have been excluded from the library because of high in vitro expression, because resolvase production would be reduced by a poor RBS. It also allows for more sensitive detection of weak in vivo promoters because resolvase expression

Author contributions: M.-J.L., J.M., J.P.N., C.O.T., A.C., and J.B.K. designed research; M.-J.L., J.M., H.M.-W., C.M., T.H., and C.O.T. performed research; C.G.O. contributed new reagents/analytic tools; M.-J.L., A.C., and J.B.K. analyzed data; and M.-J.L. wrote the paper.

The authors declare no conflict of interest.

This article is PNAS Direct Submission.

Abbreviations: IVET, in vivo expression technology; ivi gene, in vivo-induced gene; RBS, ribosome binding site; Amp, ampicillin; Rif, rifampicin.

‡To whom correspondence should be addressed. E-mail: jkaper@umaryland.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0705636104/DC1.

© 2007 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0705636104

PNAS | November 13, 2007 | vol. 104 | no. 46 | 18229–18234
would be strengthened by a good RBS. Second, two different genomic partial digests (and so two cloning sites) were used in constructing the library to increase coverage of the genome. The use of three vectors and two cloning sites means that any gene can potentially be represented in the library in six different vector contexts. Finally, the library was constructed in two strain backgrounds that differ in the sensitivity of cassette resolution. The neo-sacB cassette in one strain contains the WT res sites and in the other strain contains mutant res sites that are resolved less efficiently, so the frequency of resolution is also a variable that may aid in isolating a particular ivi gene fusion. Each of the above features, except the res site, can be determined by a single sequencing run across the fusion junction from within the tnpR gene (see Materials and Methods). Thus, we can categorize the fusions based not only on which gene is upstream of tnpR but also which RBS, genomic digest, and the exact junction sequence fused to tnpR (see SI Fig. 2). These features of the library are an advantage because genes or operons that are represented by multiple fusion types are more likely to be in vivo induced. To restate, each different fusion to a gene that is recovered from the library is unambiguously an independent isolation of that gene, whereas multiple isolates of the same fusion may be clonal derivatives of the same resolved SucR isolate.

Volunteer Study. Five adult volunteers ingested $5 \times 10^8$ cfu of the library described above (see Materials and Methods). All volunteers displayed typical immune responses to ctx V. cholerae (data not shown), and four shed SucR V. cholerae in their stools. We also attempted to isolate SucR cells from the duodenum, the site of colonization, by using gelatin string capsules (see Materials and Methods). However, only two strings yielded SucR cells, each from a different volunteer. SucR V. cholerae were isolated from stool and duodenum samples, and up to 672 individual SucR colonies from each sample were picked into microtiter well plates. More than 20,000 independent colonies were picked and stored at $-80^\circ$C for subsequent analysis.

Human IVET Fusions. To identify the putative ivi genes we sequenced the genomic DNA fused to tnpR from 8,665 SucR isolates from stool samples. Previously, sequencing IVET fusion junctions was a multistep process including plasmid preparations and transformations that limited the number of samples that could be sequenced (e.g., ref. 3). To circumvent this limitation we developed a high-throughput PCR-based approach that should be applicable to future IVET studies (see Materials and Methods). We sequenced fusion junctions from three stool samples from each of three volunteers and six samples from the fourth volunteer whose samples were more heterogeneous. We chose stool samples from early, midway, and late in infection. A total of 8,665 isolates were sequenced from stool and compared with the sequenced genomes of V. cholerae strain N16961 (6) and the bacteriophage K139 (7). About 84% of sequences for a given stool sample were of high quality, yielding 7,063 BLAST hits. Of these, 6,025 (85%) were fused in the correct orientation to tnpR for the identified promoter/transcriptional unit to be driving tnpR expression.

Analysis of the 6,025 fusion sequences revealed that they corresponded to a total of 220 genes: 217 V. cholerae genes and 3 phage K139 genes (SI Table 3). The 217 V. cholerae genes were assigned to 176 putative transcriptional units (data not shown) based on gene orientation and gene order conservation (www.tigr.org/tigr-scripts/operons/operons.cgi) and predicted terminators (6). Note that recovering a fusion to any gene in an operon identifies all genes in that operon as potentially in vivo induced during human infection. Genes or operons to which multiple unique fusions were isolated are good candidates for being in vivo induced. For example, see SI Table 4 in which multiple fusions to selected transcriptional units are indicated, both as fusions to different genes in the operon and unique fusions to each gene.

Infant Mouse Studies. Infant mice provide a useful model of V. cholerae colonization, even though this model does not reproduce all aspects of disease (reviewed in ref. 8). There is a very good correlation between the genes identified in this human study and those identified in the screen of this same IVET library in the infant mouse model (3). Of the 40 genes that were confirmed to be in vivo induced in the mouse in that study, 83% were identified in this human screen by fusions to those genes or to the putative operons containing them.

We examined expression of a number of other genes identified in the human study in the mouse model. We reasoned that genes that were isolated from multiple human volunteers were more likely to be true ivi genes than those isolated from a single volunteer, and we chose to work first with genes or transcriptional units for which fusions were isolated from at least three volunteers. The class of genes isolated from at least three volunteers consists of 79 genes in 47 putative transcriptional units (listed in SI Table 4). We examined expression of fusions to genes in this class in vitro (LB medium) and in vivo in the mouse model. The entire data set is presented in SI Table 4, and a condensed set with only those genes that are expressed at higher level in vivo than in vitro is presented in Table 1. We find that of the 40 of these 47 putative transcriptional units tested, 26 (65%) are induced in the mouse relative to in vitro growth (Table 1). These include the arg, thr, ilv, thi, and trp biosynthetic operons, genes encoding chitobiase (VC0611–VC0620) and ferrichrome (VC0200–VC0203) transporters, and VC0130, which encodes a GGDEF family protein. Of genes identified from four volunteers, 14 of the 16 tested (88%) are in vivo induced in the mouse.

As mentioned above, another class of genes/operons that we reasoned might be likely to be human ivi genes is those that were represented by more than one type of tnpR fusion in the set of 6,025 fusions. For example, four different genes in the thi operon (VC0063–VC0066) were isolated as fusions (see SI Table 4). Different fusions to some of those genes were isolated as well (see SI Fig. 2 for examples). We found that 25 of the 176 transcriptional units identified in this study were represented by three or more unique fusions to tnpR (data not shown). Fourteen of the 20 of these tested (70%) are in vivo induced in the mouse. Thus, both of these classes of genes from the human IVET screen consist largely of genes that are in vivo induced in the mouse. The others are good candidates for genes that might be expressed in humans but not in mice.

We constructed nonpolar null mutations (data not shown) in several ivi genes confirmed to be in vivo induced in the mouse (from SI Table 4) to assess their effects on virulence in competition experiments in the mouse (Table 2). We also tested the effects of null mutations in VC1349 and VC0303, which encode response regulators of unknown function, even though neither is induced in the mouse model (SI Table 4 and data not shown). Deletion of either VC1349 or VC0303 caused attenuation in vivo. Deletion of VC0616–VC0617 (chitobiase transport), VC1275 (hypothetical), VC0304 (conserved hypothetical), or VC1687 (inorganic pyrophosphatase) also caused no attenuation in vivo. Deletion of VC0201 (fluC), the ATPase for the flu ferrichrome ABC transporter (9), caused no growth defect in vitro but led to a 22-fold attenuation in the mouse. Null mutations affecting VC0488, VC0130, VC2705, and VC2646 were previously tested (3) with only VC2705 showing strong attenuation (14-fold) in the mouse. Mutations affecting the ilv operon (VC0027–VC0031) and the trp operon (VC1169–VC1174) were previously shown to be associated with decreased virulence in the mouse (10).
Table 1. Putative human ivi genes/operons that are preferentially induced in the mouse small intestine

<table>
<thead>
<tr>
<th>Gene</th>
<th>Putative operon*</th>
<th>No. genes in operon isolated as fusions</th>
<th>Description</th>
<th>% Suc&lt;sup&gt;i&lt;/sup&gt;&lt;sub&gt;v&lt;/sub&gt; in vitro&lt;sup&gt;n&lt;/sup&gt; (LB)</th>
<th>% Suc&lt;sup&gt;i&lt;/sup&gt;&lt;sub&gt;v&lt;/sub&gt; in vivo&lt;sup&gt;n&lt;/sup&gt; (mouse)</th>
<th>In vivo/ in vitro&lt;sup&gt;n&lt;/sup&gt; ratio</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VC0027</td>
<td>VC0027–31</td>
<td>3</td>
<td>Threonine dehydrogase (ilvA)</td>
<td>2.5</td>
<td>49</td>
<td>19</td>
<td>0.005</td>
</tr>
<tr>
<td>VC0063</td>
<td>VC0060–66</td>
<td>4</td>
<td>Adenylyltransferase (thif)</td>
<td>&lt;0.1&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>77</td>
<td>220&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC0130</td>
<td>VC0125–130</td>
<td>2</td>
<td>GGDEF family protein</td>
<td>0.2&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>2&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>10&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>0.1&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC0201</td>
<td>VC0200–203</td>
<td>3</td>
<td>Ferrichrome ABC transporter, ATP-binding protein (fhuC)</td>
<td>0.6&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>17</td>
<td>28&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC0300</td>
<td>VC0291–301</td>
<td>1</td>
<td>Conserved hypothetical protein</td>
<td>&lt;0.08</td>
<td>8.9</td>
<td>120</td>
<td>0.2</td>
</tr>
<tr>
<td>VC0488</td>
<td></td>
<td></td>
<td>Extracellular solute-binding protein, putative</td>
<td>&lt;0.1&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>22</td>
<td>220&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC0613</td>
<td>VC0611–620</td>
<td>8</td>
<td>β-N-acetylhexosaminidase (chb-1)</td>
<td>0.35</td>
<td>2.6</td>
<td>7.4</td>
<td>0.007</td>
</tr>
<tr>
<td>VC0614</td>
<td>VC0611–620</td>
<td>8</td>
<td>Glucosamine-specific kinase</td>
<td>3.9</td>
<td>38</td>
<td>9.6</td>
<td>0.0006</td>
</tr>
<tr>
<td>VC0616</td>
<td>VC0611–620</td>
<td>8</td>
<td>(GlicNAC&lt;sub&gt;2&lt;/sub&gt;) ABC transporter, ATP-binding protein</td>
<td>0.69</td>
<td>31</td>
<td>45</td>
<td>0.0008</td>
</tr>
<tr>
<td>VC1034</td>
<td></td>
<td></td>
<td>Uridine phosphorylase (udp-1)</td>
<td>75&lt;sup&gt;n&lt;/sup&gt;</td>
<td>100</td>
<td>1.3</td>
<td>&lt;0.05&lt;sup&gt;n&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC1072</td>
<td></td>
<td></td>
<td>Hypothetical protein</td>
<td>0.37</td>
<td>3.9</td>
<td>11</td>
<td>0.002</td>
</tr>
<tr>
<td>VC1173</td>
<td>VC1169–74</td>
<td>2</td>
<td>Anthranilate synthase component II (trpG)</td>
<td>1&lt;sup&gt;†&lt;/sup&gt;</td>
<td>16&lt;sup&gt;†&lt;/sup&gt;</td>
<td>16&lt;sup&gt;†&lt;/sup&gt;</td>
<td>0.2&lt;sup&gt;†&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC1275</td>
<td>VC1274–75</td>
<td>1</td>
<td>Conserved hypothetical protein</td>
<td>8&lt;sup&gt;§&lt;/sup&gt;</td>
<td>85&lt;sup&gt;§&lt;/sup&gt;</td>
<td>11&lt;sup&gt;§&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC1658</td>
<td></td>
<td></td>
<td>Serine transporter (sdA-C-2)</td>
<td>3&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>88&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>30&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC1687</td>
<td>VC1686–87</td>
<td>1</td>
<td>Mn-dependent inorganic pyrophosphatase</td>
<td>5&lt;sup&gt;§&lt;/sup&gt;</td>
<td>18&lt;sup&gt;§&lt;/sup&gt;</td>
<td>3.6&lt;sup&gt;§&lt;/sup&gt;</td>
<td>0.07&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC2363</td>
<td>VC2362–64</td>
<td>3</td>
<td>Homoserine kinase (thrB)</td>
<td>4.1</td>
<td>54</td>
<td>13</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>VC2373</td>
<td>VC2373–74</td>
<td>1</td>
<td>Glutamate synthase, large subunit (glb-1)</td>
<td>0.1&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>3&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>30&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>0.1&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC2411</td>
<td></td>
<td></td>
<td>Hypothetical protein</td>
<td>0.21</td>
<td>1.4</td>
<td>6.9</td>
<td>0.02</td>
</tr>
<tr>
<td>VC2419</td>
<td>VC2417–19</td>
<td>1</td>
<td>Integrase/recombinase XerD (serD)</td>
<td>5&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>34&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>6.8&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC2539</td>
<td>VC2537–39</td>
<td>2</td>
<td>Thiamine ABC transporter, periplasmic binding protein (tpbA)</td>
<td>3.4</td>
<td>54</td>
<td>16</td>
<td>0.01</td>
</tr>
<tr>
<td>VC2565</td>
<td>VC2565–66</td>
<td>1</td>
<td>ElaA protein</td>
<td>0.12</td>
<td>31</td>
<td>260</td>
<td>0.03</td>
</tr>
<tr>
<td>VC2641</td>
<td>VC2641–44</td>
<td>3</td>
<td>Argininosuccinate lyase (argH)</td>
<td>&lt;0.1&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>70&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>700&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC2646</td>
<td></td>
<td></td>
<td>Phosphoenolpyruvate carboxylase (ppc)</td>
<td>8&lt;sup&gt;§&lt;/sup&gt;</td>
<td>18&lt;sup&gt;§&lt;/sup&gt;</td>
<td>2.2&lt;sup&gt;§&lt;/sup&gt;</td>
<td>0.07&lt;sup&gt;§&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC2705</td>
<td></td>
<td></td>
<td>Sodium/solute symporter, putative</td>
<td>3&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>12&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>4&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC2712</td>
<td></td>
<td></td>
<td>Xanthine/uracil permease family protein</td>
<td>1.1</td>
<td>46</td>
<td>43</td>
<td>0.007</td>
</tr>
<tr>
<td>VC2742</td>
<td>VC2742–43</td>
<td>1</td>
<td>Ribonuclease BN (rbn)</td>
<td>1.4&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>5&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>3.6&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC0016</td>
<td>VC0013–16</td>
<td>4</td>
<td>1,4-α-glucan branching enzyme (gldB)</td>
<td>7&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>24&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>3.4&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>&lt;0.05&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC0242</td>
<td>VC0240–46</td>
<td>1</td>
<td>Hexulose-6-phosphate synthase SgbH, putative</td>
<td>2.5&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>8.4&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>3.4&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>0.1&lt;sup&gt;‡&lt;/sup&gt;</td>
</tr>
<tr>
<td>VC0629</td>
<td>VC0629–33</td>
<td>2</td>
<td>Conserved hypothetical protein</td>
<td>0.47</td>
<td>10</td>
<td>21</td>
<td>0.006</td>
</tr>
<tr>
<td>VC0631</td>
<td>VC0629–33</td>
<td>2</td>
<td>Hypothetical protein</td>
<td>31</td>
<td>95</td>
<td>3.1</td>
<td>0.02</td>
</tr>
</tbody>
</table>

This list is a subset of the list in SI Table 4 and includes only genes/operons that were recovered from at least three volunteers. There are additional mouse ivi genes in the human set, some of which are discussed in the text.

*In most cases putative operons are represented only by a single fusion to one gene of that operon. However, if a putative operon contained genes of hypothetical or diverse function, fusions to more than one gene in that operon may have been tested (i.e., VCA0629–33).

†<sup>n</sup>n = 3 for in vitro and 4 for in vivo except for VC2539 where n = 2 for in vivo data.

‡<sup>n</sup>These data are from Osorio et al. (3).

Discussion

The IVET genetic approach has been extensively used in animal models to study microbial pathogenesis. This human IVET study allows comparison of the effectiveness of IVET in identifying virulence genes in humans and in an animal model system. We found a very good correlation between results of this human screen and one using the same IVET library in an infant mouse model, validating the utility of IVET to study human pathogens in model systems. The screen in humans identified a large set of genes that may be important during human infection, including genes encoding hypothetical and conserved hypothetical proteins, regulatory proteins, potential surface expressed proteins, and chemotaxis and motility functions. This human IVET screen also demonstrates the feasibility of using the IVET approach in humans with other attenuated human pathogens.

The Human Putative ivi Gene Set. We identified 217 V. cholerae genes that are putatively induced during human infection (SI Tables 3 and 4). When categorized by functional roles (data not shown) the largest class of genes in the set (26%) encodes hypothetical and conserved hypothetical proteins, which is typical of most genomewide analyses, including IVET studies (1). Fifteen of these hypotheticals were isolated from at least three volunteers and therefore seem more likely to be induced in vivo than genes isolated from only one infection. Six of the hypotheticals tested so far are preferentially expressed in the mouse

Lombardo et al. November 13, 2007 | vol. 104 | no. 46 | 18231

PNAS
relative to in vitro (SI Table 4 and Table 1). The next largest classes are of genes encoding transport and binding proteins and proteins involved in energy metabolism and amino acid biosynthesis. These classes also tend to be heavily represented in IVET screens in animal models, underscoring the key roles of metabolic processes in colonization and infection (1).

Examining the class of genes that was isolated from at least three volunteers revealed that the majority of these genes were preferentially expressed in the mouse relative to growth in rich medium in vitro (SI Table 4 and Table 1). Thus the screen successfully identified genes expressed in the animal model of cholera infection. Further, we examined virulence phenotypes for several of these genes in the infant mouse model. Deletion of VC0201 (fhuC), an iron-regulated gene that encodes an ABC transporter component for hydroxamate compounds like ferriochrome (9, 11), led to a 22-fold attenuation in the mouse (Table 2). This finding implicates hydroxamate compounds as an important iron source during infection, an important finding because the iron sources used by V. cholerae in the host are not clear (12). Deletion of VC2705, which encodes a sodium symporter, led to a 14-fold attenuation (3). Other mutations affecting genes in this class that have been tested so far (Table 2 and ref. 3) caused no defect in infection in the mouse. These genes may be important during human infection and not mouse infection or may represent a function that is encoded by more than one gene. For example, VC1687, which encodes an inorganic pyrophosphatase (13), is highly expressed in the mouse (and poorly expressed in rich medium in vitro) but its deletion has no effect on virulence. V. cholerae has two pyrophosphatases, so the other pyrophosphatase (encoded by VC2545) is apparently sufficient. Alternatively, ivi genes that do not appear to be important in vitro may encode functions that are being expressed to prepare for conditions encountered after being shed from the host (and so perhaps not have effects on virulence). For example, chemotaxis functions are differentially regulated as V. cholerae exit the host, leading to an increased ability to infect a subsequent host (14, 15).

**Human ivi Regulators.** The 14 regulators identified in this screen (SI Table 3) include the following known regulators: VieB (VC1651), a response regulator that is expressed in vivo in the mouse and is part of a signal transduction system that contributes to regulation of cholera toxin production (16, 17); VarS (VC2453), a sensor kinase involved in virulence and quorum sensing (18, 19); and PhoR (VC0720), a sensor kinase that controls expression of the PhoBR regulon in response to phosphate and is important for colonization of the rabbit small intestine (20). Deletion of either VC1349 or VC0303, two response regulators, had no effect on virulence in the mouse model (Table 2). As discussed above, absence of a phenotype in the mouse model may indicate redundancy, importance only during human infection, or importance in the environment after being shed from the host.

**Human ivi Promoters.** An important current approach in vaccine development is the use of attenuated vaccine vectors for expression of heterologous antigens. Promoters that are induced strongly in vivo, but poorly in vitro are good candidates for driving antigen expression (21). Several fusions in the human gene set have in vivo/in vitro induction ratios in the mouse model that are >100-fold including VC0063, VC0300, VC0488, VC2565, and VC2641 (Table 1) and the previously reported VCA1008 (3).

**Virulence Genes and Pathogenicity Islands.** Several genes known to be required for human virulence were not recovered in this study. For example, no fusions to genes in the V. cholerae pathogenicity island (VPI) were isolated. The VPI includes genes encoding the TCP pilus, a type IV pilus required for human infection, and other genes important for virulence (22). Also absent are fusions to other members of the toxR regulon. The absence of these genes is not surprising given that tcpA-tmpR fusions with two of the three RBS sites (the A and B RBS constructs; see SI Fig. 2) used in the IVET library yielded a high level of resolution in vitro after growth in LB medium (in a related IVET system; ref. 23). Thus fusions of the A and B RBS type to genes regulated like tcpA would have been excluded from the library. The number of C type fusions in the library (~2,500 strains) is insufficient to expect good coverage of the genome in the C type construct, and indeed we were unable to detect fusions to tcpA in the inoculum by PCR (data not shown). There were also no fusions to genes in the VPI-2 island (24) and none to genes in the VSP-I genomic island (25). The VSP-I and VSP-II genomic islands are specific to seventh pandemic El Tor strains of V. cholerae (25). There was a single fusion in VSP-II, to VC0506, which encodes a hypothetical protein (26).

**Antisense Genes or Cryptic Genes/Promoters.** Fifteen percent of fusions isolated were in the incorrect orientation for the fused gene to drive expression of tmpR. Some are probably transcribed from a nearby convergent promoter; however, in some cases there is no candidate promoter. As discussed (3, 27, 28), these fusions may reflect cryptic protein-coding genes, or transcribed but nontranslated sequences such as antisense or small regulatory RNAs and may be worth further investigation given the increasing numbers of small RNAs with important regulatory functions (29).

**Comparison with Previous Screens.** Previous large-scale microarray and signature-tagged mutagenesis screens for V. cholerae virulence genes using various V. cholerae strains and model systems (10, 15, 30, 31) have yielded sets of genes with limited overlap with each other, highlighting the importance of using multiple approaches and model systems to examine in vivo gene expression and pathogenesis. This human IVET data set also has limited overlap with those screens (with the exception of the screen of this same IVET library in the mouse, discussed above), and thus provides, as expected, a unique view of in vivo gene expression. At least in part, this unique set is caused by the ability of IVET to identify promoters induced at any point, even transiently, in the infection, whereas microarray analysis, in contrast, compares averaged gene expression levels in a population of cells at a particular time. The IVET system also has the advantage that individual IVET fusion strains can be introduced into an animal model and the promoter activity can be monitored in both space and time during infection to gain insight into where and when that promoter is induced (e.g., ref. 23).

In conclusion, our findings support the widely held view that IVET experiments in model systems, and in the infant mouse model of V. cholerae in particular, are useful surrogates of human infection. We have identified a large number of candidate human ivi genes of V. cholerae, some of which are preferentially expressed during infection of mice. Analysis of only a small number of mutants from this data set has revealed several strongly attenuating mutations including fhuC and VC2705 (3). The data from this human study will fuel both studies of basic virulence mechanisms and vaccine development.

**Materials and Methods**

**Bacterial Strains and Growth Conditions/Inoculum.** The V. cholerae strains used in the IVET study are derivatives of GOA1264, a rifampicin-resistant mutant of CVD110, a ctc− derivative of the E7946 El Tor Ogawa strain (32). Deletion mutants for mouse studies were constructed in E7946. Unless stated otherwise, additives were used at the following concentrations: 50 μg/ml ampicillin (Amp), 25 μg/ml kanamycin, 10 μg/ml rifampicin.
obtained from all volunteers after appropriate counseling. Five volunteers ingested a single dose of \( \times 10^8 \) cells of the inoculum in sodium bicarbonate buffer and were observed for 9 days as inpatients. On day 4 all volunteers received ciprofloxacin (500 mg every 12 h) for 5 days. All stools were graded for the severity of diarrhea and were either diluted and plated to TCBS (Vibrio selective medium), LB Rif (100 \( \mu \)g/ml), and LB Rif (100 \( \mu \)g/ml) sucrose within several hours of passage or stored at 4°C before plating. The LB Rif plates provided a total \( V. \) cholerae count that was more accurate than the count provided by plating to TCBS medium. A total of 672 well isolated Suc\(^R\) colonies were picked from each of multiple stools of each volunteer into microtiter well plates containing LB Rif (100 \( \mu \)g/ml) and Amp (100 \( \mu \)g/ml) and grown overnight at 37°C. DMSO was added to 4.8%, and plates were then stored at -80°C. Each volunteer also swallowed a gelatin string capsule (Enterotest) on day 1 and/or 2 after inoculation to sample the \( V. \) cholerae population in the duodenum (33). The distal 15 cm of the recovered string was placed in sterile saline and plated as described above for stool samples.

Identification of Putative Human ivi Genes. An efficient and cost-effective method for identifying the genes fused to \( mnpR \) was developed based on the low-frequency (~1/1,000) recombination excision of the integrated \( mnpR \) fusion from the chromosome into a nonreplicating plasmid form. Primers flanking the insertion site in the library vectors (5‘-ACGTCACCCTCCTC-CACTTTTAC and 5‘-AAATTGTACGCCAGTAAAT- GTG) were used to amplify the inserts from these plasmid forms by using boiled cell suspensions as the template source and Easy-A high fidelity polymerase (Stratagene, La Jolla, CA) in a 96-well plate format. Aliquots of the PCR mixtures were then used directly as template in sequencing reactions with a primer (5‘-CTCTTATCCTCAAGCA GTCA) that reads out from the 5‘ end of the \( mnpR \) coding region, also in a 96-well plate format. Sequences were analyzed in batches of 96 by using a tailored local BLASTN program. The program first performs a BLASTN (34) search against the \( V. \) cholerae N16961 annotated ORF database (TIGR). If no hit is found, BLASTN is performed against the two chromosomes of \( V. \) cholerae N16961 [TIGR accession nos. NC 002505 and NC 002506 (6)] and also against the bacteriophage K139 genome (7) (TIGR accession no. NC 002506). If no hit is found, BLASTN is performed against the set of three independent LB Amp cultures to both LB Rif and LB Rif sucrose and dividing the LB Rif titer by the LB Rif sucrose titer. In general, fusion strains with \( \leq 10\% \) resolution in vitro were examined for in vivo resolution frequencies after infection of infant CD-1 mice, although several with higher resolution were also examined. One culture of each fusion strain was grown overnight in LB Amp with kanamycin (Kan) added to prevent contamination of the inoculum with Suc\(^R\) Kan\(^R\) resolved cells (so that resolution events occurring post inoculation would predominate). Five- to 6-day-old infant mice (four to five per fusion strain) were inoculated intragastrically with \( \leq 10^6 \) cfu of the diluted culture in 50 \( \mu \)l. After 20-24 h bacteria were recovered from small intestine homogenates, and dilutions were plated to LB Rif and LB Rif sucrose to allow calculation of Suc\(^R\) colonies/total colonies as above. The frequency of resolution in the inoculum was also determined so that rare cultures with high numbers of Suc\(^R\) cells would be detected. In vivo expression is expressed as the average percentage of Suc\(^R\) cells in the four small intestine homogenates. Thus mouse in vivo-induced genes are those in which the fusion strain displayed a higher average percentage of Suc\(^R\) cells in the small intestine homogenates than the average percentage from three in vitro-grown cultures.

Competition experiments were performed with nonpolar deletion mutant derivatives of E7946 Sm\(^R\) constructed (data not shown) by using a standard suicide vector approach (37). Mutants were competed against a lacZ\(^-\) derivative of E7946 Sm\(^R\) (carrying lacZ\(z\):res-let-res) either in vitro (LB medium) or in vivo (in the infant mouse) as described (3). Mutant and WT cells were distinguished by plating to LB medium with streptomycin and X-gal. Competitive indices were calculated as described (3).

We thank M. G. Pohl (Bioinformatics Core, University of Maryland School of Medicine) for developing the tailored myblast program, L. Sadewicz (University of Maryland Biopolymer Laboratory) for high-throughput sequence analysis, and J. W. Yoon, N. Morin, and members of J.B.K.’s and J.P.N.’s laboratories for assistance with various aspects of the human and animal studies. This work was supported by National Institutes of Health Grants R01-AI19716 (to L.B.K.), R01-AI55058 (to A.C.), and M01-RR-16500 (to the General Clinical Research Center at the University of Maryland), National Institute of Allergy and Infectious Diseases Grant N01-AI 65299 (to C.O.T.), and National Institutes of Health Training Grants T32 AI07524 (to M.-J.L.) and T32 DK067872 (to C.M.).