

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

Mary-Jane Lombardo*, Jane Michalski*, Hector Martinez-Wilson[†], Cara Morin*, Tamara Hilton*, Carlos G. Osorio[‡], James P. Nataro*, Carol O. Tacket*, Andrew Camilli[†], 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 (Suc^R). A total of ≈20,000 Suc^R isolates, those carrying putative *in vivo*-induced fusions, were recovered from volunteer stool samples. Analysis of the fusion junctions from >7,000 Suc^R 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 has 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 (*tnpR*) 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 (Suc^R). Thus, selection for Suc^R 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 Suc^R 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 *tnpR* 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 a 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

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 Suc^R isolate.

Volunteer Study. Five adult volunteers ingested 5×10^8 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 Suc^R *V. cholerae* in their stools. We also attempted to isolate Suc^R cells from the duodenum, the site of colonization, by using gelatin string capsules (see *Materials and Methods*). However, only two strings yielded Suc^R cells, each from a different volunteer. Suc^R *V. cholerae* were isolated from stool and duodenum samples, and up to 672 individual Suc^R colonies from each sample were picked into microtiter well plates. More than 20,000 independent colonies were picked and stored at -80°C for subsequent analysis.

Human IVET Fusions. To identify the putative *ivi* genes we sequenced the genomic DNA fused to *tnpR* from 8,665 Suc^R 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 chitobiose (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 neither VC1349 nor VC0303 caused attenuation *in vivo*. Deletion of VC0616–VC0617 (chitobiose transport), VC1275 (hypothetical), VCA0629 (conserved hypothetical), or VC1687 (inorganic pyrophosphatase) also caused no attenuation *in vivo*. Deletion of VC0201 (*fhuC*), the ATPase for the *fhu* 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).

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 ferri-chrome (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 vivo* 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-tnpR* 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 ($\approx 2,500$ 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 *tnpR*. 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 *ctx*⁻ 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 $\mu\text{g/ml}$ ampicillin (Amp), 25 $\mu\text{g/ml}$ kanamycin, 10 $\mu\text{g/ml}$ rifampicin

9. Rogers MB, Sexton JA, DeCastro GJ, Calderwood SB (2000) *J Bacteriol* 182:2350–2353.
10. Merrell DS, Hava DL, Camilli A (2002) *Mol Microbiol* 43:1471–1491.
11. Mey AR, Wyckoff EE, Kanukurthy V, Fisher CR, Payne SM (2005) *Infect Immun* 73:8167–8178.
12. Wyckoff EE, Mey AR, Payne SM (2007) *Biometals* 20:405–416.
13. Salminen A, Ilias M, Belogurov GA, Baykov AA, Lahti R, Young T (2006) *Biochemistry (Mosc)* 71:978–982.
14. Butler SM, Nelson EJ, Chowdhury N, Faruque SM, Calderwood SB, Camilli A (2006) *Mol Microbiol* 60:417–426.
15. Merrell DS, Butler SM, Qadri F, Dolganov NA, Alam A, Cohen MB, Calderwood SB, Schoolnik GK, Camilli A (2002) *Nature* 417:642–645.
16. Lee SH, Angelichio MJ, Mekalanos JJ, Camilli A (1998) *J Bacteriol* 180:2298–2305.
17. Tischler AD, Lee SH, Camilli A (2002) *J Bacteriol* 184:4104–4113.
18. Wong SM, Carroll PA, Rahme LG, Ausubel FM, Calderwood SB (1998) *Infect Immun* 66:5854–5861.
19. Lenz DH, Miller MB, Zhu J, Kulkarni RV, Bassler BL (2005) *Mol Microbiol* 58:1186–1202.
20. von Kruger W, Humphreys S, Ketley J (1999) *Microbiology* 145:2463–2475.
21. Kotton CN, Hohmann EL (2004) *Infect Immun* 72:5535–5547.
22. Karaolis DK, Johnson JA, Bailey CC, Boedeker EC, Kaper JB, Reeves PR (1998) *Proc Natl Acad Sci USA* 95:3134–3139.
23. Lee SH, Hava DL, Waldor MK, Camilli A (1999) *Cell* 99:625–634.
24. Jermyn WS, Boyd EF (2002) *Microbiology* 148:3681–3693.
25. Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ (2002) *Proc Natl Acad Sci USA* 99:1556–1561.
26. O’Shea YA, Finnan S, Reen FJ, Morrissey JP, O’Gara F, Boyd EF (2004) *Microbiology* 150:4053–4063.
27. Silby MW, Rainey PB, Levy SB (2004) *Microbiology* 150:518–520.
28. Camilli A, Mekalanos JJ (1995) *Mol Microbiol* 18:671–683.
29. Gottesman S (2004) *Annu Rev Microbiol* 58:303–328.
30. Xu Q, Dziejman M, Mekalanos JJ (2003) *Proc Natl Acad Sci USA* 100:1286–1291.
31. Larocque RC, Harris JB, Dziejman M, Li X, Khan AI, Faruque AS, Faruque SM, Nair GB, Ryan ET, Qadri F, *et al.* (2005) *Infect Immun* 73:4488–4493.
32. Michalski J, Galen JE, Fasano A, Kaper JB (1993) *Infect Immun* 61:4462–4468.
33. Levine MM, Kaper JB, Herrington D, Losonsky G, Morris JG, Clements ML, Black RE, Tall B, Hall R (1988) *Infect Immun* 56:161–167.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) *J Mol Biol* 215:403–410.
35. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) *Nucleic Acids Res* 34:D354–D357.
36. McNeil LK, Reich C, Aziz RK, Bartels D, Cohoon M, Disz T, Edwards RA, Gerdes S, Hwang K, Kubal M, *et al.* (2007) *Nucleic Acids Res* 35:D347–D353.
37. Donnenberg MS, Kaper JB (1991) *Infect Immun* 59:4310–4317.