Global discovery of colonization determinants in the squid symbiont *Vibrio fischeri*

John F. Brooks II, Mattias C. Gyllborg, David C. Cronin, Sarah J. Quillin, Celeste A. Mallama, Randi Foxall, Cheryl Whistler, Andrew L. Goodman, and Mark J. Mandel

*Department of Microbiology-Immunology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611; Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH 03824; and Department of Microbial Pathogenesis and Microbial Sciences Institute, Yale University, New Haven, CT 06536*

Edited by Graham C. Walker, Massachusetts Institute of Technology, Cambridge, MA, and approved October 24, 2014 (received for review August 19, 2014)

Animal epithelial tissue becomes reproducibly colonized by specific environmental bacteria. The bacteria (microbiota) perform critical functions for the host's tissue development, immune system development, and nutrition; yet the processes by which bacterial diversity in the environment is selected to assemble the correct communities in the host are unclear. To understand the molecular determinants of microbiota selection, we examined colonization of a simplified model in which the light organ of *Euprymna scolopes* squid is colonized exclusively by *Vibrio fischeri* bacteria. We applied high-throughput insertion sequencing to identify which bacterial genes are required during host colonization. A library of over 41,000 unique transposon insertions was analyzed before and after colonization of 1,500 squid hatchlings. Mutants that were reproducibly depleted following squid colonization represented 380 genes, including 37 that encode known colonization factors. Validation of select mutants in defined competitions against the wild-type strain identified nine mutants that exhibited a reproducible colonization defect. Some of the colonization factors identified included genes predicted to influence copper regulation and secretion. Other mutants exhibited defects in biofilm development, which is required for aggregation in host mucus and initiation of colonization. Biofilm formation in culture and in vivo was abolished in a strain lacking the cytoplasmic chaperone DnaJ, suggesting an important role for protein quality control during the elaboration of bacterial biofilm in the context of an intact host immune system. Overall these data suggest that cellular stress responses and biofilm regulation are critical processes underlying the reproducible colonization of animal hosts by specific microbial symbionts.

bacterial colonization | symbiosis | biofilm | chaperone | functional genomics

Humans and other animals are often sterile before birth, from which point they immediately proceed to acquire environmental bacteria (1). The bacteria that reproducibly colonize animal hosts are critical for host tissue development, immune system development, and nutrient acquisition. The selection process by which the functional symbionts take residence in the animal from among the great diversity of environmental microbes is poorly understood, so model systems have been especially valuable to examine how specific patterns of colonization are shaped by the genetic makeup of the bacteria and the host environment (2).

The light organ of the Hawaiian bobtail squid, *Euprymna scolopes*, is colonized exclusively by the Gram-negative luminous bacterium *Vibrio fischeri*. The host inhabits seawater containing $10^8$ bacteria per milliliter, with *V. fischeri* comprising at most 0.02% of the environmental population (3). *E. scolopes* hatch without symbionts, but then rapidly acquire environmental bacteria and proceed to select for *V. fischeri* in a “winnowing” process that ensures colonization by only the specific symbiont (4). The squid–*Vibrio* system thus presents an opportunity to investigate the processes that underlie acquisition of specific environmental microbes. Forward genetic studies in *V. fischeri* have proven fruitful for identification of relevant colonization factors. Previous work included identification of Tn10 transposon mutants that individually failed to colonize the squid host; that work identified the RscS (regulator of symbiotic colonization-sensor) biofilm regulator and Syp (symbiosis polysaccharide) biofilm effectors that are required for robust colonization (5, 6). Global genetic approaches have been applied in culture to identify regulators of the hallmark luminescence phenotype, of flagellar motility, and in other signaling pathways (7, 8).

We sought to advance the utility of this model system by applying a global forward genetic approach to identify animal colonization factors. Signature-tagged mutagenesis was initially developed to identify factors that are specifically defective for growth in an animal host, and the original approach has been recently updated to include global analysis using Illumina deep sequencing (9–11). We applied such a technology, insertion sequencing (INSeq), to mutagenize *V. fischeri* MJM1100 (strain ES114), an *Euprymna scolopes* squid isolate (Table S1). The global data generated from this analysis allows for the first identification to our knowledge of essential bacterial genes in this organism and a genome-wide identification of genes that are conditionally required for colonization of the squid host. The previously undiscovered constituents in known pathways and the pathways suggested by this analysis identify key processes required for reproducible host colonization.

**Significance**

Animals form associations with bacteria that play important roles in host development and fitness. The mechanisms by which animals horizontally acquire their bacterial partners from the environment are poorly understood. To address this question, we take advantage of a natural symbiosis between the luminous Gram-negative bacterium *Vibrio fischeri* and its squid host, *Euprymna scolopes*. We applied the insertion sequencing global approach and identified 380 colonization determinants in *V. fischeri*. Characterization of the factors revealed novel biofilm regulation and beneficial colonization factors at the cell envelope. To our knowledge, our study is the first global functional analysis in *V. fischeri* and expands opportunities for systems biology approaches at the host microbe interface in a valuable reductionist model of microbiota colonization.
Results

Mutagenesis of *V. fischeri* and INSeq. To identify colonization determinants, it was first necessary to develop tools to globally analyze mutant pool composition in *V. fischeri*. The *Bacteroides thetaiotaomicron* mutagenesis vector pSAM was modified so that its resistance cassette and transposase promoter functioned in *V. fischeri* (10, 12). The resulting vector, *V. fischeri* mariner transposon delivery vector (pMarVF1), can be conjugated from *Escherichia coli* into *V. fischeri*, generating stable erythromycin-resistant insertions (Fig. S1). With this approach, we generated a library of over 41,000 independent erythromycin-resistant insertions (*Tn5*) in MJ1100. This is the “input” library for our analysis (Fig. 1A). Using the INSeq high-throughput method, the transposon-chromosome junctions were subsequently captured, subjected to limited amplification, and sequenced (10, 12) (SI Materials and Methods and Table S2), facilitating the identification of over 41,000 independent genomic insertion sites with high concordance across 10 technical replicates (Fig. 1B and Fig. S2).

Essential and Nonessential Genes in Culture. The mutagenesis approach affords the opportunity to identify potentially essential genes as those with reduced transposon insertion in the input mutant population. We quantified the transposon abundance (i.e., Illumina read counts) from the 5’-most 90% of each gene, and normalized those counts to the respective sample in a counts-per-million (cpm) measure. Genes that exhibited a low cpm (normalized to the size of the gene) represented likely essential genes and operons (including nonessential genes that exhibit polar effects on the downstream essential genes). To identify these loci, the transposon counts were plotted against the size of each respective gene. Mariner transposons insert at TA dinucleotides, so for gene size we used the count of such dinucleotides in the gene. 

We observed three clusters of genes (Fig. 1C): two clusters of likely essential genes included those with no transposon hits and those with a low number of normalized hits (0.03–4.5, depending on gene size). The third cluster exhibited significantly higher levels of transposon counts, and the genes in this cluster were therefore labeled as putative nonessential genes. Genes with fewer than nine TA dinucleotides were too small to reliably determine essentiality.

Onto the plot in Fig. 1C we mapped the *V. fischeri* genes that were orthologs of essential genes in *E. coli* MG1655, the organism for which the best curated list of essential genes exists (13). Most *E. coli* essential genes mapped to the clusters with no or relatively few transposon counts in *V. fischeri* (175 of the 230 orthologs encoded in *V. fischeri*), supporting their assignment as putative essential genes in *V. fischeri*. For the *V. fischeri* genes that are orthologs of essential *E. coli* genes, 55 fell in the higher (putative nonessential) gene cluster. For 36 of these genes, 85–100% of hits fell in the 3’ half of the gene, suggesting that the mutation allele still encodes a (partially) functional truncated protein. Therefore, we classified these genes as “putative essential” in *V. fischeri*. The remaining 19 genes had a more even distribution of hits across each gene, arguing that the gene is not essential in *V. fischeri*. The 19 essential *E. coli* genes that we predict to have nonessential *V. fischeri* orthologs include multiple genes whose products function as part of the same complex or biochemical pathway, supporting this assignment. Independent analysis using published software further confirmed their assignment as nonessential, as described in SI Materials and Methods. The genes that are predicted to be nonessential in *V. fischeri* include genes that have been demonstrated to be nonessential in the related species *Vibrio cholerae*: genes for riboflavin biosynthesis (*ribABDE*), phospholipid biosynthesis (*psbB*), other metabolic functions (*asd, can, cca*), cell division (*muk*), and secretion (*secDF*) (Table S3). Mutants in multiple *V. fischeri muk* genes were identified in a previous study, and in the VFS (*Vibrio fischeri* sublibrary) collection 9 of the 19 predicted nonessential genes harbor Tn5 transposon insertions (Table S3) (8). Included in the list of essential *E. coli* genes that are not essential in *V. fischeri*—and that are essential in *V. cholerae* or for which there exist conflicting data in different strains or studies—are multiple genes involved in cell division (*ftsK, mrdB, nodA, zipA*), LPS biosynthesis (*kdsA, kdsB*), and ubiquinone biosynthesis (*ubid*) (Table S3). Comparative analysis of the essential genes therefore suggests conserved and unique features of *Vibrio* and *V. fischeri* biology for
these factors. Importantly for this study, these results highlight that the INSeq analysis identifies relevant signal within a mutant library, and we therefore proceeded to examine genes that were depleted following squid colonization.

**Identification of Squid Colonization Factors.** For experimental inoculation of *E. scolopes* bacterial concentrations of $10^7–10^8$ cfu/mL are typically used (14, 15). It was demonstrated that higher inoculum concentrations led to a greater diversity of strains within the light organ, so we used an inoculum of $2 \times 10^7$ cfu/mL (16, 17). To ensure that sufficient strain diversity could be achieved in the light organ, we colonized squid with a defined 96-mutant library and then analyzed the individual animal output samples by INSeq. Each animal contained over 80 of the initial 96 mutants (Fig. S2A), revealing that even with some initial bottleneck we can obtain a large number of independent mutant strains in an individual animal. We estimated that with 250 animals, we could conservatively sample 20,000 mutants (250 animals $\times$ 80 mutants minimum per animal). We therefore sought to analyze the distribution of the input mutant library following colonization in six replications of 250 animals each, which would give average coverage of at least 35-fold per nonessential gene. In many systems, the natural route of inoculation is bypassed to investigate colonization in vivo, in contrast to inoculation of *E. scolopes* where bacteria are introduced to the seawater and then acquired by the host’s natural mechanisms. We observed some noise in the squid output libraries (Figs. S2B and S3), which likely reflects some variation in the extent to which the host bottleneck influenced symbiont diversity across the large number of animals in our analysis.

To avoid the study of mutants that were simply defective for growth, the input library was passaged in LB salt (LBS)-rich medium and the 115 genes that exhibited >10-fold mutant depletion after 15 generations were excluded from further analysis (Fig. 24). We proceeded to use the dynamics of known colonization mutants to train our analysis to identify true, novel colonization factors. First, we observed that in most cases, mutants in known colonization factors were observed in every squid output library, and they were diminished in each library with respect to their relative abundance in the input library (Fig. 2C and Fig. S3). In contrast, 967 genes were represented by mutants that were not sampled in at least one squid output library, and because known squid colonization factors were not highly represented in this dataset, we suspect that some of these factors are artifacts (i.e., seawater growth defective; or low representation in the input library and therefore dramatic sampling bias in the squid). For subsequent analysis, we therefore took a conservative approach and focused on the 2,291 genes for which mutants were sampled in all squid libraries, and for which representation in each output library diminished by at least 1.8-fold (Fig. 2B). This set of 380 squid-depleted mutants captured over 75% of known mutants during the first two of the three defined stages of squid colonization (i.e., initiation and accommodation; Fig. 2C and Fig. S3), validating our analysis criteria were achieved. The INSeq squid-depleted mutants identified only 20% of known persistence-stage colonization factors (Fig. 2C), and finer-scale analysis of these factors revealed significant variation across the
different replicates (Fig. S3), which in most cases excluded them from our selection criteria.

To understand global patterns in the dataset, we categorized genes by the COG (clusters of orthologous groups) classification. First, we examined the *V. fischeri* essential genes. The categories that were most represented included translation, cell envelope, coenzyme transport and metabolism, lipid transport and metabolism, and replication (Fig. 2D). These assignments track with essential cellular processes, validating the use of COG categories as a tool to understand mutant dynamics in *V. fischeri*. We proceeded to categorize the squid-depleted genes in a comparable manner. In contrast to the overall essential genes, the colonization factors exhibited a markedly different pattern (Fig. 2D). Over 20% were not included in known groups (i.e., they were referred to as R or S). Categories that included a large number of the colonization factors included those for signal transduction, motility, amino acid transport and metabolism, inorganic ion transport and metabolism, and cell envelope integrity (Fig. 2D). The COG analysis highlighted broad-scale classes of genes that are required for squid colonization. We then proceeded to examine individual factors both in vivo and for specific behaviors in vitro.

Validation of Individual Candidates in 1:1 Competition Assays. The INSeq analysis involved competition of over 41,000 mutants at high inoculum concentrations. To identify colonization factors that are required across a broad range of inoculation concentrations and conditions, we examined 20 factors predicted by INSeq to have a fitness deficit in the animal. We introduced each mutant in competition with a marked wild-type strain at a 1:1 ratio and then examined the relative ratio of the mutant following colonization. Nine of the strains exhibited fitness deficits in the 1:1 competition assay relative to wild type, ranging from 2- to 65-fold depletion in the host (Fig. 3A and Fig. S4). These nine strains include mutants in genes encoding two copper efflux systems (CusC and CopA); two protein quality control factors (cytoplasmic chaperone DnaJ and periplasmic endoprotease DegS; an inner membrane protein predicted to assist in the secretion of autotransporters (TamB/YtfN); a predicted lysine 2,3-aminomutase (YjeK/EpmB); and three poorly characterized proteins (YdhC, YafD, and YhcB) (Fig. 3A). We conducted the same competitions under squid-free culture conditions to control for the host environment (Fig. 3B); two of the nine mutants (ydhC and yjeK/epmB) exhibited competitive defects in culture that fully explained their respective squid colonization deficiencies, arguing that these genes encode factors that are not specifically required for colonization of the host environment.

*V. fischeri* must synthesize a symbiotic biofilm to passage through the squid ciliated epithelial field to the light organ pore. Biofilm assembly in vitro correlates strongly with the ability to synthesize a symbiotic biofilm to passage through the squid ciliated epithelial field to the light organ pore. This assembly is known to be critical for symbiotic development and host colonization specificity (15, 18). DnaJ/DnaK Regulate Symbiotic Biofilm Development. Elaboration of the Syb symbiotic biofilm is critical for squid colonization (15, 18). We therefore sought to further examine the phenotype of dnaJ, which exhibited no wrinkled colony formation in the Syb biofilm assay (Fig. 3B). We constructed a dnaJ deletion and it exhibited the same biofilm defect; complementation with the wild-type allele restored full wrinkled colony development (Fig. 4A). DnaJ (Hsp40) is conserved from bacteria through metazoans and must synthesize a symbiotic biofilm to passage through the squid ciliated epithelial field to the light organ pore.

<table>
<thead>
<tr>
<th>Competitive Index</th>
<th>Squid (in vivo)</th>
<th>Culture (in vitro)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>WT</td>
<td>cusc</td>
</tr>
<tr>
<td>b</td>
<td>Depletion</td>
<td>0.7</td>
</tr>
<tr>
<td>c</td>
<td>Normalized to in vitro</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Fig. 3. Validation of host colonization factors identified by INSeq analysis. (A) Competition (1:1) of the indicated mutant strain versus a LacZ-marked wild-type strain in squid. Hatching squid were inoculated with 2 × 10^7 cfu/mL bacteria, washed at 3 and 24 h, and assayed at 48 h. The competitive index (CI) is the Log_{10} (mutant/wild type), where the mutant/wild-type output ratio is normalized to the input ratio as detailed in SI Materials and Methods. Dots represent the CI for individual squid, and bars represent the median. The value below the strain represents the fold depletion of the strain (e.g., a median CI of −1 would equal a 10-fold depletion). (B) Competition (1:1) of the indicated mutant strain versus a LacZ-marked wild-type strain in culture. Log phase cells were mixed, diluted, and grown for 15 generations in LBS medium. Dots represent the CI for three independent replicates, and bars represent the median. The competitive fold depletion was normalized to the in vitro competition as described in the SI Materials and Methods, and the normalized in vivo depletion is shown in the gray shaded box. (C) Wrinkled colony biofilm analysis at 24 h and 48 h after plating the indicated strains containing the biofilm-inducing pKG11 plasmid on LBS-tetracycline. (D) Single-strain growth curves in LBS medium. Strains were grown in a microplate for 30 h. The gray filled curve is shown similarly in all plots and represents the wild-type growth. Mutant growth curves are shown with the black lines.

formation, a behavior that is known to be critical for symbiotic development and host colonization specificity (15, 18).
in vivo. We proceeded to test directly whether DnaJ influences biofilm development in the host mucociliary field. We colonized the host tissue (21). Detoxification of intracellular copper is achieved by efflux via P-type ATPase and RND (resistance-nodulation-cell division) family efflux systems (22). We observed that V. fischeri lacking the P-type ATPase CopA or components of the RND efflux system CusCFBA exhibited significant colonization defects in the INSeq analysis and in 1:1 colonization stages. Oxygen in squid is carried by hemocyanin, a metal-ligand protein that uses a pair of copper ions to reversibly bind the oxygen. Copper coordination is predicted to be conserved in E. scolopes, and there is intimate communication between host and microbe, including known involvement of hemocyanin and oxygen (23). Copper may therefore be critical to the dialogue that leads to establishment of a secure and robust partnership.

We describe 380 candidate colonization factors, and as proof of principle we rediscovered over 75% of the known colonization initiation and accommodation factors (Fig. 2C). Mutants that affect the later persistence stage were not well represented in this analysis. We therefore predict that most of the previously unidentified factors will map to the initiation or accommodation stages, a claim that is supported by the discovery of three biofilm (i.e., initiation) regulators in the initial set of validated factors.

There is an increasing appreciation that metal concentrations are critical during host colonization. It is established that copper is an effective antimicrobial agent ex vivo, and there is evidence that it may play a similar role in vivo. Pathogens that lack copper efflux systems colonize poorly in the mammalian lung, possibly because of an inability to detoxify increased levels of copper in the host tissue (21). Detoxification of intracellular copper is achieved by efflux via P-type ATPase and RND (resistance-nodulation-cell division) family efflux systems (22). We observed that V. fischeri lacking the P-type ATPase CopA or components of the RND efflux system CusCFBA exhibited significant colonization defects in the INSeq analysis and in 1:1 colonization stages. Oxygen in squid is carried by hemocyanin, a metal-ligand protein that uses a pair of copper ions to reversibly bind the oxygen. Copper coordination is predicted to be conserved in E. scolopes, and there is intimate communication between host and microbe, including known involvement of hemocyanin and oxygen (23). Copper may therefore be critical to the dialogue that leads to establishment of a secure and robust partnership.

Our screen identified a role for the two-partner secretion system TamAB (YtfMN) during host colonization. This system was identified as a colonization determinant in signature-tagged mutagenesis screens in Proteus mirabilis and Klebsiella pneumoniae (24, 25). From work in Enterobacteriaceae the TamAB system has been suggested to function in the tranlocation of autotransporters; however, V. fischeri ES114 genome does not have any annotated autotransporters (26). This suggests the presence of an unannotated autotransporter in V. fischeri—which our data would argue play an important role during host colonization—or a novel role for the TamAB secretion system. In other systems it is not clear how TamAB impacts host colonization, and the V. fischeri model provides an opportunity to examine this further.
The discovery of genes with unknown or predicted functions as required for robust host colonization provides an opportunity to interrogate how characterized gene products influence colonization behavior. Some of these genes (yhcB and yakJ) exhibit no defects in culture and influence host colonization with such dramatic phenotypes that they exert them within 48 h during both massive competition and during 1:1 competitions. The Vibrio–squid system is therefore a sensitive one in which to interrogate the function of uncharacterized gene products during host association.

Previous reports in bacteria and yeast suggest a possible role for various chaperone systems in biofilm development. Increased expression of chaperone genes is observed in environmental biofilms (27) and in Pseudomonas aeruginosa tobramycin-treated biofilms (28). However, the functional role of these chaperones has been difficult to examine. DnaK influences biofilm development in *Listeria monocytogenes* (29), and treatment of *Hilarius sommi* biofilms with anti-GroES blocking antibodies leads to reduced biofilm formation (30). Furthermore, depletion experiments in *Candida albicans* have identified roles for Hsp90 (HtpG) and Hsp104 (ClpB) in regulating biofilm formation, arguing for a broad connection between protein folding and biofilm development. Here we demonstrate that the DnaJ/DnaK chaperone system regulates symbiotic biofilm formation in *V. fischeri*, and that aggregation in host tissue is abolished in the absence of the system. Prior reports have connected nitric oxide signaling to the size of the bacterial biofilm-like aggregates at this stage (31, 32). Further investigation of the basis of the DnaJ/DnaK aggregation defect should examine how microbial proteostasis regulates biofilm formation upon encountering host immune challenges.

### Materials and Methods

Detailed methods are provided in SI Materials and Methods.

### Mariner Transposition and INSeq Analysis in *V. fischeri.*

The *V. fischeri* mariner transposon delivery vector pMarVF1 was constructed to facilitate INSeq analysis in *V. fischeri* ES114. The suicide delivery vector carries an INSeq-compatible transposon, which encodes an erythromycin-resistance cassette and a backbone that includes an R6K origin of replication, RP4 transfer origin, beta-lactamase (i.e., ampicillin resistance), and expresses the Himar1C9 transposase from the ES114 *nrdr* promoter.

### Squid Colonization.

Colonization of hatching squid is detailed fully in SI Materials and Methods. For the assays performed, the inoculum concentrations, inoculation time, and total colonization time are as follows: INSeq screen: 2 × 10<sup>5</sup> CFU/mL, 3 h and 48 h; in vivo 1:1 competitions: 2 × 10<sup>6</sup> CFU/mL, 3 h and 48 h; single-strain colonization assays (DnaJ): 2 × 10<sup>5</sup> CFU/mL, 3 h and 24 h; and visualization of aggregates: 2 × 10<sup>6</sup> CFU/mL, 3 h and 3 h.

### ACKNOWLEDGMENTS.

We thank Karen Visick, Caitlin Brennan, and Gregory Tyson for contributing reagents or preliminary data; Michael Hadfield and the Kewalo Marine Laboratory for *Euprymna* field resources; and Natacha Ruiz and Karla Satchell for comments on the manuscript. This work was supported by National Science Foundation (NSF) Grant IOS-0843633 (to M.J.M.), National Institutes of Health Grant DK089121 (to A.L.G.), and NSF Grant IOS-1258990 (to C.W.). J.F.B. was supported by National Institute of General Medical Sciences Cellular and Molecular Basis of Disease T32 Training Grant GM08061.

---