

Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners

E. G. Ruby^{*†}, M. Urbanowski[‡], J. Campbell[§], A. Dunn[¶], M. Fainil^{||}, R. Gunsalus^{**}, P. Lostroh[‡], C. Lupp^{*}, J. McCann^{*}, D. Millikan^{*}, A. Schaefer^{*}, E. Stabb[¶], A. Stevens^{||}, K. Visick^{††}, C. Whistler^{*}, and E. P. Greenberg[‡]

^{*}Pacific Biomedical Research Center, Kewalo Marine Laboratory, and W. M. Keck Microbial Communities and Cell Signaling Program, University of Hawaii, Honolulu, HI 96813; [‡]Department of Microbiology and W. M. Keck Microbial Communities and Cell Signaling Program, University of Iowa, Iowa City, IA 52242; [§]Integrated Genomics, Chicago, IL 60612; [¶]Department of Microbiology, University of Georgia, Athens, GA 30602; ^{||}Department of Biology, Virginia Polytechnic Institute, Blacksburg, VA 24061; ^{**}Department of Microbiology and Molecular Genetics, University of California, Los Angeles, CA 90095; and ^{††}Department of Microbiology and Immunology, Loyola University Chicago, Maywood, IL 60153

Contributed by E. P. Greenberg, December 31, 2004

Vibrio fischeri belongs to the Vibrionaceae, a large family of marine γ -proteobacteria that includes several dozen species known to engage in a diversity of beneficial or pathogenic interactions with animal tissue. Among the small number of pathogenic *Vibrio* species that cause human diseases are *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Vibrio vulnificus*, the only members of the Vibrionaceae that have had their genome sequences reported. Nonpathogenic members of the genus *Vibrio*, including a number of beneficial symbionts, make up the majority of the Vibrionaceae, but none of these species has been similarly examined. Here we report the genome sequence of *V. fischeri* ES114, which enters into a mutualistic symbiosis in the light organ of the bobtail squid, *Euprymna scolopes*. Analysis of this sequence has revealed surprising parallels with *V. cholerae* and other pathogens.

genomics | pili | symbiosis | toxins | toxin-coregulated pilus

The marine bacterium *Vibrio fischeri* is best known as the specific symbiont in the light-emitting organs of certain squids and fishes (1), where it produces luminescence by expressing the *lux* operon, a small cluster of genes found in several of the Vibrionaceae. Luminescence is controlled by acyl-homoserine lactone quorum sensing, which was discovered in *V. fischeri* but is a common feature of host-associated bacteria in a number of genera (2). The Vibrionaceae comprise several dozen species that are often found associated with animal tissue (3). Among the small number of pathogenic *Vibrio* species that cause human diseases are *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Vibrio vulnificus*, the only members of the Vibrionaceae that have had their genome sequences reported (4–6). To date, no representative of the much more numerous benign and beneficial Vibrionaceae has been examined at the genome level.

Unlike many symbiotic bacteria, *V. fischeri* has been studied at the physiological and genetic levels for decades, has a well described environmental biology, and is easily examined by using molecular genetics (7). Thus, this particular nonpathogenic member of the Vibrionaceae is an ideal candidate for comparative genome analyses with pathogenic vibrios. The best understood of the *V. fischeri* symbiotic associations are those with sepiolid squids. These symbioses involve monospecific populations of *V. fischeri* cultured extracellularly, but within epithelium-lined crypts, in a specialized host organ. The squid associations have been extensively studied because of the ease of initiating the association and of observing developmental changes in both partners (8). Nevertheless, important questions remain concerning the genetic and metabolic mechanisms by which *V. fischeri* and other symbiotic bacteria adjust to the special environment of host tissues. To better understand these symbiotic activities in *V. fischeri*, and to begin to identify features common to beneficial and pathogenic bacteria, we sequenced the genome of strain ES114, the model light-organ symbiont of the squid *Euprymna scolopes* (9).

Materials and Methods

V. fischeri genomic DNA was mechanically sheared, and 2- to 3-kb fragments were isolated. The ends of the fragments were filled by using Klenow fragment and ligated into *Sma*I-digested pGEM3 to produce a high-copy-number, small-insert library (10). Higher molecular mass genomic DNA was partially digested with *Sau*3A to construct a cosmid library containing 30- to 35-kb inserts in Lorst6 (11). Undigested, unsheread DNA was used as template for PCR amplification of chromosomal regions not represented in the plasmid or cosmid libraries. Whole-genome shotgun sequencing was performed on \approx 35,000 plasmids and 400 cosmids, as well as gap-spanning PCR products. The average PHRAP consensus quality was 82. Automated contig-assembly algorithms were ineffective in determining the number and orientations of the highly homologous, tandemly repeated, ribosomal RNA operons; thus, these regions were assembled manually from sequenced PCR products. The genome, with an average coverage of 10 \times , was assembled into three contigs. Potential ORFs were identified by using both CRITICA (12) and a program developed at Integrated Genomics (Chicago). The finished genome underwent a round of initial automated annotation, followed by a manual gene-by-gene curation. The results of this analysis can be found at www.ergo-light.com.

Results and Discussion

General Features. As in each of the three sequenced pathogenic *Vibrio* species (4–6), *V. fischeri* has two chromosomes (Fig. 1). Genes encoding representatives of 25 classes of cellular function can be found on each of these replicons (Fig. 2). Strain ES114 is also characterized by the presence of a 45.8-kbp plasmid designated pES100. Carriage of a plasmid that is homologous to pES100 is common among other symbiotic strains of *V. fischeri*, but it is not required for host association (13). The sequence of pES100 suggests that, like a similarly sized replicon in *V. vulnificus* YJ016 (4), it is a conjugative plasmid, much of whose sequence encodes a putative type IV secretion system. Thus, such plasmids are common to both beneficial and pathogenic *Vibrio* species.

One of the striking characteristics of the *V. fischeri* genome is the G+C content of the DNA (Table 1), which is the lowest of the 27 species of Vibrionaceae (14). Despite this very low G+C content, *V. fischeri* is more closely related to the higher G+C pathogenic *Vibrio* species than to any other sequenced bacterium (www.ergo-light.com). The genome-wide value of

Abbreviations: Chr, chromosome; CTX, cholera toxin; TCP, toxin-coregulated pilus.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. CP000020 (large chromosome), CP000021 (small chromosome), and CP000022 (plasmid)].

[†]To whom correspondence should be addressed. E-mail: egruby@wisc.edu.

© 2005 by The National Academy of Sciences of the USA

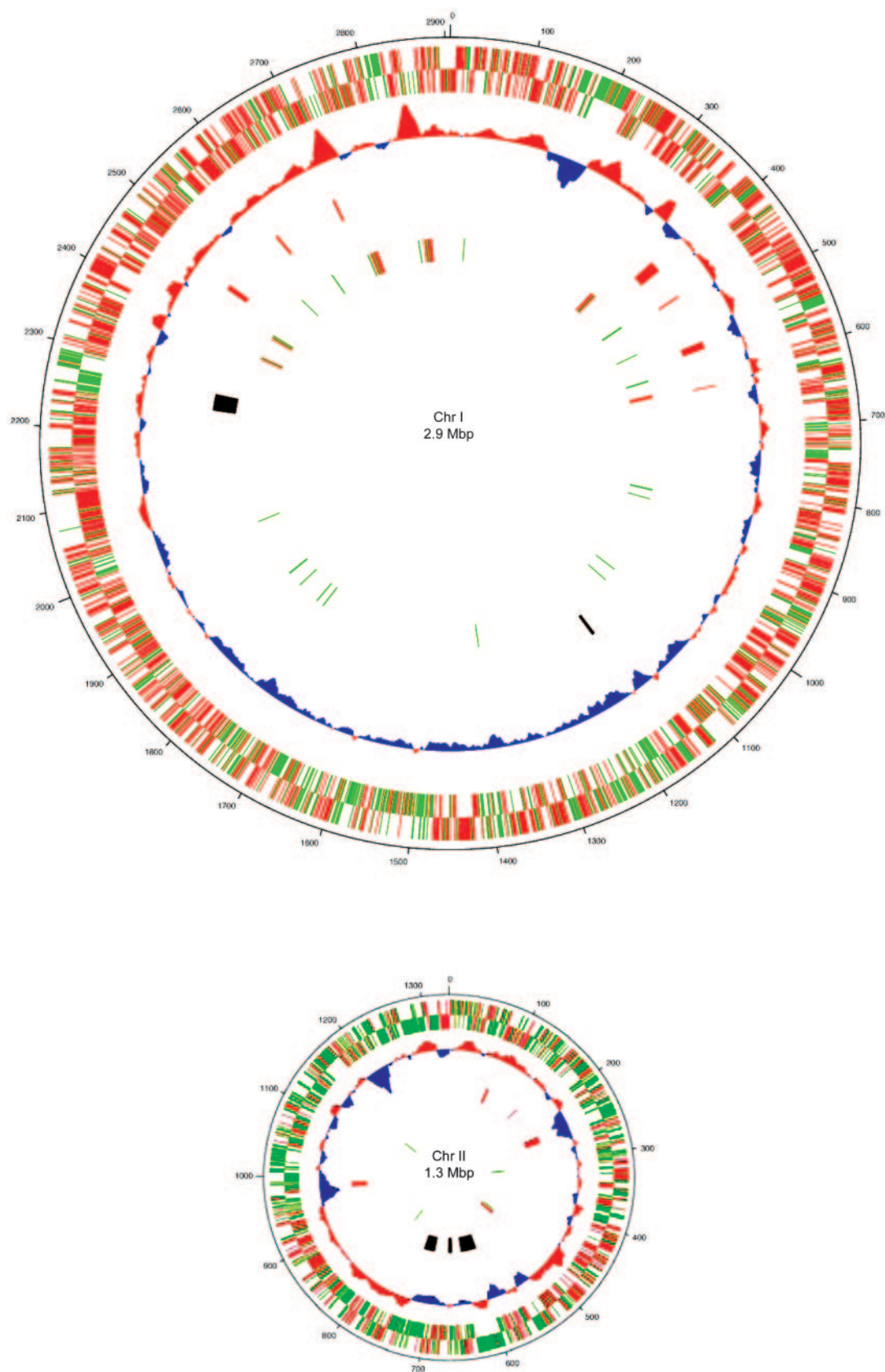
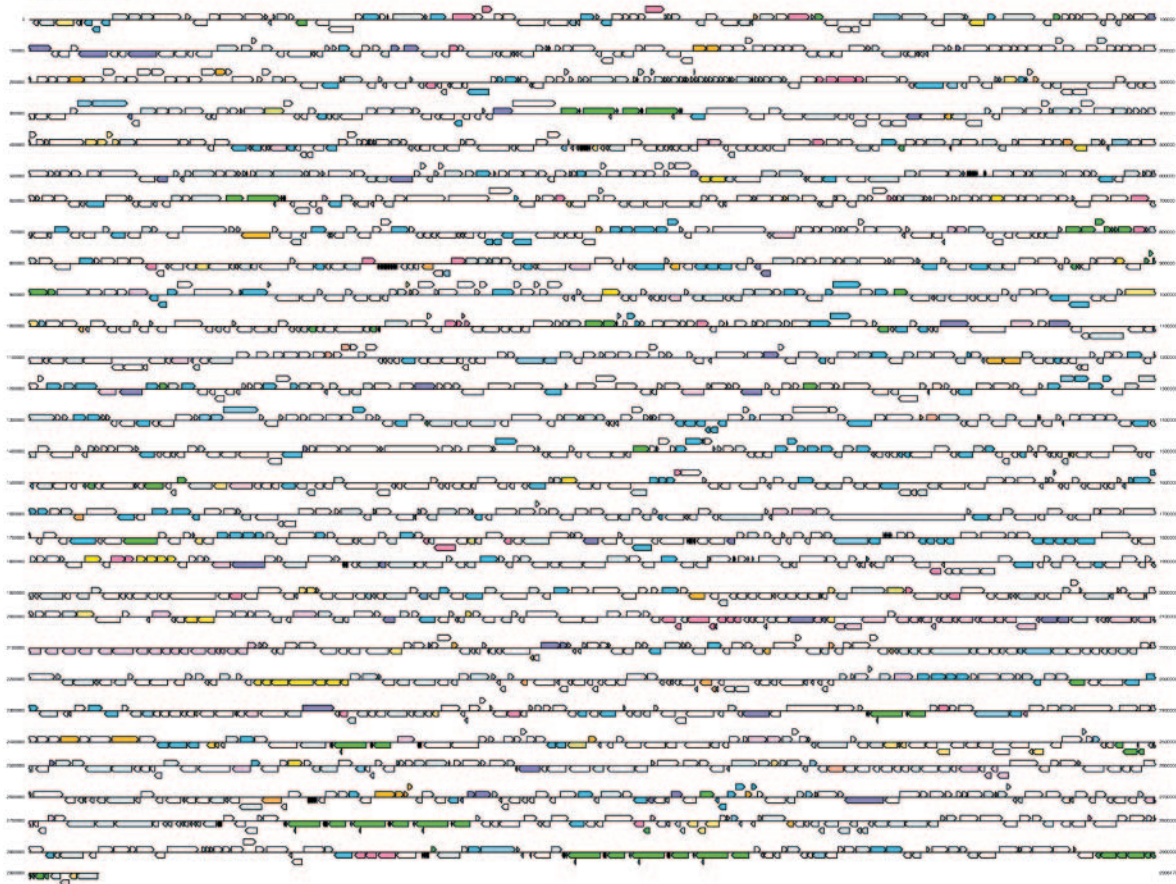


Fig. 1. Genome maps of Chr I (2.9 Mbp) and Chr II (1.3 Mbp) of *V. fischeri* ES114. (From the inside) Ring 1, *rrn* operons (red); tRNAs (green). Ring 2, "foreign" elements (black). Ring 3, type IV pilus loci (red). Circle 1, G+C content over a 200-kb window with 5-kb steps; red and blue denote G+C content higher and lower than average, respectively. Circle 2, homology on + and - strands to *V. cholerae* N16961 ORFs, determined by using FASTA; red, *e* value < -30; green, *e* value > -30. Circle 3, location in kilobases.

Chromosome 1



Chromosome 2

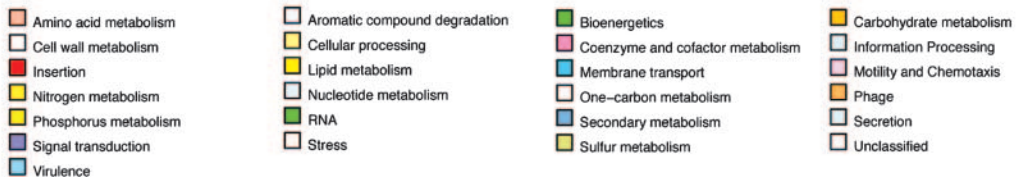
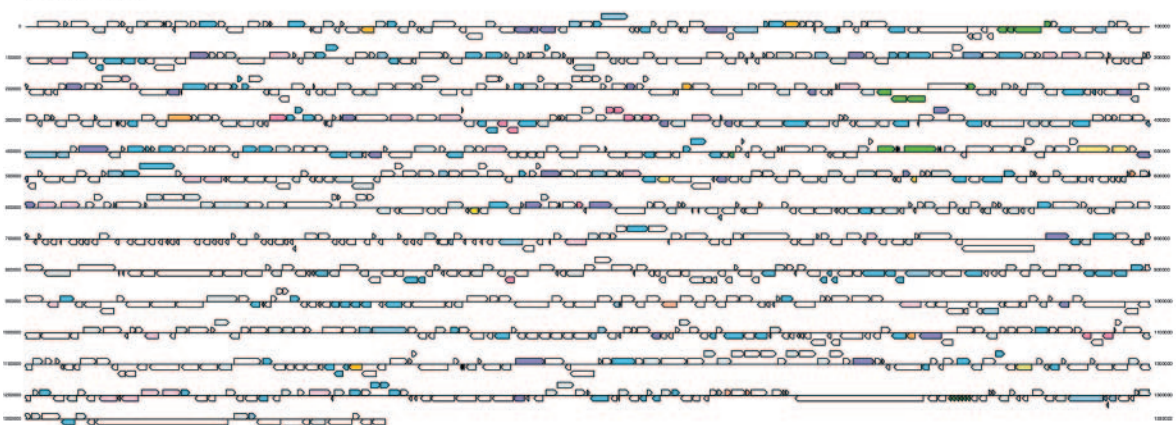


Fig. 2. *V. fischeri* E5114 ORFs on the + and - strands, color-coded to functionality (see key).

Table 1. Comparison of general features of the *V. fischeri* ES114 and *V. cholerae* N16961 genomes

Property	Species	
	<i>V. fischeri</i>	<i>V. cholerae</i>
Chromosomes (plasmid)	2 (1)	2 (0)
Size, bp	4,284,050	4,033,464
Number of ORFs	3,802	3,915
G + C percent	38.3	47.5
Number of rRNA operons		
Chr I	11	8
Chr II	1	0
Number of tRNA genes		
Chr I	108	94
Chr II	11	4
ORFs with no similarity	186 (4.7%)	363 (9.3%)
Chr I	59 (2.2%)	210 (7.6%)
Chr II	101 (8.6%)	153 (13%)
Plasmid	26 (47%)	–

Data are from the ERGO Light database (www.ergo-light.com).

38.3% G+C (Table 1) is similar to the G+C values of each of the three individual replicons, suggesting they have all had a long history in this species. The distribution of 16S ribosomal RNA (*rrn*) operons in *V. fischeri* is similar to that reported for *V. parahaemolyticus* and *V. vulnificus*. Specifically, there is one *rrn* on the smaller chromosome (Chr II) of *V. fischeri*, whereas the 11 remaining operons are on the larger chromosome (Chr I) (Fig. 1). Thus, the absence of an *rrn* operon on the small chromosome in *V. cholerae* (5) appears atypical in this genus. In *V. fischeri*, most of the 11 *rrn* operons are clustered in three loci (Fig. 1). Such extended series of sequential *rrn* operons, creating patches of higher G+C content, have not been observed in the other *Vibrio* species. It is also of interest to note that the chromosomal density of apparent ORFs is almost 10% greater in *V. cholerae* compared to *V. fischeri* (Table 1).

An analysis of the two chromosomes of *V. cholerae*, the first sequenced *Vibrio* genome, revealed that ORFs located on Chr II were twice as likely to be unique to this organism than those on Chr I (5). At that time, it was not clear whether this result reflected a dearth of genes reported from other Vibrionaceae or indicated that the smaller replicon was a repository for genes unique to *V. cholerae*. A similar analysis of the two chromosomes of *V. fischeri* indicates that, even with the genomes of four members of the genus *Vibrio* in the database, there is an \approx 4-fold greater percentage of unique genes on Chr II of this species (Table 1). Thus, the smaller chromosomes characteristic of this genus may turn out to be a rich source of genes that define the unique potential of individual *Vibrio* species, and perhaps their specific lifestyles.

Mobile Elements. As in most bacteria, *V. fischeri* carries evidence of mobile genetic elements on its chromosomes. Although little is known about the importance of these elements in *V. fischeri*, in other bacteria they play a role in obtaining genes encoding virulence factors or resistance to environmental stresses (15, 16). Evidence of a retron, an integron, and three phage-like loci is found in the genome (Fig. 1). The most intriguing of these foreign elements is a cholera toxin (CTX) phage-like gene cluster on Chr II (Fig. 3). This element is composed of eight ORFs, including four homologs of CTX-phage genes (17): *cep*, *orfU*, *ace*, and *zot*. The *V. fischeri* locus differs from the well studied *V. cholerae* El Tor CTX phage in two important ways: it is missing the downstream RS2 cluster believed to be required for phage excision and multiplication

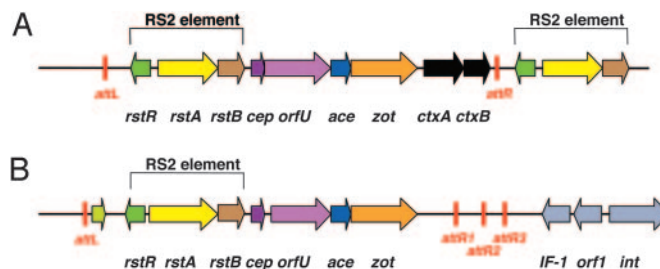


Fig. 3. Comparison of the *V. cholerae* CTX phage locus (A) with a homologous locus in *V. fischeri* ES114 (B). ORFs encoding the CTX genes *ctxA* and *ctxB*, and the second RS2 element downstream of the locus, are missing in *V. fischeri*. Identifiable *attL* and *attR* sites are indicated.

(18) and, in place of the CTX genes (*ctxAB*), there are two apparently truncated ORFs. Interestingly, the *V. fischeri* locus shares greater sequence similarity to the *V. cholerae* CTX phage than the severely diminished homologous locus present in *V. parahaemolyticus* RIMD 2210633.

Pilus Gene Clusters. Extracellular pilus structures are common among bacteria, and have been implicated in diverse colonization functions (19). Several kinds of pili have been described in *Vibrio* species, some of which are required for pathogenesis (20, 21). The *V. fischeri* genome contains 10 separate pilus gene clusters, including eight type-IV pilus loci (Table 2). Four of these clusters are found on Chr I of all of the sequenced *Vibrio* species. They include a mannose-sensitive hemagglutinin (MshA), a homolog of the PilT pilus involved in twitching motility (22), and two *Vibrio* PilA homologs (21, 23). Two of the remaining four pilus gene clusters are paralogs (one on each *V. fischeri* chromosome) that are homologous to the Flp1 type IV-B tight adhesion (*tad*) pilus family. Members of this evolutionarily diverse family are found in many bacteria (24), including *V. parahaemolyticus* and *V. vulnificus*, but not the sequenced *V. cholerae* strain. The last two type-IV pilus loci of *V. fischeri* have homologs only in *V. cholerae*: one encodes a highly diverged PilA-like gene (7), and the other is a toxin-coregulated pilus (TCP)-encoding locus (25) (Fig. 4). The two

Table 2. Putative pilin loci in *V. fischeri*

Pilin name	Pilus gene locus* (kb)	Class	Presence in other <i>Vibrio</i> spp. [†]
Chr I			
MshA	VF0355-0371 (390)	Type IV	Vc, Vp, Vv
PilT	VF0431-0432 (460)	Type IV	Vc, Vp, Vv
Flp1A	VF0510-0523 (550)	Type IVB	Vp, Vv
PilA1	VF0571-0572 (630)	Type IV	Vc, Vp, Vv
PilA	VF2184-2191 (2,400)	Type IV	Vc, Vp, Vv
CsgA	VF2405-2411 (2,700)	Curli	None
Chr II			
PilA2	VFA0148 (166)	Type IV	Vc [‡]
Flp1	VFA0217-0232 (240)	Type IVB	Vp, Vv
TcpA	VFA0865-0875 (980)	Type IVB	Vc
pES100			
VirB2	VFB38-44; 54-55 (26)	Conjugative	None

*Data are from the ERGO Light database; value in parentheses is the relative location along each contig (see Fig. 1). ORFs designated "VF" are located on Chr I, those designated "VFA" are located on Chr II, and those designated "VFB" are located on pES100.

[†]Compared to the sequenced genomes of *V. cholerae* N16961 (Vc), *V. parahaemolyticus* RIMD 2210633 (Vp), and *V. vulnificus* CMCP6 (Vv) in the ERGO Light database.

[‡]Low ($e = 10^{-4}$) but detectable amino acid sequence identity.

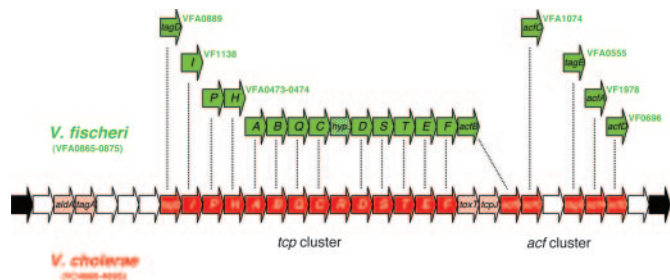


Fig. 4. The TCP gene cluster in *V. fischeri* ES114 (green) and *V. cholerae* N16961 (red). *V. fischeri* has no identifiable homologs of *aldA*, *tagA*, *toxT*, or *tcpJ*, and the *V. fischeri* ORF labeled "hyp." has no detectable similarity to *V. cholerae* *tcpR*.

non-type-IV pilus clusters are not found in any other sequenced *Vibrio* species, and include homologs of the curl-encoding genes of enteric bacteria (26) and a putative conjugative pilin encoded on pES100.

Mutations in only two pilin genes, both on Chr II (Table 2), have been analyzed for their effects on symbiosis. The first is the Flp1 paralog, which is required for achieving normal colonization levels in the light organ.^{††} The second is the PilA2 pilus, which facilitates efficient light organ colonization by bacteria (7). The presence of multiple pilus gene clusters in the *V. fischeri* genome suggests that different pili may be expressed to aid this bacterium either in the diverse environments it inhabits or during the multiple stages of its development as a symbiont (27).

The TCP Genes in *V. fischeri*. In *V. cholerae*, the TCP is an important virulence factor (28). The genes encoding its synthesis are located in a single large cluster (Fig. 4) that is believed to have entered the genome as a *Vibrio* pathogenicity island (VPI) (29, 30). Homologs of most of the TCP genes are present in the genome of *V. fischeri*. A majority of the genes in the *tcp* cluster exhibit synteny between the two species (Fig. 2); however, eight of the homologs are located elsewhere in the *V. fischeri* genome. Although the reasons for the different gene arrangements in *V. cholerae* and *V. fischeri* are unknown, the G+C content and apparent absence of flanking insertion elements in the *V. fischeri* cluster suggest that it was not horizontally acquired in the recent past (Fig. 5). In contrast, the presence of insertion elements, as well as the G+C content of the *V. cholerae* VPI, support the idea that this region is foreign, and originated in an unusually low G+C genome like that of *V. fischeri*.

Homologs of Toxin-Encoding Genes. Although *V. fischeri* is not known to be pathogenic, and strain ES114 is a beneficial symbiont, its genome carries homologs of *Vibrio* genes that may have toxin activity (Table 3, which is published as supporting information on the PNAS web site). As mentioned earlier, these include two CTX phage-encoded genes, *zot* (zona occludens toxin) and *ace* (accessory cholera enterotoxin), the latter of which has been found only in the *V. cholerae* and *V. fischeri* genome sequences. The proteins coded for by these genes have been shown to contribute to the structure of CTX phage (17), and their possible roles as toxins remain controversial (31, 32). At this time, it is not known whether the *V. fischeri* homologs of these two genes are expressed in this species, or whether they might play a role in this bacterium's symbiotic associations.

All sequenced *Vibrio* species carry genes encoding another putative toxin called RTX (repeats in structural toxins). RTX activity in *V. cholerae* appears to affect regulators of host-cell

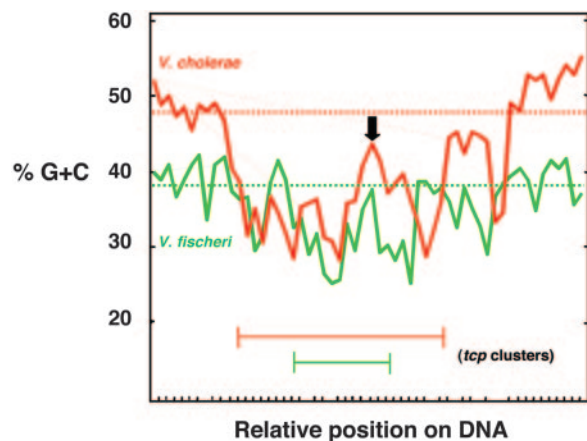


Fig. 5. G+C content of ORFs in and around the TCP gene clusters of *V. fischeri* ES114 (green) and *V. cholerae* N16961 (red) plotted against their relative linear position on the chromosome. The average G+C content of the two species' genomes are indicated by the colored dotted lines, and the arrow points to the locations of their TcpA pilin ORFs, around which the rest of the sequences have been aligned.

actin polymerization, causing dramatic changes in cytoskeletal structure, leading to the loss of tight-junction integrity (31, 33). The expression or potential activity of a *V. fischeri* RTX protein has not yet been investigated, but it is intriguing that symbiotic infection of the squid host results in a set of controlled changes in actin deployment in epithelial cells surrounding the bacteria in the light organ (34, 35). Because these changes are a part of the host's normal developmental program, the RTX protein may provide an important signal in symbiosis. Similarly, it has been discovered recently that *V. fischeri* secretes a degradation product of peptidoglycan that is responsible for inducing normal tissue development in the nascent light organ (36). This extracellular product is identical in structure to tracheal cytotoxin (TCT), produced by *Bordetella pertussis* and *Neisseria meningitidis*, indicating that bacterial virulence factors not only are context specific, but also may be required symbiotic signals. The protein believed to be responsible for TCT production may be encoded by a distinct transglycosylase gene that is present in these two pathogens and *V. fischeri*, but is not found in the other sequenced *Vibrio* species (A. Schaefer, unpublished data). The carriage by *V. fischeri* of genes encoding RTX and TCT suggests that the activities these effectors encode may result in either a beneficial or a pathogenic outcome, depending on the host species or tissue location colonized.

Conclusions. Countless bacterial species interact with animals and plants in persistent associations that are often essential to their host's existence. Such symbiotic associations may share similarly derived colonization factors with pathogens. If we are to understand the unifying themes underlying these contrasting bacteria-host interactions, we must begin to use comparative genomic approaches with closely related pathogenic and beneficial microbial species. Such studies within the genus *Vibrio* may help reveal not only the evolutionary origins of host-targeted virulence factors, but also those mechanisms by which pathogens commonly associate with marine invertebrate reservoirs as benign or even beneficial symbionts.

We thank the following members of Integrated Genomics, Inc., for their work on the project: A. Lapidus and L. Chu for sequencing; E. Goltsman, N. Larson, and G. Pusch for assembly, and I. Anderson, V. Kapratl, A. Bhattacharyya, G. Reznik, M. D'Souza, T. Walunas, Y. Grechkin, and N. Kyrpides for initial annotation. This sequencing project was supported by a grant from the W. M. Keck Foundation.

^{††}Feliciano, B. & Ruby, E. G. (1999) *Abstr. Gen. Meeting Am. Soc. Microbiol.* **99**, 462 (abstr.).

