ABSTRACT

Title COMPARATIVE GENOMIC ANALYSIS OF *VIBRIO CHOLERAE* O31: CAPSULE, O-ANTIGEN, PATHOGENESIS AND GENOME

Yuansha Chen, Ph. D., 2006

Directed by Professor J. Glenn Morris, Program in Marine Estuarine Environmental Sciences

Vibrio cholerae is the causative agent of cholera. In order to understand the genetic basis underlying the emergence of novel epidemic strains of *V. cholerae*, the genetics of surface polysaccharide biogenesis, and the role of lateral gene transfer in the evolution of this species, we investigated.

NRT36S and A5 are both NAG-ST producing, cholera toxin negative, serogroup O31 *V. cholerae.* NRT36S is encapsulated and causes diarrhea when administered to volunteers; A5 is acapsular and does not colonize or cause illness in humans. The structure of the capsular (CPS) polysaccharide in NRT36S was determined by NMR. The gene cluster of CPS biogenesis was identified by transposon mutagenesis combined with whole genome sequencing data. The CPS gene cluster shared the same genetic locus as that of the O-antigen of lipopolysaccharide (LPS) biogenesis gene cluster. The LPS biogenesis regions in A5 were similar to NRT36S except that a 6.5 kb fragment in A5 replaced a 10 kb fragment in NRT36S in the middle of the LPS gene cluster.

The genome of NRT36S was sequenced to a draft containing 174 contigs plus the superintegron region. Besides confirming the existence of NAG-ST, we also identified

the genes for a type three secretion system (TTSS), a putative exotoxin, and two different RTX genes. Four pili systems were also identified. Therefore, the genome of non-O1 *Vibrio cholerae* NRT36S demonstrates the presence of pathogenic mechanisms that are distinct from O1 *V. cholerae.*

We conclude that lateral gene transfer plays a critical role in the emergence of new strains. The co-location of CPS and LPS could provide a mechanism for simultaneous emergence of new O and K antigens in a single strain. Our data also highlights the apparent mobility within the CPS/LPS region that would provide a basis for the large number of observed *V. cholerae* serogroups and the emergence of novel epidemic strains.

COMPARATIVE GENOMIC ANALYSIS OF *VIBRIO CHOLERAE* O31: CAPSULE, O-ANTIGEN, PATHOGENESIS AND GENOME

By

Yuansha Chen

Dissertation submitted to the faculty of the Graduate School of the

University of Maryland, College Park, in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

2006

Advisory Committee:
Professor J. Glenn Morris, Jr., Chair
Associate Professor O. Colin Stine
Associate Professor Judith A. Johnson
Professor Sam W. Joseph
Professor Shiladitya DasSarma
Assistant Professor Afsar Ali

PREFACE

My dissertation contains the full text of three manuscripts that will be published in scientific journals.

Chapter II contains the full text of the manuscript entitled "The capsule biogenesis genes are embedded in the LPS region in non-O1 *Vibrio cholerae* NRT36S", returned for revision, 2006. I wrote the first draft of the manuscript and circulated it to my co-authors. I incorporated their revisions through 12 drafts. I performed the transposon mutagenesis with the Km-2 transposon and advice from Afsar Ali. I picked and grew the mutants and stored them in -80ºC. I extracted genomic DNA from the translucent mutants. I mapped the position of the transposons in the mutants. I performed all of the sequencing and the subsequent analyses of the excised clones including reading, assembly and annotation. I combined data from 454 with my own data and filled gaps in the capsule/LPS region by designing primers, PCR amplifying and sequencing the products. Again, I performed the assembly, and annotation and comparison of the sequences to GenBank. I performed the serum killing assays under the direction of Judith Johnson. I conceived of the HPLC analysis and directed its completion by the Biopolymers Core. I performed the gel electrophoresis and immunoblotting. I prepared samples for the EM experiments and gave them to Pinaki Panigrahi for embedding and sectioning. Pinaki and I then looked at the samples together under EM and took pictures. I cloned *galE* gene and performed the complementation experiment. Peter Bystricky, Jacob Adeyeye and Allen Bush conducted the work towards resolving the capsule structure of *V. cholerae* NRT36S.

Chapter III contains the full text of the manuscript entitled "The genome of non-O1 *Vibrio cholerae* NRT36S demonstrates the presence of pathogenic mechanisms that are distinct from O1 *Vibrio cholerae*", returned for revision, 2006. I grew the bacteria and isolated the DNA. The genome was sequenced by the company 454 Life Science. It was automatically annotated by Gordon D. Pusch in the National Microbial Pathogen Data Resource. I compared the genomes, assembled the differences and performed further manual annotation. I compared the genomes with the program MUMMER; edited the results with Microsoft Excel and analyzed the difference. I compared the proteomes of the two strains with a stand alone Blast program. When using these programs, I built up the local databases. I wrote the first draft of the manuscript and incorporated edits and revisions from my co-authors through 10 drafts.

Chapter IV contains the full text of the manuscript "Genetic variation of capsule/lipopolysaccharide biogenesis in two serogroup O31 *Vibrio cholerae* isolates", in preparation, 2006. I conceived of the idea to compare the CPS/LPS region of these genomes. I developed primers and amplified across the region to find the differences. I sequenced, assembled and confirmed the assembly for the disparate region. I manually annotate the LPS region in NRT36S and A5. I retrieved data from GenBank for the phylogenetic trees. I analyzed the GC content in the CPS/LPS regions. I wrote the first draft of the manuscript and incorporated suggestions and comments from my co-authors J. Glenn Morris, Jr., O. Colin Stine and Judith A. Johnson.

DEDICATION

To my parents

ACKNOWLEDGEMENTS

I would like to thank Dr. Morris and Dr. Stine for their guidance and financial support provided throughout my graduate study and work towards the Ph. D. degree. I am grateful to Dr. Johnson for her supervision in conducting experiments and her help in making corrections to my dissertation. I also thank Dr. Ali for his guidance in the molecular genetic. I would like to thank Dr. DasSarma for his help in Bioinformatics. My thanks also go to Dr. Joseph for his suggestions. I also thank Dr. Stine's laboratory members for their help and friendship.

Several cooperators have contributed to the work presented in my dissertation. Dr. Alan Bush and his group helped solving the structure of capsule polysaccharide in NRT36S. Ross Overbeek and Gordon Pusch in the National Microbial Pathogen Data Resources (NMPDR) and Beenish Bhatia helped with annotating the genome of NRT36S. Dr. Pinaki help doing the EM. I would like to thank all of the cooperators for their effort.

I would like to thank my family for their love and encouragement. I thank my grandparents for their unconditional love. I thank my parents for having worked and striven for the best of their children. I also thank my brothers and sisters for their support and love through the years. I would also like to thank my two sons for the enjoyment they bring to me.

I am deeply grateful to my husband for his love, encouragement and help. He has been with me and supported me for every step I go forward. Thanks him for all the discussion and suggestions on my research.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION AND BACKGROUND

VIBRIO CHOLERAE

Microbiology and Ecology

Vibrio cholerae is the causative agent of the devastating epidemic disease cholera. It is a motile, curved gram-negative rod; capable of both respiratory and fermentative metabolism. There are several biochemical features that can distinguish *V. cholerae* from the Enterobacteriacean. It is oxidase positive; it can grow in nutrient broth with added NaCl; it does not produce gas from glucose, and it ferments sucrose to give a yellow colony on TCBS agar. *V. cholerae* is a heterogeneous species. Some serogroups cause epidemics of the disease cholera, others are responsible for intestinal infections in small outbreaks or in an individual; some cause extraintestinal infections, and the rest are considered harmless environmental isolates.

The natural habitat for *V. cholerae* is an estuarine environment, but it can persist in both marine and fresh water. It can exist in free living form, associated with phytoplankton and zooplankton, in biofilms, and even in a so call "viable but non-culturable (VBNC)" form. *V. cholerae* can be readily isolated from the natural coastal water. The ecology of it is not fully understood. But studies conducted by Colwell's group in Chesapeake Bay, Maryland identified several environment parameters that affect the distribution *V. cholerae* in its natural reservoir (1-3). The two parameters affect the distribution of *V. cholerae* the most are temperature and salinity. When water temperatures were warmer than 19°C and the salinity was between 2-8 ppt, about 80% of the samples collected were positive for *V. cholerae* (3). Only non-O1 *V. cholerae* was cultured from these environmental samples. O1 *V. cholerae* was not isolated by

traditional culture method but was detected in about the same frequency, only when a direct fluorescent-antibody assay was used. That suggests O1 *V. cholerae* may exist in nature in a VBNC state. VBNC is a survival strategy for bacteria when they encounter stress conditions (4). Bacteria are dormant in this state. They become smaller in size; their metabolism is very slow; and they significantly reduce their macromolecules including lipid, carbohydrate, proteins and DNA.

Colwell et al's findings are consistent with the results of investigating the distribution of toxigenic (O1 and O139) *V. cholerae* in the aquatic environment around the cholera endemic area. In India and southern Bangladesh, cholera erupts seasonally, twice a year, once before, and once after the monsoon season (5). Toxigenic *V. cholerae* are rarely isolated by culture from coastal waters in these areas in between the outbreaks, and they are cultured only infrequently even during epidemic period (6). In the contrast, non-O1, non-O139 *V. cholerae* are isolated in culture from almost every sample collected from the same area (7). But using fluorescent monoclonal antibody against O1 serogroup, toxigenic O1 could be detected in the aquatic environment in the cholera epidemic area (6). Now, it is known that toxigenic *V. cholerae* exist as a VBNC form and was not picked up by traditional laboratory culture method (8). VBNC *V. cholerae* O1 recovers in fully virulence when inoculated into rabbit intestine (9).

An important survival mechanism for *V. cholerae* is to live in biofilms. A recent study by Alam et al. (7) surveyed the bay of Bengal for *V. cholerae*, using a combination of methods including direct fluorescent antibody count and fluorescent microscope. They found a large number of O1 *V. cholerae* in thin films of exopolysaccharides (biofilm). In

biofilms the cells are clumped and their shapes do not show the typical curved rod of the free living form. Survival is better in biofilms compared to free living (10-12).

Matz et al. found that embedding in biofilms was a strategy for *V. cholerae* to avoid protozoan grazing (13). In their study, they compared the amount of planktonic versus biofilm embedded *V. cholerae* grazed by protozoa and they found the planktonic *V. cholera* was quickly grazed by protozoa. But *V. cholerae* embedded in biofilm was protected from protozoan grazing. Matz et al. also found that there were two protecting factors: i) Biofilms provide a physical barrier for the grazers. ii) *V. cholerae* embedded in biofilms secretes a factor, under the control of quorum sensing, that is against protozoan predation. The nature of the factor secreted is yet to be investigated.

The current serotyping scheme of *V. cholerae* is based on the O-antigen of the lipopolysaccharide (LPS) with more than 200 serogroups reported (14). Only serogroups O1 and O139 of *V. cholerae* cause epidemic cholera and frequently carry cholera toxin. Because of the severe symptoms and high mortality of cholera, and also because its great tendency to spread and cause world-wild explosive outbreaks, the disease cholera and serogroup O1 and O139 *V. cholerae* have been studied extensively. There have been 7 pandemics of cholera recorded, with the fifth and sixth pandemics caused by O1 classical and the seventh pandemic caused by O1 El Tor, and the previous ones are believed also to be caused by O1. Besides these seven pandemics, there may be also an 8^{th} pandemic of cholera beginning, which started in the 1990's in India and Bangladesh; caused by a new serogroup O139, designated as O139 Bengal. O139 Bengal is very similar to O1 El Tor genetically except for the region encoding its surface polysaccharides (15-17). Other serogroups, including O37 Sudan, have sometimes caused cholera like disease in small

outbreaks. O37 Sudan was found to have arisen from O1 classical by switching its O-antigen (18). O1 and O139 frequently carry cholera toxin. Thus O1 and O139 are also referred to as toxigenic *V. cholerae*. Occasional non-toxigenic O1 isolates are found in the environment. O1 isolates can be further subdivided into serogroups Inaba and Ogawa and biotypes classical and El Tor.

Most of non-O1, non-O139 *V. cholerae* in the environment do not carry cholera toxin genes, but may carry other toxins and be harmful if ingested. An earlier survey to the Chesapeake Bay in Maryland isolated 65 isolates of non-O1 *V. cholerae*. About 87% of them displayed toxicity for Y-1 adrenal cells and majority of them were positive for fluid accumulation in ileal loops experiments, suggesting the presence of enterotoxins. Non-O1, non-O139 *V. cholerae* do occasionally cause diarrhea, and sometimes are also isolated from extraintestinal infections including blood, wounds, ear, and other sites (19).

Genetic Backbone

The genome of *V. cholerae* isolate N16961, serogroup O1, biotype El Tor, has been sequenced (20). The genome is 4,033,460 base pairs (bp). There are two chromosomes in *V. cholerae* with a total of 3885 genes identified. Most of the genes are similar to *Escherichia coli* genes, but 12.8% of the genes show highest similarity to the other *V. cholerae* genes and suggest recent duplication. Chromosome I harbors most of the genes essential for cell functions such as metabolism and pathogenesis. Genes encoding the O1 specific LPS are also on chromosome I. On chromosome II, the genes are mostly hypothetical proteins with unidentified functions. Because of the presence of many plasmid origin genes (20), chromosome II was suggested to be derived from a megaplasmid that was captured by an ancestral *Vibrio* species. Chromosome II harbors a

large integron (superintegron) which is about 125 kb and encodes 215 open reading frames, each one flanked by a *V. cholerae* repeat.

The genomes of *Vibrio vulnificus* and *Vibrio parahaemolyticus* were also sequenced (21, 22). Both of these two species also have two chromosomes. Chromosomes I in both *V. vulnificus* and *V. parahaemolyticus* have similar sizes to Chromosome I in *V. cholerae*. Chromosomes II in *V. vulnificus* and *V. parahaemolyticus* have similar sizes (1.9 mb), but they are bigger than the size of chromosome II in *V. cholerae* (1.0 mb). Comparative genomic analysis showed that *V. vulnificus* and *V. parahaemolyticus* have a higher degree of conservation in gene organization. *V. cholerae* has extensive genome rearrangement compared to *V. vulnificus* and *V. parahaemolyticus* (21, 22). Makino et al. compared genes between *V. cholerae* and *V. parahaemolyticus* (21, 22). There is 16.2% and 29.5% of genes specific to *V. parahaemolyticus* and *V. cholerae* respectively in the chromosome I. There is 41.9% and 56.8% of genes specific to *V. cholerae* and *V. parahaemolyticus* respectively in Chromosome II.

House-keeping genes, conserved in the genome, are used as indicators of the genetic relationship among different isolates of *V. cholerae*. Two techniques, multi locus sequencing typing (MLST) and multi locus enzyme electrophoresis (MLEE), have found *V. cholerae* to be a genetically diverse species (23-25). In contrast, the epidemic strains were found to be closely related, forming a tight clonal cluster in the phylogenetic tree, separate from the environmental isolates. Using MLEE, Beltran et al. studied the polymorphism of 17 enzymes in an extensive collection of 397 *V. cholerae* isolates, including the O1 classical, O1 El Tor pandemic strains, O139 epidemic strains, and environmental strains (23). They found most of the O1 epidemic strains and all O139

epidemic strains have the identical electrophoretic types (ETs), while some O1 epidemic strains differ to this ET in only one enzyme locus. The O37 isolates from Sudan epidemic also differed from the main pandemic ETs at a single locus. The environmental stains included in this study are diverse and form many clusters, separating from the epidemic clusters. This result was consistent with the findings by Wachsmuth's group at the CDC (26). It is also consistent with the finding from other groups that O139 may have arisen from an O1 El Tor strain by acquiring new LPS biogenesis region (15, 17, 27).

Sequence typing has been used to determine the genetic relationship among the *V. cholerae* isolates from the 6^{th}, 7^{th} pandemics and the US Gulf coast. Reeves et al. initially studied the polymorphism of one housekeeping gene *asd* and concluded that the 6^{th}, 7^{th} pandemic and US Gulf isolates arose from different clones (28). Later on, the same group, by studying 4 more housekeeping genes, concluded that the epidemic clones are actually very closely related (24).

There were also studies looking at the diversity among the O139 epidemic isolates. Farfan et al. looked at the six house-keeping genes in the O139 Bengal isolates and found the 29 isolates fall into three different groups and therefore concluded that they may derive from different ancestral clones (29). But since they did not include many isolates of the other serotypes or from other sources as reference, the variation among those O139 Bengal isolates may just represent the variation after the O139 epidemic strain arose and their conclusion that the origin is polyclonal is questionable. Garg et al. analyzed nine housekeeping loci in the 96 *V. cholerae* O139 isolates from the same hospital and identified 51 different sequence types (25). In summary, the genetic backbones in all the

epidemic strains are very similar, and may have separated from the environmental strains long time ago.

Virulence Factors

The pathogenesis mechanism in the pandemic *V. cholerae* has been investigated extensively. The major virulence factors are cholera toxin (CT) and toxin co-regulated pilus (TCP) (30-34). CT is encoded in the genome of a filamentous phage CTX∅, incorporated into the chromosome I of *V. cholerae* (32). CT consists of five B subunits and one A subunit (35, 36). During infection, B subunits bind to the GM1 ganglioside receptors in the host small intestinal mucosa; subunit A then travels into cells, increases the level of cAMP via G-protein, and subsequently interferes with the function of ion channel for Cl^-. This leads to net water flux into the lumen of the human intestine and causes diarrhea (37-39).

TCP is a type IV pilus and the major colonization factor for *V. cholerae*. It is also the receptor for CTX∅. Some strains only have TCP but lack CTX∅ (40). TCP is encoded in a 39.5 kb region of the *V. cholerae* genome, designated VPI-1 (30). All the strains that demonstrated the ability to cause epidemic cholera are positive for both CT and TCP. Besides CT and TCP, other toxins have been reported, including Ace, Zot, RTX, etc. (19).

The impact of non-O1 and non-O139 *V. cholerae* may have been overlooked. During cholera epidemics period, they are isolated from up to 13% of patients with cholera-like disease (41). In contrast to the epidemic strains, pathogenesis by non-O1, non-O139 *V. cholerae* is not well understood. There are different toxins reported and sometimes results are controversial from studies conducted with different isolates. In an

animal model study and a volunteer study, a heat stable enterotoxin (NAG-ST) was found to be a critical virulence factor to cause diarrhea by an O31 strain NRT36S (41). But NAG-ST along is not enough to cause disease since a similar O31 strain A5, also NAG-ST positive, is not pathogenic. The colonization factor that possibly made the difference in terms of pathogenesis between NRT36S and A5 was not identified.

SURFACE POLYSACCHARIDE

General Classification

Like other gram-negative bacteria, the surface of *V. cholerae* is coated with complex polysaccharides. Three major surface polysaccharides exist: lipopolysaccharide (LPS), which is covalently bound to the cell outer membrane by a lipid A anchor; a capsule, which normally has higher molecular weight and exists as moiety closely associated with the cell; and an exopolysaccharide, which is a loose slime outside the cell that forms an intercellular matrix in biofilms. Exopolysaccharide are expressed in the cells that display a rugose (wrinkle) phenotype. Rugose is discussed in the section of "phase variation". This exopolysaccharide is not with the scope of this study.

LPS is attached to the outer membrane in gram-negative bacteria and is also referred to as endotoxin. In epidemic *V. cholerae* O1, it is the outer most layer of the cell and has direct contact to the host environment. LPS was shown to adhere to the intestinal mucosa and may contribute to colonization during infection (42, 43). Infection with *V. cholerae* can result in protective immunity against re-infection with the same serogroup (44, 45). The protective immunity comes from both anti-bacterial and antitoxin responses with the most important immunity being the antibacterial immunity. The antibacterial antibodies are mostly absorbed by the LPS, strongly suggesting that the protective immunity mostly

comes from LPS (38). The current most effective vaccine is a combination of killed whole bacterial cells and purified toxin subunit B, to activate both vibriocidal and antitoxin antibodies.

LPS consists of three parts including lipid A, core and O side chain. The highly antigenic O side chain is often referred to as O antigen. The lipid A and core are conserved, while the O antigens are highly variable. The genetic location of the O antigen biogenesis gene clusters is conserved at a specific site of the genome in *V. cholerae*. O-antigen regions for 7 serogroups of *V. cholerae* have been identified and sequenced and they are all between the two genes *gmhD* and *rjg* (46-49). Examination of the gene clusters in different serogroups of *V. cholerae* found that the genetic composition between these flanking genes is highly variable. For example, the O antigen region in serogroup O1 has been replaced by an entirely different fragment in O139 and O37 and there is very little common sequence in this region shared by the three serogroups. Thus the epidemic strains have changed their O-antigens while maintaining their toxic genetic backbone, as we see from the rise of *V. cholerae* O139 Bengal and O37 Sudan from O1 progenitors. This potential for change probably extends to non-toxigenic strains as well. Sequence from O antigen gene clusters from 4 additional serogroups, O5, O8, O37, O108, by Aydanian et al., (46) found no conserved organizational motif, and only a very few genes shared by these 4 different clusters (46). In contrast, the O139 O antigen region is found to be very similar to O22. The organization and the content of the genes for O antigen biogenesis are highly conserved between these two serogroups with a small divergent region associated with an insertion sequence (49).

The second major surface antigen, the capsule, is composed of high molecular weight polysaccharide and forms a dense thick coat outside of the bacterial cells. How the capsule attaches to the cell membrane is not well understood. The capsule helps the bacteria to escape the host immune system by masking the LPS and hence avoid triggering the non-specific defense system. The capsule also protects the bacteria from complement-mediated phagocytosis and complement-mediated killing. Therefore capsule is a virulence factor itself; its virulence role is essential for pathogenesis in non-O1 *V. cholerae* (50) and some pathogenic bacteria like *Streptococcus pneumoniae* (51) and *V. vulnificus* (52, 53). Encapsulated pathogens can become invasive and cause septicemia due to increased resistance to complement-mediated killing. The capsule is a critical virulence factor in extraintestinal infections caused by *V. vulnificus* and non-O1 *V. cholerae* (41, 50, 54, 55). A case report of *V. cholerae* O139 that caused septicemia in immuno-compromised patient raises concerns for vaccine development with encapsulated *V. cholerae* such as O130 strains (17, 56).

Capsule Biogenesis

Capsule biogenesis has been studied in many species including *E. coli*, *Salmonella typhi*, *Neisseria meningitides*, *Haemophilus influenzae*, *Klebsiella pneumoniae* and *Streptococcus pneumoniae*. Among them, capsules in *E. coli* are studied most extensively and serve as model systems for the other species. Capsule biogenesis in the other species is more or less similar to *E. coli*. The capsule types in *E. coli* display striking diversities. *E. coli* has more than 80 capsule types and they are classified into 4 groups according to a scheme proposed by Whitfield's group (57). This classification scheme considers both the thermo-stability and genetic properties.

Groups 1 and group 4 capsules are thermo stable and share many common features genetically. In group 1 and group 4, a proportion of the K (capsule) antigen is attached to the oligosaccharide core and is termed K_{lps}, while the rest of the K-antigen is not attached and forms a capsule (58). In most cases, there is also an additional O-antigen attached to the same core in these two groups. Occasionally, there is no other O-antigen except the K_{lps} present in *E. coli* that express group 4 capsules. The capsule gene clusters for groups 1 and 4 reside in a region define where the O antigen cluster resides for groups 2 and 3 of *E. coli*. It maps closely to the *his* operon at 45 min in the genome. Group 4 capsules can co-express with colanic acid while group 1 capsules do not. Colanic acid is an exopolysaccharide expressed by many *E. coli* (59). The biosynthesis gene cluster of colonic acid is close to the *his* operon; it is lost and replaced by the group 1 capsule gene cluster in some *E. coli*. Besides genes at the 45 min locus, there are other genes related to the groups 1 and 4 capsule synthesis at the 22 min position. Genes in the 22 min locus are redundant for group 1 capsule biogenesis, but they are necessary for the group 4 capsule biogenesis.

Groups 2 and 3 capsules are thermo-labile. Their gene clusters in *E. coli* occupy the same location near the *ser* operon in the genome mapped to 68 min. The capsule clusters in these two groups are very similar in terms of their organization; both consist of 3 regions, with the two conserved regions flanking the serogroup specific region. The genes in the two conserved regions are responsible for the polysaccharide process and exportation. In the serogroup specific region, the gene content is very different in group 2 and group 3 (60). First the genes of the serogroup specific region encode enzymes that synthesize different sugar subunits and repeating units. Second, groups 2 and 3 capsules

also differ in their thermo regulatory properties. Group 2 capsules are not expressed in low temperatures below 20°C (61, 62). Group 3 capsules are express under all growth conditions.

There are limited strategies employed by the bacteria to synthesize polysaccharides. The biosynthesis pathways are found to be conserved. Whitfield et al. gave several excellent reviews on the capsule biogenesis in *E. coli* (57, 63, 64). Although there are extensive studies on the capsule biogenesis pathways, a lot still remain unknown or unclear, partly because the capsule biosynthesis is a complicated suite of the functions of multiple proteins with complex interactions.

Capsules can consist of either a homopolymer or a heteropolymer. The synthesis pathways then differ by the nature of the polymer. The sugars constitute the repeating unit and the polysaccharides are linked and polymerized. The capsule then needs to be exported across the cell envelop, and to the cell surface.

Capsules in groups 1 and 4 consist of heteropolymers. Individual sugars are added to a lipid carrier in the cytoplasm membrane to compose an oligo sugar subunit, by specific glycosyltransferases. The repeating oligo sugar units are then translocated to the periplasmic by a translocase *wzx*, and are polymerized by a polymerase *wzy*. The exportation of capsule polymer involves gene products of *wza*, *wzb* and *wzc*. The current model favors the forming of a multiple protein complex spanning the entire cell envelop. The requirement for *wza*, *wzb* and *wzc* is a common character shared by groups 1 and 4 capsules and differentiates them from groups 2 and 3 capsule. In group 1, *wza*, *wzb* and *wzc* are located together with the rest of the capsule genes near the *his* operon. But in

group 4, *wza*, *wzb* and *wzc* are at the 22min locus, separated from the rest of the capsule genes.

Groups 2 and 3 capsules in *E. coli* employ a different synthesis strategy to groups 1 and 4. *Wzx* and *wzy* genes are absent from groups 2 and 3 capsule gene clusters. Instead, groups 2 and 3 depend on an ABC transporter system to carry the polymerized polysaccharide across the inner membrane. Chain elongation in these two groups of capsule is possibly processive by the glycosyltransferase activity in the cytoplasmic site of the inner membrane (57). The polymer is then transported to the periplasmic area by the ABC transporter. The ABC transporter consists of two proteins, Wzm and Wzt. Wzm is a membrane protein while Wzt is an ATPase. The capsule polysaccharide is then exported to the outer surface of the bacteria. The current model suggests this last step depends on the function of two additional gene products, *kspE* and *kspD*.

Capsules have been studied in other pathogenic bacteria too. There have been more than 90 capsule types identified for *S. pneumoniae* (65). Capsule is an essential virulence factor in *S. pneumoniae*. The classical experiments in the study of its virulence ultimately lead to the discovery that genetic material is DNA (51). The genetic loci of all 90 capsule types of *S. pneumoniae* had been identified and sequenced (66, 67). With one exception, all loci are in the same location of the genome, between genes *dexB* and *aliA*. The organization of the capsule clusters in *S. pneumonia* is similar to the group 2 capsule in *E. coli*, with conserved regions flanking the serogroup specific region. The exception in *S. pneumonia* is type 37. The genes to synthesize type 37 capsule is located far away from *dexB* and *aliA*. But interestingly, there is a silent copy of capsule gene cluster between *dexB* an *aliA* for pneumococci that expresses type 37 capsule.

Capsule gene clusters in both *Haemophilus influenzae* type b and *Neisseria meningitidis* group B are organized remarkably similar to group 2 capsules in *E. coli*. But something very unique is found in *N. meningitidis*. The capsule biogenesis region is organized in 5 sectors in group B *N. meningitides* (68) (69, 70). Similar to group 2 *E. coli*, the polysaccharide processing and transportation genes are conserved and flanking the capsule type specific genes. The two regions in the middle, regions A and D are type specific. Interestingly, region D is related to lipooligosaccharide synthesis and mutations in this region result in a truncated lipooligosaccharide.

Capsules of *Vibrio* Species

There are limited studies on the capsule biogenesis in the genus *Vibrio*. Capsules are important virulence factors in *V. vulnificus*, *V. parahaemolyticus* and non-O1 *V. cholerae*. There are many capsule types in *V. vulnificus*. Wright et al. identified a genetic locus for capsule biogenesis in *V. vulnificus* M06-24 (71). The presence of *wza* gene in this locus suggests it is similar to group 1 capsule in *E. coli*. Smith el at also identified a capsule gene cluster in *V. vulnificus* 1003 (72). The capsule genes in 1003 include a *wzx* and a *wzy*, suggesting it is also a group 1 capsule. In both of these cases, the capsule gene clusters are not complete and the understandings to capsule biogenesis in *V. vulnificus* are impaired.

More than 85% of non-O1 *Vibrio cholerae* isolates have a capsule (73). The capsule is a critical virulence factor in extraintestinal infections. Johnson et al. found that the loss of capsule in a naturally encapsulated pathogenic strain *V. cholerae* NRT36S confers less virulence in mouse models (50). However, the pandemic strain that most associated with cholera, serogroup O1, lacks a capsule and only has an LPS. This raises

an interesting question whether the loss of function for the capsule increases virulence in diarrhea disease.

The newly emerged epidemic strain *V. cholerae* O139 has a capsule. But O139 capsule is an O-antigen capsule, in which the O-antigen and the capsule share the same repeating unit. One repeating unit attaches to the LPS core and behaves as an O-antigen, the same repeating units also polymerizes but does not seem to attach to the core, and hence behaves as a capsule (74-76). Genetically, the gene cluster for capsule polysaccharide biogenesis in O139 replaces the gene clusters for the O-antigen gene cluster in O1 and resides in the same position on the genome, between the genes *gmhD* and *rjg* (49). Therefore, the same polysaccharide in *V. cholerae* O139 represents both capsule and O-antigen and resembles the group 4 *E. coli* capsule (64), yet is different from the most of the group 4 capsule because *V. cholerae* O139 does not have another O-antigen attached to the core.

In *V. cholerae* the atypical capsule of O139 is the only well studied capsule, other capsules and their biogenesis are essentially unknown. However, since the majority of the isolates of *V. cholerae* have capsules, the elucidation of capsule biogenesis in this species is a critical part of understanding the evolution of the surface polysaccharides, to understanding the internal relationship between the capsule and LPS, and furthermore to understanding the genetic basis underlying the emergence of novel epidemic strains. As indicated from the other species, surface polysaccharide evolves in a similar and module-like pattern as pathogenicity islands. Whether this is true in the highly diverse species of *Vibrio cholerae* needs to be answered.

PHASE VARIATION

Types of phase variation

Phase variation in bacteria refers to bacteria switching between two phenotypes, usually differences in colony morphology, and that switching is inheritable and reversible. Bacteria can switch between opaque and translucent, rough and smooth, or dry and moist. These morphology changes generally reflect the structural changes on the bacterial surface, including polysaccharide, pilus, outer membrane protein, or cell wall composition. Although the frequency of phase variation varies, it is higher than the frequency of point mutation. Phase variation is especially well recognized in pathogenic bacteria and is often associated with changes in virulence. The reason bacteria undergo phase variation is not clear. It may be a survival strategy for bacteria, perhaps to avoid host immune system, to alter attachment, and sometimes for improved interaction between bacteria to the environment.

Phase variation occurs when expression level changes in one or more proteins due to controlled modification of their DNA. It can be associated with changes in the expression of exopolysaccharide, capsule, LPS, outer membrane proteins, or granules in the cells. The changes could be absence or presence, higher or lower levels of expression, or structural changes.

The production of an exopolysaccharide can cause phase variation in bacteria. Colonies of *Vibrio* sp. can display rugose (wrinkle) or smooth phenotype (68). Both rugose and smooth variants of *V. cholerae* are virulent. Rugose phenotype is caused by an exopolysaccharide excreted by the cell. This exopolysaccharide forms slime and causes cells to clump together. As a result, colonies have a wrinkled surface and are

more resistant to biocides. Genes for exopolysaccharide production have been identified in *V. cholerae* (77-79). They are very similar to the exopolysaccharide genes in *V. parahaemolyticus*. These genes are under the control of *opaR* (80). Exopolysaccharide is important in biofilm formation. Under stressful conditions such as protozoan grazing, *V. cholerae* may switch to the rugose phenotype (13).

The status of encapsulation in bacteria also affects their colony morphology. Colonies are more opaque when encapsulated and translucent when not. Encapsulated non-O1 *V. cholerae* can spontaneously switch from opaque to translucent and reversibly switch back (50) and the opacity is correlated to the capsule amount when examined by electron microscope or by the combination of a stereoscope and transmitted oblique illumination technique (13). The similar kind of phase variation happens in *V. vulnificus* and *V. parahaemolyticus* (71, 72) (81). Phase variations due to the capsulation statues are also reported from many other bacteria including *Neisseria meningitides* (82, 83) and *Streptococcus pneumonia* (84).

Phase variation may result from the alteration in third surface polysaccharide, LPS. Although LPS is a structural part of the cell membrane in gram negative bacteria and bacteria lacking the entire structure are not viable, there are rough or semi rough strains of bacteria that only express the lipid A and core of the LPS and do not express or only express a very short O-side chain. The modifications in bacteria LPS side chain length affect their sensitivity to serum killing. Lipooligosaccharide (LOS) in *Campylobacter jejuni* mimics human ganglioside (GM) and causes Guillain-Barré syndrome. Linton et al. reported *C. jejuni* can expressed GM1 or GM2 like LOS and the two types of LOS are phase variants (85).

Fourth common cause of phase variation is because of the differential expression of proteinaceous structures on the cell surface. That includes various types of pilus and flagella. Pili act as adhesins for bacteria to interact with the environment or act as bridges between different cells. *E. coli* phase varies its type 1 pili (86). *S. enterica* encodes at least 11 pili, 4 of them were found to undergo phase variation (87). *N. gonorrhoeae* phase vary its pili. There are multiple pilus subunit genes in the genome and only one is expressed at a time. *N. gonorrhoeae* can phase vary a large number of distinct pili (88). Flagellum is another proteinaceous structure on bacteria cell surface. It can function for bacteria motility and adhesion. Flagella can undergo phase variation in *C. jejuni* and *H. pylori* (89, 90). As a result, motilities are affected in those bacteria undergo flagella phase variation.

The opacity change in colony morphology can also be caused by the phase variation of outer membrane proteins. *N. gonorrhoeae* can not only phase vary its pili, but a subset of surface proteins, known as opacity proteins (*opa*). There are multiple copies of *opa* in each *N. gonorrhoeae* strain and each *opa* gene phase varies (91, 92). Each *opa* gene contains two hyper-variable regions. Each gene can switch between on and off expression state, and can also change the content in its hyper-variable region. Therefore, the potential of *N. gonorrhoeae* to undergo phase variation is enormous. The S-layer protein in *Campylobacter fetus* is another example to go extensive phase variation. S-layer proteins form a thick matrix outside the cell of *C. fetus* and completely cover the LPS and hence protect the cell from antibody mediated opsonization. There are 6-9 copies of S-layer protein genes in the genome, clustering closely together. Only one gene is expressed in a cell. Recombination between different S-layer protein gene copies can

alter the protein expressed (93). Another membrane protein was found to associate with colony opacity in *V. cholerae* too. A 27 kDa outer membrane protein was found present in the translucent variant but absent from the more opaque variants of both O139 and O1 serogroups (13).

Another fifth factor reported to affect colony opacity and it is not on the cell surface. Poly-β-hydroxybutyrate forms granules in the cell and the amount of it present is positively correlated with the colony opacity in *V. cholerae* (13).

Molecular Mechanisms of Phase Variation

Bacteria employ different molecular mechanisms to produce phase variation. The most common theme identified is slipped-strand mispairing (SSM). This mechanism involves DNA sequence consisting of tandem repeats that have 1 to 7 nucleotides in the repeating unit. During DNA replication, mispairing can happen between mother strand and daughter strand of DNA, give rise to a reduced or increased number of tandem repeat in daughter DNA. If SSM happens in the promoter region, transcription can be turned on or off depends on the number of repeats. If SSM happens in the coding region, it could introduce a premature stop codon, the reading frame is shifted. In *H. influenzae*, the genes *hif* encodes fimbriae. There is a dinucleotide TA repeats in the promoter of *hif* genes, between the -10 and -35 positions. The repeat numbers varies from 9, 10 or 11. The expression level of *hif* genes change, depends on the number of repeats present (94). Phase variation by SSM mechanism has also been identified in other species including *H. pylori* (95), *N. gonorrhoeae* (96), *N. meningitides* (97), *Bordetella pertussis* (98) etc.

Homologous recombination is another common mechanism for phase variation. Genes that associated with this molecular mechanism of phase variation have multiple

copies, but only one copy is active and the others of them are silent because of the absence of promoters. When gene in the expression site recombines with a silent gene in the repertoire, it results in the expression of a different protein. This mechanism is well illustrated in the phase variation of *N. gonorrhoeae* pilin proteins (99, 100). The pilin proteins of a type IV pilus in *N. gonorrhoeae* are conserved for two thirds of the N terminal and vary at the C terminal. The pilin proteins are encoded by *pil* genes. There are several loci of the *pil* genes in the genome but only one locus is the expression site and termed *pilE*. The other loci are silent and termed *pilS*. *PilS* locus consist most of the variable genes. During DNA replication, genes in the *pilE* locus can recombined with genes in the *pilS* loci and express different pilus structure. S-layer proteins in *C. fetus* also phase varies under a similar mechanism (93).

There are other two mechanisms of phase variation that are based on genetic changes to the DNA. First, inverting the direction of the promoter for the operon encoding type 1 pilus, type 1 pili in *E. coli* phase vary (101). Second, site specific insertion and precise excision of the insertion element also cause phase variation. This mechanism maybe the cause of phase variation in capsule production in some *N. meningitides* (82).

Besides changing the DNA sequence itself, phase variation can also happen by epigenetic regulation. This mechanism involves modifying DNA by methylation. Methylation normally happens to the regulatory site. These regulatory sites are DNA sequence targets of the regulatory proteins directly regulate transcription. The binding of regulatory proteins to the target site either activate or repress the transcription. The methylation status decides the binding status of regulatory protein the regulatory element.

This in turn affects the expression of the regulated gene. It is found that the expression of pili in *S. entericca* Typhimurium and the P pili in *E. coli* are under this regulation (102, 103).

LATERAL GENE TRANSFER

Genomic Variation in *V. cholerae*

The genomes of *V. cholerae* exhibit substantial variation among conserved housekeeping genes, as indicated by multilocus enzyme electrophoresis (23) and multi locus sequencing typing (24, 25, 46). In addition, studies on the pathogenic isolates of *V. cholerae* from the 6th, 7th and a U.S. coast epidemic suggest that, although the clones were closely related, there was evidence for homologous recombination (24, 29, 104). Garg et al. analyzed nine housekeeping loci in the 96 *V. cholerae* O139 isolates from the same hospital and identified 51 different sequence types. Farfan et al. looked at the six house-keeping genes in the O139 Bengal isolates and found the 29 isolates fell into three different groups and therefore concluded that they may derive from different ancestral clones (29). By extending the study to include environmental isolates, Aydanian et al. found 157 different genotypes in the MLST study with seven house keeping genes in 167 isolates of *V. cholerae*. The number of alleles with two of more nucleotide difference is much higher than expected from point mutation alone, suggesting lateral gene transfer. In contrast, a study using MLST has found the variation in the genomes of pandemic *V. parahaemolyticus* to be very limited and there was no evidence for recombination (105). This paradox suggested that the high plasticity of genome may be a unique character of *V. cholerae*, or at least it is not a common theme among all *Vibrio*.

In addition to the variation seen in conserved housekeeping loci, the *V. cholerae* genome also has several highly variable regions where non-homologous genes may be present or absent. The regions include the *Vibrio* pathogenicity island (VPI) encoding TCP (31, 106), the cholera toxin gene cluster, the superintegron (20) and the surface polysaccharide biogenesis regions. The high plasticity of the *V. cholerae* genome are clear evidence that lateral transfer plays an important role in the evolution of this species. Some of the lateral gene transfer mechanisms are known in *V. cholerae*, others remain to be investigated. VPI is linked with a phage integrase, which suggests VPI may have been acquired into the genome by phage. Cholera toxin is encoded in a prophage CTX∅. CTX∅ has been isolated as a free phage particle from the supernatant of *V. cholerae* culture (32). Genes in the superintegron region may have been recruited by the integron structure. The mechanism for the plasticity in the surface polysaccharide biogenesis regions is not well understood.

Mechanism of Lateral Gene Transfer

Exogenous DNA can enter a bacterial cell via several paths including transduction, conjugation and transformation. All of these mechanisms have been shown to exist in *V. cholerae*. Transduction is conducted by bacterial phage. After phage transfect a host cell, it can have its genetic material replicated by the host replication system. The genetic material is then packed into its protein coat. The host cell then lyses and release phage particles. The phage particles are ready for a new round of transfection. There are two potential outcomes after a phage transfection. First, phage can accidentally pack the host DNA and carry with it into the new host. The genetic material carried from previous host

can recombined with its new host DNA by homologous recombination if their similarity permits. Second, the phage DNA can insert into the host DNA by lysogeny. Therefore, bacteria phage can introduce both its own genetic material and the genetic material from its previous host.

Phage can insert into the host genome by lysogeny. In this mechanism, phage plays a critical role in *V. cholerae* pathogenesis. Cholera toxin is encoded in a prophage CTX∅. CTX∅ has been isolated as a free phage particle from the supernatant of *V. cholerae* culture (32). The incorporation of CTX∅ into *V. cholerae* genome is beneficial for the spread and propagation of the bacteria and differentiate epidemic *V. cholerae* to the non-epidemic *V. cholerae*. The mobility of CTX∅ raises concerns for vaccine development, since certain vaccines of *V. cholerae* is a recombinant whole cell with mutated toxin genes (19). Besides CTX∅, other virulence genes in *V. cholerae* may also related to phage, including VPI.

Conjugation is the second major mechanism for exogenous DNA to enter a bacterial cell. Conjuative plasmids are responsible for conjugations to happen. DNA is mobile if they are in a plasmid. Plasmid can enter bacterial cell by conjugation or transformation. *V. cholerae* has a sex factor similar to F factor, a P plasmid (107). The conjugation plasmid F factor in *E. coli* can integrate into the genome stably and form a strain capable of high frequency recombination. Unlike F factor, P factor can not integrate into the genome stably. There are also a few plasmid belonging to the IncC incompatibility group, termed R factors, exist in *V. cholerae* (108, 109). These plasmids are responsible for some of the increase of antibiotic resistance in *V. cholerae*. Antibiotic resistances

genes introduced by plasmid have posed an enormous impact in health care and economic.

The third major mechanism for bacteria to take up exogenous DNA is natural transformation. One immediate benefit of this is probably to use the uptake DNA as nutrient (110). In order for DNA to transform into bacteria, cells need to enter a competent state. *V. cholerae* was considered not naturally competent up until recently. Schoolnik's group showed that *V. cholerae* could be induced into a competent state by the presence of chitin (111) plus other factors including overgrowth and nutrient limitation. The competence state requires the expression of a type IV pilus and a DNA binding protein as well as several regulatory cascades. Chitin is abundant in its natural aquatic environment of *V. cholerae*. Therefore *V. cholerae* is likely to be naturally competent on a regular basis.

After entering the cell, in order to be inherited stably, foreign DNA needs to be integrated into the host genome. This is achieved by homologous recombination or illegitimate recombination. Homologous recombination is reciprocal; it requires the points of recombination of the DNA being exchanged have high similarity. Homologous recombination depends on the activity of various Rec proteins. In *E. coli*, it was found that the presence of a short nucleotide sequence, *chi*, promotes recombination (112, 113). Similarly, a DNA-uptake signal sequence has been identified from *H. influenza,* the *Pasteurellaceae* (114-116), *N. gonorrhoeae* and *N. meningitides* (117, 118). These short sequences interact with both dsDNA exonuclease and RecBCD helicase. These sequences attenuate DNA exonuclease activity, thus avoiding degradation of the invading DNA. At the same time, these sequences provide a target for the RecBCD helicase and

promote recombination. DNA-uptake sequence can be identified by an exquisite experiment designed by Sourice et al. (115), in which the presence of such sequence in random clone will prevent the formation of rolling circle plasmid, which depends on the activity of host exonuclease to excise the linear replication form into circular monomers.

Illegitimate recombination, on the other hand, does not depend on the DNA similarity. Multiple copies of the same element can integrate into the genome by insertion. This procedure is normally depends on the activity of transposase, which maybe carried in phage DNA, insertion sequences, or transposons. The insertion of CTX∅ is a good example of an illegitimate recombination. Insertion sequences are frequently found in the region of polysaccharide biogenesis. For example, an insertion sequence (IS1358) is present in the O1, O139 and O22 serogroup defining LPS region. Another insertion sequence is present in the O37 LPS biogenesis region.

Integrons are gene capture systems that recombine exogenous gene cassettes and direct their expression. They are first described by Stokes et al. in 1989 and since then investigated extensively because its role in antimicrobial resistance (119). Integrons contain three key elements: i) an integrase gene, which recombines genes in a RecA independent manor; ii) a primary integration site *attI*; and iii) a promoter, which directs the transcription of a captured gene (120). Integrons can link to a mobile genetic element such as insertion sequences, transposons and conjugative plasmids. The mobilized integrons contribute significantly to the microbial antibiotic resistance. These integrons are classified into five groups according to their integrase gene sequence (120): Class1 integrons are associated with Tn402 transposons; Class 2 integrons are related to Tn7 transposons; Class 3 integrons may be located in a plasmid that is yet to be characterized.

Class 4 integrons in found as a subset of SXT element in *Vibrio cholerae* (121), confers trimethoprim resistance; Class 5 is in a plasmid carried transposon, found in *Vibrio salmonicida*, also confers trimethoprim resistance. The mobilized integrons generally contains one to several gene cassettes downstream of their *attI* sites. The gene cassettes all contain an imperfect repeat called *attC* site. The sequences in different *attC* sites are not very conserved. DNA recombination in integrons involves both the *attI* site and the *attC* site.

A superintegron was first discovered from *V. cholerae* (122). It is located on chromosome II. Superintegrons share same genetic structures as integrons. The integrase in the superintegron is related to the ones in the mobile integrons. However, superintegron are different from the mobile integrons, which normally capture one to a few gene cassettes in that, a superintegron captures many gene cassettes. The function of most of the gene encoded in the superintegron is not well characterized. The *attC* sites of different gene cassettes in a superintegron are highly similar. In *V. cholerae*, the *attC* sites are termed *Vibrio cholerae* repeats (VCRs). The VCRs are 120-124 base pairs long and share more than 80% identity to each other. Superintegrons have been discovered in many diverse bacteria species, including all of the *Vibrio* species examined, and some *Pseudomonas* sp., *Xanthomonadaceae* sp. etc. It is proposed that mobile integrons may have involved from superintegrons with the capture of integrase gene and the *attI* site by a mobile genetic element (120). This proposal has been supported by the experimental evidence that an integron mobilized on a plasmid can randomly recruit gene cassette from the superintegron of *V. cholerae* (123). The presence of a superintegron provides an

onsite gene capture system for *V. cholerae* and may have contributed significantly to the evolution of this species, as will be shown in chapter 3.

STUDY AIMS AND HYPOTHESIS

The emergence of the O139 and O37 groups of *V. cholerae* indicated that new epidemic strains emerge by changing their surface polysaccharides including LPS and CPS, but with a relatively stable toxigenic background. This phenomenon raises many questions including: 1) What is the genetic basis underneath the emergence of the new epidemic strain; 2) Will novel epidemic strains of *V. cholerae* emerge in the future? 3) What should we expect from the novel strain(s)? Does the capsule present a potential threat in the new strains?

The predominant epidemic strain of serogroup O1 does not have a capsule, therefore the impact and genetic basis of capsules has received little attention. O139 arose from O1 with a new O-antigen and a capsule, both consisting of the same repeating unit. The capsule of O139 comes as a surprise and raises a potential septicemia threat. The threat would have been greater if the capsule and the O-antigen were different entities, in which one antibody could not account for both antigens. The demand to understand the genetics of capsule biogenesis in *V. cholerae* is urgent and critical. This study was aimed at understanding, a) the genetic basis of capsule biogenesis in *V. cholerae*; b) the role lateral gene transfer plays in the emergence of new capsule and LPS genes, and in subsequently converting a strain to be pathogenic. I initiated our study with the following hypotheses: **A**: In the genome of non-O1 *V. cholerae*, there is a distinct region for capsule biogenesis, different from the lipopolysaccharide (LPS) biogenesis region.

B: Horizontal gene transfer plays a critical role in the evolution of *V. cholerae*, especially in maintaining diversity of surface polysaccharide.

CHAPTER II: THE CAPSULE BIOGENESIS GENES ARE EMBEDDED IN THE LPS REGION IN NON-O1 *VIBRIO CHOLERAE* NRT36S

ABSTRACT

V. cholerae serogroup O31 NRT36S, a human pathogen that produces a heat-stable enterotoxin (NAG-ST), is encapsulated. The structure of the capsular (CPS) polysaccharide was determined by high resolution NMR spectroscopy and shown to be a complex structure with four residues in the repeating subunit. The gene cluster of capsule biogenesis was identified by transposon mutagenesis combined with whole genome sequencing data (GenBank accession number DQ915177). The capsule gene cluster shared the same genetic locus as that of the O-antigen of lipopolysaccharide (LPS) biogenesis gene cluster. Other than *V. cholerae* O139, this is the first *V. cholerae* CPS for which a structure has been fully elucidated and the genetic locus responsible for biosynthesis identified. The co-location of CPS and LPS biosynthesis genes was unexpected, and would provide a mechanism for simultaneous emergence of new O and K antigens in a single strain.

INTRODUCTION

Vibrio cholerae has three forms of surface polysaccharide, although some strains do not express all three forms: a lipopolysaccharide (LPS) inserted in the outer membrane, a capsule composed of high molecular weight polysaccharide that forms a dense thick coat outside of the bacterial cells, and a loose slime-like exopolysaccharide. Unlike *V. cholerae* of serogroup O1, which causes cholera, most non-O1 isolates have capsular polysaccharide (CPS) in addition to lipopolysaccharide (LPS). The LPS of *V. cholerae* is a protective antigen for cholera (19, 38), with over 200 serogroups identified based on the

O-antigen of the LPS. The O-antigen biogenesis loci of 4 serogroups (O1, O139. O22, O37) have been sequenced and characterized, and have been found to reside between two genes, *gmhD* and *rjg*, in the genome (18, 41, 48, 49, 124). More than 85% of non-O1 *V. cholerae* isolates have a capsule that is critical for virulence in extraintestinal infections (125). However, in contrast to *E. coli*, in which extensive work has been done on capsule structure and genetics (with associated classification into groups by Whitfield and Roberts (64)), structures and the genetics of CPS in *V. cholerae* are poorly understood.

The one strain for which data on capsule structure and genetics are available is the newly emerged epidemic strain *V. cholerae* O139. This strain has a capsule that appears to have arisen from the replacement of the O1 antigen biosynthetic region with a new gene cluster in the genetic background of an O1 strain (41, 49), resulting in emergence of a strain to which the human population did not yet have immunity. The capsule in O139 is unusual in that it shares the same repeating subunit as the O-antigen (76, 126, 127). Therefore, the polysaccharide in O139 appears as both capsule and LPS and resembles the K_{LPS} in the group 4 *E. coli* capsule (64).

There are limited studies on the genetics of polysaccharide biogenesis for the genus *Vibrio*. In *V. vulnificus* the CPS is a primary virulence factor and hence has been the target of more intensive study (128). An operon including genes *wza*, *wzb* and *wzc* was identified as part of the CPS genes for *V. vulnificus* strain M06-24 (71, 129), consistent with the presence of a group 1 capsule. The genetic loci for CPS were also identified in another strain of *V. vulnificus* 1003 (72). A wzx/wzy system was present for polymerization and exporting the CPS. However, the genetic region responsible for LPS biosynthesis has not been identified in *V. vulnificus*.

The elucidation of capsule structure and biogenesis is critical to understanding the evolution of surface polysaccharide and the internal relationship between the capsule and LPS in this species. It also has clear implications for understanding the behavior of this species within human populations, as the ability to change these surface antigens to avoid host immunologic detection is a key feature underlying the ability of *V. cholerare* to cause pandemic disease. Here, we report the covalent structure and studies of the biogenesis of the capsule in *V. cholerae* NRT36S.

RESULTS AND DISCUSSION

Structure of CPS

Chromatographic results. Analysis of the polysaccharide hydrolysate by HPAEC-PAD (high performance anion exchange chromatography pulsed amperometric detector) indicates that it contains three major peaks in approximate molar ratios 1:2:1. From comparison of the retention times of these peaks and the standards run under identical conditions it was determined that the NRT36S CPS contains one galactose (Gal), two glucosamine (GlcN) and one rhamnose (Rha) residues. Carbohydrate analysis by GC of the trimethyl silyl methyl glycosides performed at the Complex Carbohydrate Research Center (CCRC) likewise indicates this same monosaccharide composition. The results of methylation analysis, also performed at CCRC, are given in Table 2.1. These results indicate that the major components of the hydrolyzed sample of methylated CPS sample represent 3-linked Rha, 4-linked GlcNAc, 4- and 6-linked GlcNAc and galactose which is 3-linked, 4-linked (and small percentage 3- and 4-linked).

NMR results. NMR spectra of the native polysaccharide are complex, showing a number of peaks in the anomeric region which are not in simple stoichiometric ratios.

Likewise, the acetyl methyl region (2.0 – 2.3 ppm) shows approximately 10 peaks not in simple ratios suggesting that the polysaccharide may be heterogeneously substituted with O-acetyl functions. Therefore the sample was treated with aqueous ammonium hydroxide, which is expected to cleave O-acetyl groups by mild base catalyzed hydrolysis. The NMR spectra of the resulting sample (de-O-acetyl polysaccharide) showed only two peaks in the acetyl methyl region (2.06 and 2.09 ppm) and a greatly simplified pattern in the anomeric region with four distinct signals in the C-H HSQC spectrum (4.5 – 5.5 ppm ^{1}H, 95-105 ppm ^{13}C). We show below that the peaks at 2.06 and 2.09 arise from N-acetyl groups.

Under the assumption that the HPAEC-PAD result indicates four sugar residues in the repeating subunit of the polysaccharide, the four signals were arbitrarily assigned identifying letters, A, B, C and D for the purpose of individual sugar identification using homonuclear ^{1}H spin correlation. Residue C is identified as rhamnose by the characteristic methyl resonance of the 6-deoxy sugar at 1.33 ppm in combination with the equatorial configuration of H2 indicated by its small homonuclear coupling constants. The anomeric configuration is identified as β - by large NOE (nuclear Overhauser) between H1, H3 and H5 as well as by $^{1}J_{CH}$ = 162 Hz. Residues B and D are identified as amino sugars by the characteristic chemical shift of C2 and as glucosamine by homonuclear coupling values. Residue B has the α-anomeric configuration as indicated by small J_{H1-H2} and by $^{1}J_{CH}$ = 172 Hz while residue D has the β- configuration as indicated by large J_{H1-H2} and by $^{1}J_{CH}$ = 158 Hz. The fourth residue, A, is identified as α - galactose on the basis of the chromatographic data, small J_{H1-H2} and $^{1}J_{CH}$ = 168 Hz. The central part of the HSQC (heteronuclear single quantum coherence) spectrum of the de-

O-acetyl polysaccharide is shown in Figure 2.1 and the complete ^{1}H and ^{13}C resonance assignments are given in Table 2.2. The glycosidic linkages between the four residues were determined from HMBC (heteronuclear multiple bond coherence) and NOESY (nuclear Overhauser spectroscopy) data as indicated in Table 2.3. While no HMBC peaks could be observed for the D-B linkage, the NOE data clearly indicate a β-1-6 linkage and the downfield chemical shift of B-C6 at 68.9 ppm confirms this linkage assignment. The proposed structure of the tetrasaccharide repeating unit is given in Figure 2.2.

Having determined the structure of the sugar backbone, we turned to the acetylated forms of the polysaccharide. While interpretation of their complex spectra was difficult, it was possible given the basic sugar structure. Base hydrolysis milder than that required to produce the de-O-acetyl polysaccharide yielded a sample with NMR spectra having peaks in stoichiometric ratios (mono-O-acetyl polysaccharide). The single O-acetyl group is assigned to the 2-position of rhamnose (C), a position resistant to base hydrolysis due to the absence of a neighboring hydroxyl function. A complete set of homonuclear and heteronuclear NMR spectra of a sample of this form of the polysaccharide in H_2O solution allowed assignment of signals of the amide protons of residues B and D confirming that they are N-acetyl amino sugars. The complete assignment of the NMR spectrum of the mono-O-acetyl polysaccharide is given in Table 2.4.

The acetate methyl region of the NMR spectrum of the native, untreated, polysaccharide shows a number of peaks including those assignable to the amides of residue B and D and of the 2-O-acetyl group of residue C along with smaller peaks

indicating partial O-acetylation at other positions. Using HMBC spectra, it was possible to correlate methyl proton signals, through carbonyl carbon resonances, to sugar ring protons indicating positions of acetyl substitution. In addition to the 2-O-acetyl of residue C, it was possible to identify an O-acetyl group on the 3-position of residue B to the extent of about 50% as indicated in Table 2.5. Anomeric resonances of residue A were split into three peaks in the native polysaccharide suggesting partial O-acetylation of that residue but the exact positions could not be definitively assigned.

In considering our interpretation of the structural data, it should be pointed out that the methylation data (Table 2.1) are not entirely consistent with the NMR data and the structure proposed in Figure 2.1. While the methylation data indicating 3-substituted rhamnose, 4- and 4,6- substituted glucosamine are consistent with the proposed structure, the latter implies terminal galactose. The methylation data indicate 3 and 4 substituted galactose and we are not able to identify any substituents expected to survive the strongly basic conditions of per-O-methylation. The NMR data can be used to rule out any common C,H substituents such as pyruvate or lactate and ^{31}P spin-echo-difference spectra showed no indication of phosphates. Sulfate analyses were likewise negative. At this point, the best explanation we can offer is undermethylation resulting from some steric effects in this novel structure.

The structure proposed here for the NRT36S capsule repeating subunit is very similar to that reported for the *V. cholerae* O6 lipopolysaccharide (130). The polysaccharide backbones are identical but with stoichiometric O-acetylation at both C2 of rhamnose and at C3 of α-GlcNAc and the side chain is α-GlcA for this latter structure. While no genetic data are available for this O6 strain, we predict considerable homology

of the O-antigen clusters suggesting the possibility of lateral transfer between these two strains.

Transposon mutagenesis and mutant selection

The conjugations between wild type *Vibrio cholerae* NRT36S and donor strain *E. coli* S17λpir/putKm-2 generated 20,615 mutants of NRT36S, each carrying a single copy of the transposon Km-2 in its genome. Among these mutants, 411 colonies displayed a translucent phenotype on LB agar. This phenotype suggests that genes involved in capsule biogenesis have been disrupted by the transposon (50).

DNA analysis of mutants

Genomic DNA was isolated from the translucent mutants and analyzed by inverted PCR and sequencing, identifying 13 unique insertion sites in 11 genes. Since NRT36S can also undergo spontaneous phase variation between translucent and opaque colony morphologies, isolates with the 13 insertions were tested for complement resistance (50). Nine insertions in eight genes were sensitive to serum killing and showed no reversion to the opaque morphology (Table 2.6). Isolates with insertions in the other 3 genes reverted to opaque colonies and were resistant to serum killing and therefore were excluded from further analysis. Only four of the stable mutant genes related to sugar modification and processing and were considered as putative structural genes for the biogenesis of the NRT36S capsule. The function of the other genes was unclear. One of the putative structural genes had a homolog in the fully sequenced genome of *V. cholerae* N16961, a serogroup O1 pandemic strain. Orf23, a homolog of VC0262, an UDP-glucose 4-epimerase (*galE*) was disrupted in translucent colony TR3. TR3 was restored to opaque phenotype and resistant to serum killing when complemented with gal*E* gene. The other

three structural genes identified by transposon mutagenesis did not have homolog in the genome of *V. cholerae* N16961. In translucent colony TR17, a glycosyltransferase gene was disrupted and a rhamnosyltransferase gene was disrupted in both TR43 and TR287. An ABC transporter system integral membrane protein gene *wzm* was disrupted in TR296. The VC0262 homolog and the 3 other genes are typical of genes commonly found in polysaccharide biogenesis.

Immuno blotting and size exclusion chromatography

Size exclusion chromatography (SEC) data indicated that the molecular weight of the NRT36S capsule is greater than 670k Dalton (Figure 2.3A). The antiserum raised against the whole cell of *V. cholerae* NRT36S did not detect anything close to that molecular weight of CPS in the immuno blot (Figure 2.4), indicating that the antibodies did not react with the capsule. This result was consistent with the previous finding (131). Nevertheless, the antibodies detected some polysaccharides that form a ladder pattern in the molecular weight range of 20k to 40k Dalton (Figure 2.3B). We believed these were the LPS. Interestingly, the amount of reactive LPS to the antibody was reduced in mutants TR3, TR17, and TR296. Analysis of the capsule preps by SEC from the mutants showed that three of the mutants, TR3, Tr287 and TR43 had completely lost the high molecular weight peaks corresponding to the capsule while in TR17 and TR296, the amount of capsule was significantly reduced (Figure 2.3).

Electron microscopy

We evaluated thin sections of wild type *V. cholerae* NRT36S and several translucent mutants stained with polycationic ferritin by electron microscopy (EM). Representative profiles are shown in Figure 2.5. As seen before ((50), NRT36S

displayed a heavy, complete capsule surrounding the cell. TR3 did not have a complete capsule, but had some patches of capsule materials. Both of TR17 and TR296 had a much thinner capsule compared to opaque NRT36S. EM pictures for all three mutants were consistent with the amounts of capsule observed by SEC (Figure 2.3B).

Sequencing of the *V. cholerae* NRT36S genome

V. cholerae NRT36S genome was sequenced by the company 454 Life Sciences. The sequencing runs have generated 1,082,967 reads and output 104,531,256 bps of sequences. The estimated coverage depth was 26X. The draft genome consisted of 184 contigs with total length of 3.9 million base pairs. The average GC content for the draft genome was 47.5%. The draft genome was annotated. For the purpose of the discussion in this paper, only those features related to the polysaccharide biogenesis will be discussed.

Genetics of the polysaccharide biogenesis

O-antigen region. In previous studies, the O-antigen biogenesis genes for *V. cholerae* had been identified to cluster at one locus in the genome, between genes *gmhD* and *rjg* (18, 47-49). After aligning the contigs of the draft genome of *V. cholerae* NRT36S to the fully sequenced genome of *V. cholerae* N16961 (20), we found that 3 contigs (contigs 34, 19, 78) of NRT36S could partially align to the O-antigen region of N16961. Contig 34 contains *gmhD* gene and contig 78 contains *rjg* gene. Contig 19 falls into the middle of contig 34 and 78. The sequence information from analyzing the transposon mutagenesis mutants was able to pick up two more contigs (contig 98 and 43) and connect them to contig 19. Therefore *gmhD* and *rjg* were separated by 5 contigs in the NRT36S genome. The gaps between these contigs were filled and we ratified this

region, between *gmhD* and *rjg*, as the putative O-antigen biogenesis region (Figure 2.6) for *V. cholerae* NRT36S. The sequence between *gmhD* and *rig* was deposited into GenBank (accession number DQ915177).

CPS region. We located the capsule biogenesis genes identified by transposon mutagenesis in the NRT36S genome (Table 2.6). To our surprise, the 4 putative capsule structural genes identified by transposon mutagenesis were all located between the genes *gmhD* and *rjg*, the region considered to encode O-antigen biogenesis (Figure 2.6.). Theses four genes were knocked out by at least one of 5 independent transposition events and caused the translucent phenotype associated with the loss of the capsule in each case. Therefore, we believe that the O-antigen biogenesis region in *V. cholerae* NRT36S is also the capsule biogenesis region.

Global features. The locus of CPS/O-antigen was 49,916 base pairs in length between genes *gmhD* and *rjg*. There were 46 open reading frames (orf) (Figure 2.6). The annotation for each orf is listed along with a match from GenBank and its percent amino acid identity/positive, the species and the E-value (Table 2.7). Twelve genes were glycosyltransferases, 16 genes were recognized as pathway genes for synthesis of the nucleotide sugar precursor for external polysaccharide, and 6 other genes were recognized as polysaccharide processing and translocation genes. The function of the other 12 genes was unknown. A JUMPstart site (132) was located just downstream to the *gmhD* gene. The GC content of this region was 41.2%, lower than the 47.5% GC content of the genome. The disruption by transposon in orf5 (*wzm*), an ABC transporter gene, orf8, a glycosyltransferase gene, orf23, an UDP-glucose-epimerase, (*galE*) and orf43, a rhamnosyltransferase gene had caused the non-encapsulation of NRT36S (Figure 2.3).

The complementation by *galE* gene reverted the translucent mutant TR3 to opaque phenotype and the complemented colonies were resistant to serum killing.

Glycosyltransferases. There were 12 glycosyltransferase genes identified. The precise function of most of them remained to be elucidated. Orf45 (*wecA*) was an undecaprenylphosphate N-acetylglucosamine 1-phosphate transferase gene. WecA was putatively the initial transferase to catalyze the transfer of N-acetylglucosamine 1-phosphate to UndP in the capsule polysaccharide synthesis.

Synthesis genes. The structural data for *V. cholerae* NRT36S indicate that the capsule contains, one residue each of rhamnose and galactose and two N-acetyl-glucosamine residues, genes for whose synthesis are present in the CPS region. There are two sets of genes that are almost identical (orf1-4, and orf33-36) for L-rhamnose synthesis; they were *rmlB*, *rmlA*, *rmlC*, and *rmlD* in the order. L-rhamnose is commonly present in bacterial polysaccharides and the genes to synthesize it are normally clustered (133). Orf43 may be the rhamnosyltransferase to catalyze the addition of rhamnose to the CPS backbone. The disruption orf43 by the transposon both resulted in the loss of the capsule. Orf23 was gene *galE*; its product UDP-glucose 4-epimerase catalyzed the conversion of UDP-glucose to UDP-galactose. The disruption of *galE* gene in mutant TR3 caused the loss of the capsule. Orf24 (*wbeW*) transfers galactose to the capsule complex. Orf 41, a sugar O-acetyltransferase homologue could be involved in the observed O-acetylation of the capsule, but this modification of bacterial polysaccharides is not well understood and other genes may be involved as well. Orf11, 12, 22, 37, 41 and 44 were also putative pathway genes for the synthesis of nucleotide sugar precursors, but their precise functions were not clear to us.

Translocation and processing genes. An ABC-2 type transporter system consisted of *wzm* and *wzt* were present in the CPS/O-antigen region. When *wzm* was disrupted by transposon mutagenesis, the mutant was non-capsulated (Table 2.6). Orf38 was predicted as a polysaccharide translocase gene *wzx*. Orf40 was predicted to have several transmembrane domains by the Dense Alignment Surface (DAS) program (134) and were assigned as putative *wzy*. Three genes *wza*, *wzb* and *wzc* were also present in the CPS/O-antigen region. The proteins Wza, Wzb and Wzc in *E. coli* formed a system that was involved in the exportation of *E. coli* group 1 capsular polysaccharides (135).

The disruption of an ABC transporter system integral membrane protein gene *wzm* had significantly reduced the amount of capsule (Figure 2.3) in our experiments and resulted in the translucent colonies that were susceptible to serum killing. Examination of the CPS region also revealed the existence of *wzt*, which is another component of the ABC transporter system. Our results suggest that the processing and translocation of the capsule in *V. cholerae* NRT36S involves the ABC transporter system. There was a recent report that an ABC transporter system was involved in the transportation of hetero-polysaccharides in the O-antigen of *E. coli* O52 (136). Our results may be another case where an ABC transporter system was involved in the transportation of a hetero-polysaccharide.

Sharing of the same region by CPS and LPS. The O-antigen genes had been identified for 4 serogroups in *V. cholerae*, including O1, O139, O22 and O37 (figure 2.7) (18, 47-49, 124). In these serogroups, the gene cluster for O-antigen biogenesis all resided between the genes *gmhD* and *rjg*. In our study of *V. cholerae* NRT36S genome, there was a gene cluster identified as the LPS core biogenesis region upstream of the

gmhD gene (data not shown). There was also another gene cluster in the genome that was identified as the rugose-associated exopolysaccharide biosynthesis region (data not shown). The homolog of these genes were recognized as exopolysaccharide genes that related to the rugose phenotype in *V. cholerae* O1 El Tor (77). Besides these regions, i. e., the LPS core genes and the rugose exopolysaccharide genes, there were no other significant gene clusters for polysaccharide synthesis in the genome of *V. cholerae* NRT36S. All of this evidence supported the conclusion that the CPS region, i.e. the region between *gmhD* and *rjg* genes, is indeed also the O-antigen gene cluster. The immuno blot showed that the LPS had been altered in the non-encapsulated mutants. That not only confirmed the sharing of the CPS and LPS region, but also indicated that some genes may be shared by the biogenesis of the two polysaccharide structures.

To our surprise CPS and O-antigen shared the same genetic locus in *V. cholerae* NRT36S. This differs from the organization of CPS and O-antigen gene clusters in *E. coli*. In *E. coli*, CPS gene clusters and the O-antigen gene clusters are different (64). The CPS genes from other gram-negative bacteria including *Haemolyticus influenza*, *Salmonella* typhi and *Neisseria meningitidis* have been cloned and characterized (69, 137-141). The organizations of most of them resemble *E. coli*. No LPS genes have been reported to embed inside the CPS gene cluster for these species except in *N. meningitides*. In *N. meningitides* group B, the mutations of two CPS biosynthesis genes (*synX* or *synC*) and another gene next to the CPS region, *galE* gene, were shown to affect the lipooligosaccharide structure (138, 141). The CPS and O-antigen in group 4 *E. coli* consist of the same molecule. This arrangement is also seen in *V. cholerae* O139 where CPS and O-antigen are encoded by the same genetic locus and have identical repeating

subunits. NRT36S is an O31 serogroup. Kondo's group found that the LPS of O31 in *V. cholerae* contains L-glycero-D-mannoheptose, glucose, fructose, galactose, glucosamine and an unknown amino sugar A2 (142). L-rhamnose was not found in the LPS, while our study found L-rhamnose in the CPS, suggested that CPS and LPS are two distinct structures in *V. cholerae* NRT36S. The antiserum against NRT36S detected LPS but not the capsule, again suggested that the O-antigen and CPS were two different entities. Our finding represents a new type of genetic organization of polysaccharide genes and raises a question of differential regulation of the genes for expression of capsule and O-antigen polysaccharides.

The genes for bacterial surface polysaccharide biogenesis were typically found in a cluster with an atypical GC content compared to the rest of the genome (143). It had been suggested that bacteria could convert to a new serogroup by acquiring a new O-antigen biogenesis region. There was abundant evidence that *V. cholerae* O139 arose from an O1 strain by receiving a new O-antigen gene cluster (15, 27, 41). The sharing of CPS and O-antigen in *V. cholerae*, as indicated in our findings, makes this region one interchangeable unit. It is possible that *V. cholerae* can acquire a new gene cluster and give rise to a new strain with an O-antigen and K-antigen at the same time, both unrecognizable to the host immune system. This, in turn, may be a key element in its ability to cause pandemic disease, permitting rapid emergence of new strains that can escape immunologic detection by host populations.

EXPERIMENTAL PROCEDURES

Bacterial strains and culture conditions

V. cholerae NRT36S is an isolate originally cultured from a Japanese patient with travelers' diarrhea. It is serogroup O31, cholera toxin (CT) negative and produces a heat stable enterotoxin NAG-ST (144). When fed to volunteers this strain caused diarrhea, including, in one patient, a 5.3-liter diarrhea purge (41). Wild type *V. cholerae* NRT36S produces a capsule (50). This strain is resistant to polymycin B but sensitive to kanamycin.

Transposon mini-Km2 was carried in the plasmid putKm-2 and maintained in the host strain *E. coli* S17λpir. Mini-Km2 was found to randomly transpose into the genome of the recipient strain with a single transposition. Mini-Km2 encodes a gene for kanamycin resistance. (145, 146)

Cultures were maintained in L broth with 15% glycerol at -70°C. Bacteria were cultured in the L broth or agar at 37°C unless otherwise stated. Appropriate antibiotics were added in concentrations: 50 mg/ml Kanamycin, 50 unit/µl polymycin B.

Isolation and purification of CPS

Frozen stock of NRT36S was streaked for isolation on L agar in 150-mm Petri dishes and incubated overnight at 37 ^{0}C. A single bacterial colony from the plate was inoculated into 10 ml of L-broth for 18 h of growth at 37 ^{0}C. One ml of the culture was then inoculated into one liter L-broth and incubated overnight at 37 ^{0}C. Bacterial cells from eight 1-liter batches of culture were pelleted at 10,000g and re-suspended in 120 ml of 0.5X phosphate-buffered saline pH 7.5 and shaken at 200 rpm on a rotary shaker for 2 h at room temperature. The bacterial suspensions were centrifuged to remove cell debris

at 16,000g for 20 minutes at 4 ^{0}C and the supernatant dialyzed with multiple changes of distilled water and concentrated two fold by ultra filtration (100,000-nominal-molecular-weight stirred cell; Amicon, Beverly, Mass). The retentates were then ultra centrifuged at 154,000g for 2h at 20 ^{0}C and the supernatants were removed and digested with RNase A (100 ug/ml) and DNase 1 (50 ug/ml plus 1mM $MgCl_2$) for 2h followed by a 3h digestion with proteinase K(250 ug/ml) and phenol-chloroform extraction. The aqueous layer was dialyzed as described above, and the resultant sample was lyophilized. Purity of the CPS was assessed by bicinchoninic acid protein assay (MicroBCA, Pierce Chemical Co., Rockford Ill.), and Limulus amoebocyte lysate assay (Sigma Chemical Co., St. Louis, Mo.) and by NMR.

HPAEC-PAD

Carbohydrate composition was determined by HPAEC on Dionex HPAEC. A 2 mg sample of polysaccharide was hydrolyzed in 1ml of 1 M trifluoroacetic acid (TFA) at 100°C for 3 hours. The sample was dried with nitrogen and redissolved in 200 μL of dH_2O and submitted to HPAEC-PAD analysis with a Dionex CarboPack PA-10 (4 mm) column and detected on ED40 pulsed amperometric detector. The mobile phase used was 15 mM NaOH at 1 ml/min. The retention times were matched with those of standard monosaccharides run under identical conditions.

Deacylation and preparation of samples for NMR

A sample of 15 mg of CPS in 0.6 ml D_2O was subjected to deacylation by 0.3 M ammonium hydroxide for 6 days at 25 C. The 1D ^{1}H spectrum was monitored daily and the sample was lyophilized, exchanged with 99.9% D_2O and made up with 0.6 ml pure D_2O (de-O-acetyl polysaccharide). A second sample was prepared with milder base

treatment in 0.03 M ammonium hydroxide for 5 days at 4 C (mono-O-acetyl polysaccharide). For detection of amide protons, a sample was subjected to deacylation (without D_2O exchange) and made up with 0.6 ml 90:10 (v/v) H_2O:D_2O.

Transposon mutagenesis

Conjugations were performed between *V. cholerae* NRT36S as the recipient strain and *E. coli* S17λpir/putKm2 as the donor strain. Ten µl overnight culture of *V. cholerae* NRT36S was spotted on LB agar and let dry, 10 µl overnight culture of S17λpir/putKm2 was then spotted on top. After overnight incubation at room temperature, the mixture was re-suspended in 1 ml of LB broth; 50 µl of the suspension was plated onto LB agar with kanamycin and polymycin B to select for *V. cholerae* mutants.

DNA analysis

DNA flanking the transposon in the mutants was amplified and sequenced by a modified inverse polymerase chain reaction (PCR) protocol (147). Genomic DNA was isolated with PrepMan^{TM} (Applied Biosystems) according to the manufacturer's instruction. Two µl genomic DNA of the above preparation was digested with 5 units of *Nla* III (New England Biolab) in a 20 µl reaction overnight followed by denaturing at 65°C for 15 minutes. Two µl of the digested DNA was self-ligated with 5 units of T4 DNA ligase (Invitrogen) in a 10 µl reaction. One micro liter of this solution was used as PCR template. Two primers were designed to anneal to the transposon mini-Km2, pointing outwards to amplify the flanking sequence of the mutant genomic DNA. The sequences of the primers are, L8 (reverse), GTACCGAGCTCGAATTCGGCCTAG; and L9 (forward), GGAGAAAACTCACCGAGGCAGTTC. PCR was performed in a 30 µl reaction containing 100 µM of each dNTP, 1.5 mM of $MgCl_2$, 1x PCR buffer

(Invitrogen) and 1 unit of Taq DNA polymerase (Invitrogen). PCR products were purified with the Multiscreen PCR plates (Millipore) and sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). The resulting fragments were separated and recorded in an ABI 3730x1 automatic sequencer (Applied Biosystems). DNA sequence was then analyzed by the PHRED and PHRAP software (148-150).

Complement resistance and Electron Microscopy

Translucent colonies were challenged with complement in human serum as previously described (50). Electron Microscopy (EM) was also performed as described (50).

Size Exclusion Chromatography (SEC)

Capsule preps were analyzed by SEC using a Beckman Coulter 32 Karat HPLC, with TSK gel column (JOSOHAAS; G3000SWxL; 10um; 30cmx7.5mm), and detected at 200nm wavelength. Purified NRT36S capsule was the same sample as for doing NMR. Capsule preps were prepared as followed: The amount of 10^9 cells was harvested into 0.5X PBS and shaken for 2 hours in a rotary shaker at 250rpm followed by centrifuge at 12000g for 20min. The supernatant was treated with Dnase I and Rnase, followed by protease. The supernatant was then extracted with phenol-chloroform and precipitated with ethanol. The pellet was re-suspended in water and 1/3 of the amount was loaded.

Immuno Blotting

Immuno blotting was performed as described (47). Circa $5x10^6$ bacterial cells were treated with protease. Washed whole cell lysates were run on 16% SDS-polyacrylamide gel and transferred to Immun-Blot PVDF membrane (BioRad, Hercules, CA). Blots were

blocked in PBS contains 3% non-fat dry milk and then incubated for 1h in 1:1000 rabbit antiserum specific for *V. cholerae* NRT36S. The blots were washed three times with PBS and incubated with alkaline phosphatase-conjugated goat anti-rabbit immunoglobin G (Sigma) at 1:10,000 in PBS for 1h. The blot was washed five times with PBS and developed with Western Blue colorimetric detection solution (Promega).

Sequencing of *V. cholerae* NRT36S genome

The genome of NRT36S was sequenced by the company 454 Life Science (454 Life Science, Branford, CT) (151). The contigs of the draft genome was compared and aligned to the fully sequenced genome of *V. cholerae* N16961 (20) by Blastn (152). Gaps between contigs were filled only for the capsule biogenesis region, which contained the genes identified by transposon mutagenesis. Primers were designed for PCR to amplify the fragments of the gaps. PCR products were then sequenced.

Sequence analysis and annotation

Open reading frames were predicted by the program GLIMMER (153) using the DasSarma Laboratory Autoannotation Pipeline (DLAP) (http://halo.umbi.umd.edu/. DasSarma et al., manuscript in preparation). The settings in Glimmer were as in default, with the minimum gene size to be 90 bps and overlapping to be less than 30 bps. BlastX program (152) was used for a similarity search against the protein database in NCBI. We also used Artemis (154) to edit and confirm the results of GLIMMER.

ACKNOWLEDGEMENT

We would like to thank Dr. Shiladitya DasSarma and Beenish Bhatia for the kindly helping in annotating the genome of *V. cholerae* NRT36S using their server and their script to automate the annotation process. We also thank the company 454 life science

for sequencing the genome of *V. cholerae* NRT36S. We acknowledge support by NSF Grant MCB—212702 and of NIH grant GM-60791.

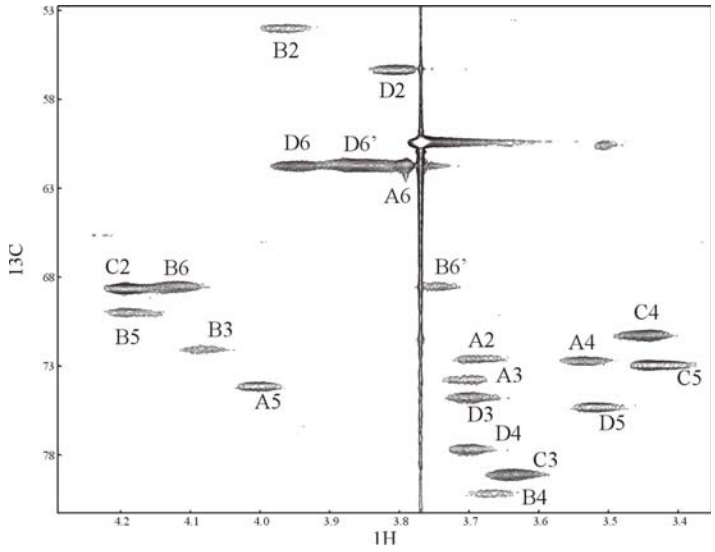


Figure 2.1. HSQC spectrum of the de-O-acetylated capsular polysaccharide from NT36S. Signals from the anomeric and methyl group region are not shown. The strong signal at 3.78, 60 ppm is a low molecular weight impurity.

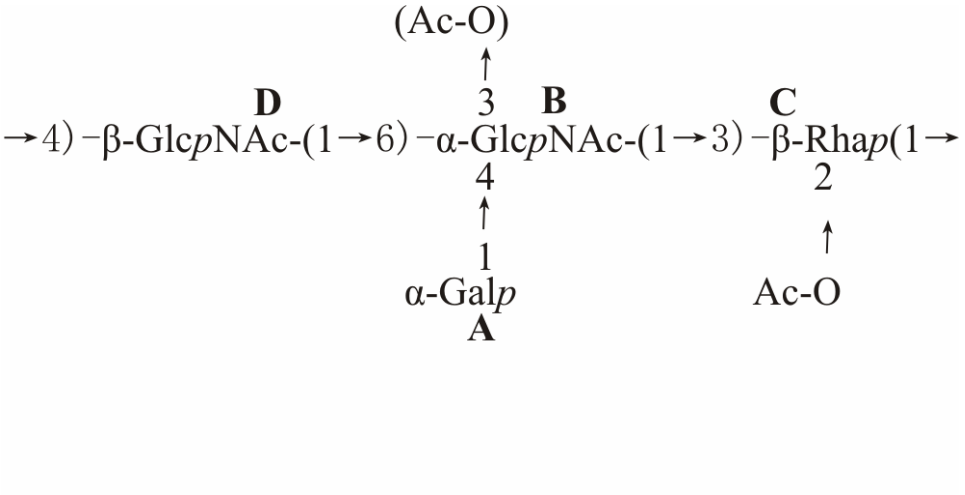


Figure 2.2. Proposed structure of NRT36S CPS repeating unit. Parentheses indicate partial O-acetylation of residue B at the 3-position.

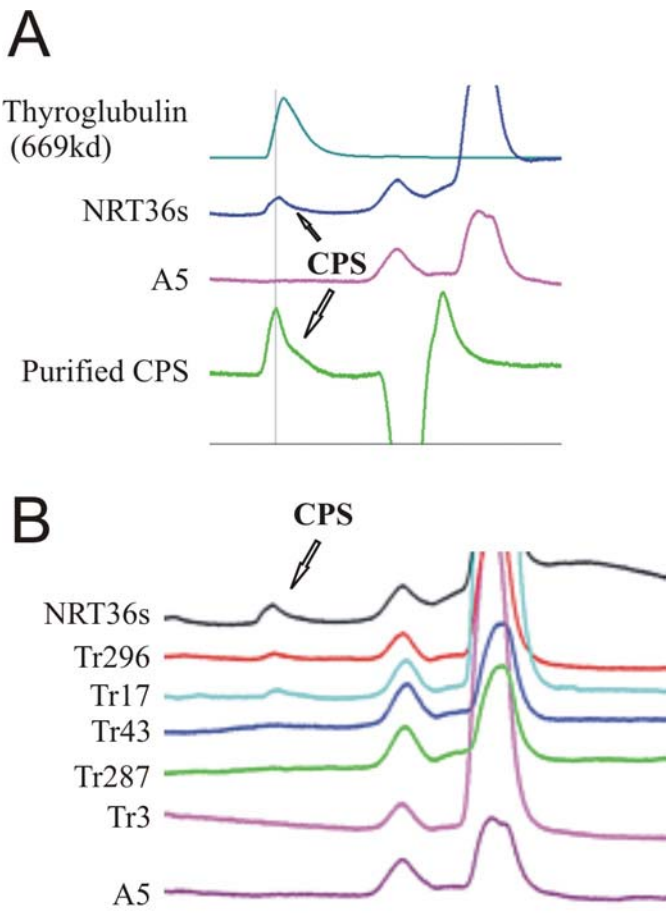


Figure 2.3. Size Exclusion Chromatography of the capsule prep.

A). Capsule prep from control (NRT36S and A5) compared to purified CPS of NRT36S. The size was estimated by thyroglobulin. NRT36S is O31 and encapsulated. A5 is also O31 but acapsular. The arrow indicates the peak of the capsule at about 13.2 minutes retention time, which corresponds to about 670k Dalton molecular weight. **B).** Capsule prep from mutants and control. Tr296, Tr17, Tr43, Tr287 and Tr3 are translucent mutants generated by transposon mutagenesis.

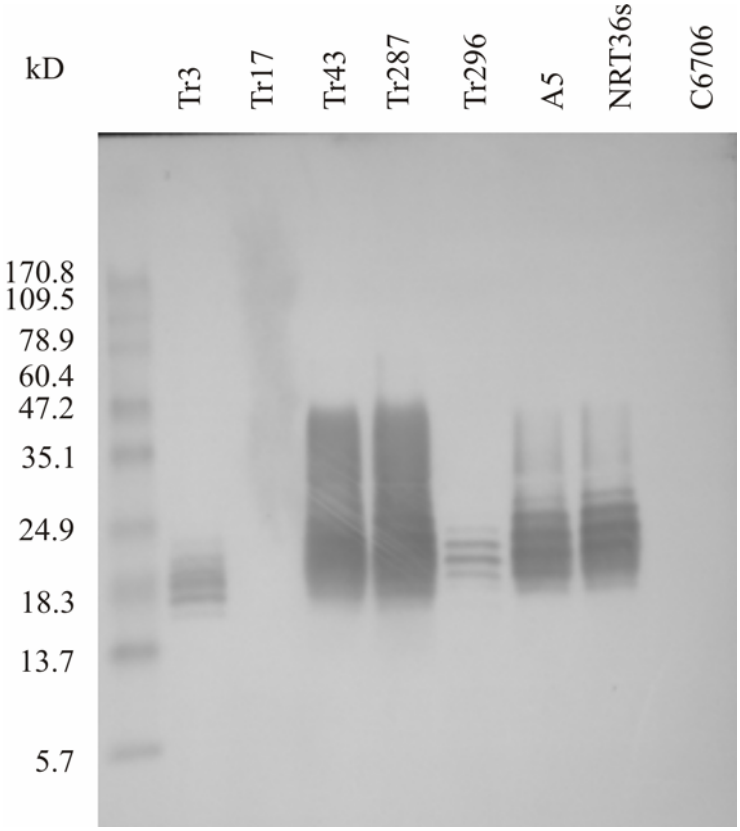


Figure 2.4. Immuno blotting. $5X10^6$ cells were washed with 0.5X phosphate buffer saline and treated with DNase I, RNase and protease. Washed whole cell lysates, separated on 16% SDS-polyacrylamide gel, were probed with antiserum specific for *V. cholerae* NRT36S. Tr3, Tr17, Tr43, Tr287, Tr296 are translucent mutants of NRT36S generated by transposon mutagenesis. A5 and NRT36S are both serogroup O31. A5 is natural acapsular variant of NRT36S. C6701 is a *V. cholerae* O1 isolate.

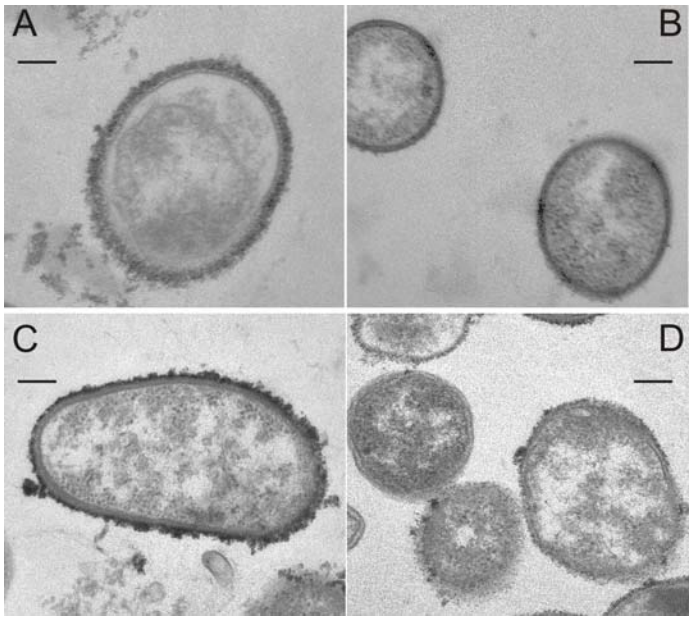


Figure 2.5. Thin sections of *V. cholerae* NRT36S and its translucent mutants stained with polycationic ferritin. (A) Wild type NRT36S; (B) Tr3; (C) Tr296; (D) Tr17. Bar, 200nm.

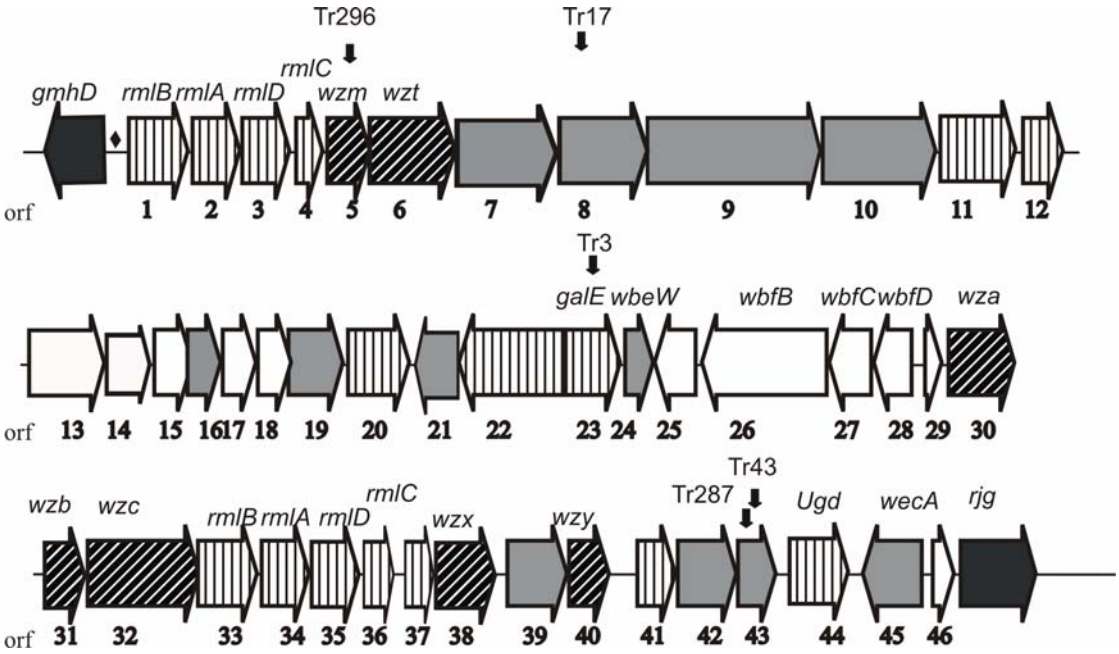


Figure 2.6. Map of the CPS/O-antigen region of *V. cholerae* NRT 36S. Diamond indicates JUMPstart site. Black arrows indicate transposon insertion sites in the translucent mutants. These insertions caused the loss or reduction of capsule polysaccharide. Design patterns of open reading frames indicate different classes of genes: vertical lines, pathway genes; diagonal lines, processing and transportation genes; grey box, glycosyltransferase; white box, functions not clear.

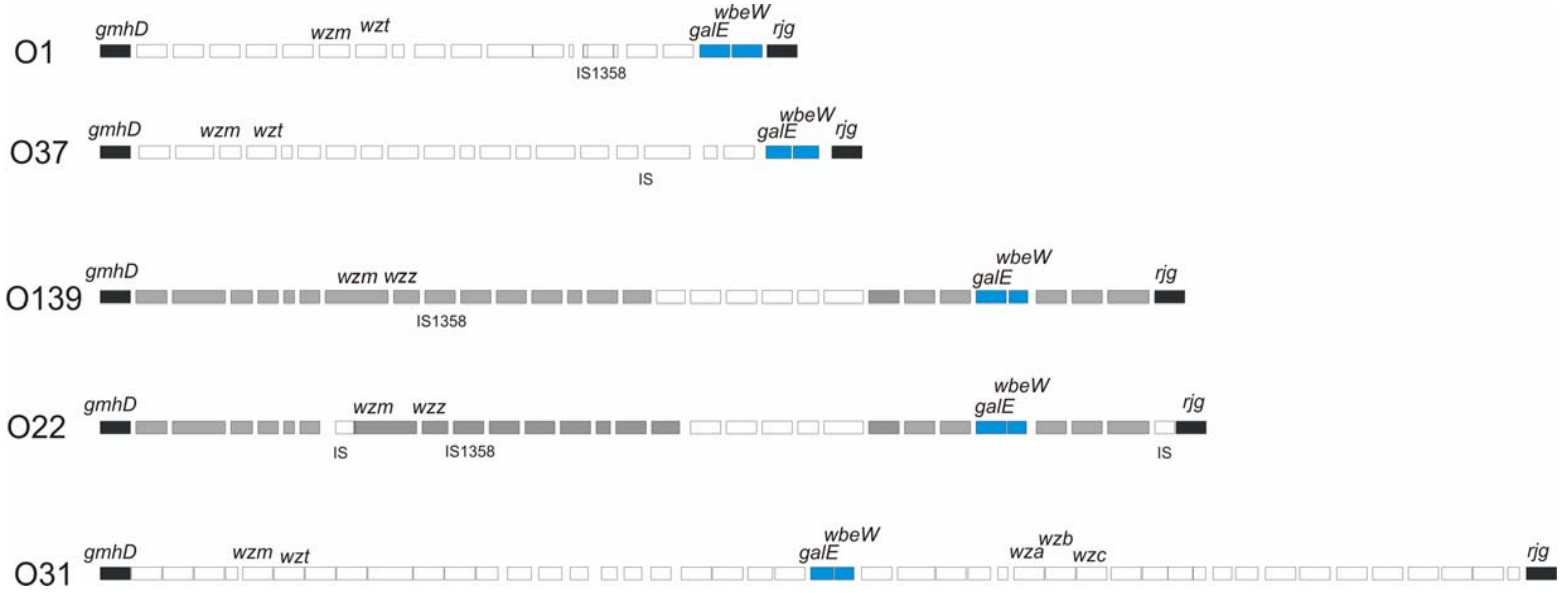


Figure 2.7. A schematic representation of CPS/O-antigen biogenesis regions in *V. cholerae*. The CPS/O-antigen biogenesis genes are between *gmhD* and *rjg* (black boxes). *GalE* and *wbeW* (blue boxes) are conserved in all the serogroups. *Wzm* gene is conserved in all serogroups; *wzt* is conserved in O1, O37, and O31 serogroups. The CPS/O-antigen regions between O139 and O22 are very similar; grey boxes indicate conserved genes; white boxes indicated genes that are not conserved between O139 and O22.

Table 2.1. Results of methylation analysis.

Glycosyl residue	Percentage present
3-linked Rha	25
terminal Gal	4
3,4-linked Rha	2
3-linked Gal	13
4-linked Gal	8
4-linked Glc	3
2- and 4-linked Gal	3
3- and 4-linked Gal	4
3- and 6-linked Gal	4
4- and 6-linked Gal	1
terminal GlcNAc	2
4-linked GlcNAc	17
4- and 6-linked GlcNAc	14

Table 2.2. Complete assignment of the NMR spectra of the de-O-acetylated polysaccharide

NRT36 (deOAc) ~50 °C (321 K)	H-1	H-2	H-3	H-4	H-5	H-6, H-6’
	C-1	C-2	C-3	C-4	C-5	C-6
A: **α-D-Gal**	5.19	3.66	3.67	3.50	3.97	3.79
	101.9	72.7	73.8	72.7	74.1	61.7
B: α-D-GlcNAc	5.01	3.93	4.05	3.64	4.16	4.09, 3.71
	95.5	53.9	72.0	80.2	69.9	68.6
C: β-L-Rha	4.85	4.16	3.61	3.42	3.40	1.33
	101.3	68.6	79.1	71.2	72.9	17.6
D: β-D-GlcNAc	4.42	3.78	3.67	3.67	3.49	3.83, 3.92
	102.3	56.3	74.7	77.6	75.3	61.7

Table 2.3. Residue linkages for the capsular polysaccharide of *V. cholerae* NRT36S

Linkage	HMBC	NOESY
A-B	AC1-BH4 AH1-BC4	
B-C	BH1-CC3	BH1-CH2 BH1-CH3
C-D	CH1-DC4	CH1-DH4
D-B		DH1-BH6 DH1-BH6’

Table 2.4. Complete assignment of the NMR spectra of the mono-O-acetylated polysaccharide. Shifts enclosed in parentheses are those of the 3-O-acetylated B residue in the native polysaccharide.

NRT36 Mono-O-Acetyl ~50 °C	H-1	H-2	H-3	H-4	H-5	H-6, H-6’
	C-1	C-2	C-3	C-4	C-5	C-6
A: α-D-Gal	5.15	3.63	3.67	3.51	3.95	3.76
	101.9	72.7	73.8	72.7	74.1	61.7
B: α-D-GlcNAc	4.99 (5.01)	3.91 (4.05)	3.99 (5.27)	3.63 (3.87)	4.14	4.09, 3.71
	94.9 (94.1)	53.9 (52.6)	71.6 (74.3)	80.3	68.6	68.4
C: β-L-Rha	5.04	5.53	3.80	3.48	3.49	1.36
	99.9	69.9	76.3	71.6	73.2	17.6
D: β-D-GlcNAc	4.39	3.77	3.66	3.69	3.43	3.90, 3.88
	10 2.1	56.3	74.7	77.8	75.1	61.7

Table 2.5. Complete assignment of the NMR spectra of the NAc and OAc.

assignment of NAc's, OAc's	N<u>H</u>	C=O	<u>C</u>H$_3$
B: α-D-GlcNAc	7.79	176.88	22.93; 2.03
D: β-D-GlcNAc	8.09	175.24	23.19; 2.08
C: β-L-Rha OAc		174.22	21.26; 2.21

Table 2.6. CPS genes identified by transposon mutagenesis

Clone #	Gene in CPS region	Putative gene functions	Best hit (AAI)
TR2	not	CYS regulon transcriptional activator	*V. cholerae* (100%)
TR23	not	FadR fatty acid metabolism rgulator protein	*V. cholerae* (100%)
TR286	not	Adenylate cyclase	*V. cholerae* (100%)
TR301	not	Ubiquinol cytochrome C reductase	*V. cholerae* (100%)
TR3	Orf23 (*galE*)	Nucleoside-diphosphate sugar epimerase	*V. cholerae* (98%)
TR17	Orf8	Glycosyltransferase	*Nitrosospira multiformis* (40%)
TR43 TR287	Orf43	Rhamnosyltransferase	*Shewanella sp.* (66%)
TR296	Orf5 (*wzm*)	ABC transporter system integral membrane protein	*Raoultella terrigena*(55%)

Table 2.7. Genes in the CPS/O-antigen biogenesis region

orf	symbol	annotated	%AA identity /positive	species	E-Value
0	*gmhD*	ADP-L-glycero-D-mannoheptose-6-epimerase	99/99	*Vibrio cholerae*	1.00E-180
1	*rmlB*	DTDP-D-glucose-4,6-dehydratase	99/99	*Vibrio cholerae*	0
2	*rmlA*	Glucose-1-phosphate thymidylyltransferase	100/100	*Vibrio cholerae*	1.00E-167
3	*rmlD*	DTDP-6-deoxy-L-mannose-dehydrogenase	99/99	*Vibrio cholerae*	1.00E-169
4	*rmlC*	DTDP-6-deoxy-D-glucose-3,5-epimerase	100/100	*Vibrio cholerae*	1.00E-104
5	*wzm*	ABC transporter system integral membrane protein	55/77	*Raoultella terrigena*	1.00E-81
6	*wzt*	ABC transporter system ATPase component	43/59	*Nitrosospira multiformis*	8.00E-86
7		glycosyltransferase	32/49	*Rubrobacter xylanophilus*	2.00E-24
8		glycosyltransferase	40/57	*Nitrosospira multiformis*	1.00E-125
9		glycosyltransferase	41/57	*Burkholderia fungorum*	1.10E-127
10		glycosyltransferase	38/56	*Burkholderia sp.*	8.00E-56
11		dTDP-glucose-4-keto-6-deoxy-D-glucose reductase	39/64	*Actinobacillus actinomycetemcomitans*	4.00E-21
12		probable acetyl transferase by domain	30/46	*Cellulophaga sp.*	0.049
13		hypothetical protein			
14		3-hydroxybutyryl-CoA dehydrogenase	48/68	*Pseudoalteromonas haloplanktis*	2.00E-70
15		Hypothetical protein			
16		glycosyltransferase	27/44	*Syntrophus aciditrophicus*	9.00E-06
17		Hypothetical protein			
18		hypothetical protein			
19		putative glycosyl transferase	55/69	*Pseudomonas fluorescens*	8.00E-97
20		UDP-N-acetylglucosamine 2-epimerase	67/82	*Yersinia intermedia*	1.00E-147
21		glycosyltransferase	96/97	*Vibrio cholerae*	2.00E-93
22		nucleoside-diphosphate sugar epimerase	99/99	*Vibrio cholerae*	0
23	*galE*	UDP-glucose 4-	98/99	*Vibrio cholerae*	0

		epimerase VC0262			
24	*wbeW*	galactosyl-transferase VC0263	99/99	*Vibrio cholerae*	4.00E-91
25		Trypsin-like serine proteases	96/97	*Vibrio cholerae*	1.00E-123
26	*wbfB*	hypothetical protein	99/99	*Vibrio cholerae*	0
27	*wbfC*	hypothetical protein wbfC, periplasmic	97/97	*Vibrio cholerae*	1.00E-141
28	*wbfD*	hypothetical protein wbfD	98/98	*Vibrio cholerae*	1.00E-112
29		hypothetical protein	52/58	*Vibrio cholerae*	6.00E-12
30	*wza*	Periplasmic protein involved in capsular polysaccharide export	67/82	*Vibrio splendidus*	1.00E-151
31	*wzb*	Protein-tyrosine-phosphatase	75/88	*Vibrio vulnificus*	6.00E-59
32	*wzc*	Putative tyrosine-protein kinase Wzc	75/87	*Vibrio vulnificus*	0
33	*rmlB*	dTDP-D-glucose-4,6-dehydratase	98/98	*Vibrio cholerae*	0
34	*rmlA*	glucose-1-phosphate thymidylyltransferase	100/100	*Vibrio cholerae*	1.00E-167
35	*rmlD*	dTDP-6-deoxy-L-mannose-dehydrogenase	99/99	*Vibrio cholerae*	1.00E-169
36	*rmlC*	DTDP-6-deoxy-D-glucose-3,5-epimerase	92/94	*Vibrio cholerae*	1.00E-88
37		O-acetyltransferase	38/58	*Enterococcus faecalis*	8.00E-17
38	*wzx*	O-antigen translocase	30/48	*Pelodictyon luteolum*	2.00E-39
39		glucosyltransferase	34/53	*Pseudoalteromonas tunicate*	1.00E-39
40	*wzy*	Putative saccharide polymerase			
41		Putative sugar acetyltransferase	51/72	*COG0110, Cytophaga hutchinsonii*	7.00E-43
42		glycosyltransferase	29/53	*Cytophaga hutchinsonii*	5.00E-41
43		Rhamnosyltransferase	66/77	*Shewanella sp.*	1.00E-113
44	*Ugd*	UDP-glucose 6-dehydrogenase	81/90	*Vibrio sp.*	0
45	*wecA*	Undecaprenylphosphate N-acetylglucosamine 1-phosphate transferase	96/99	*Vibrio cholerae*	1.00E-157
46		Hypothetical protein	96/96	*Vibrio cholerae*	2.00E-63
47	*rjg*	Predicted exonuclease of the beta-lactamase fold involved in RNA processing	98/99	*Vibrio cholerae*	0

CHAPTER III: THE GENOME OF NON-O1 *VIBRIO CHOLERAE* NRT36S DEMONSTRATES THE PRESENCE OF PATHOGENIC MECHANISMS THAT ARE DISTINCT FROM O1 *VIBRIO CHOLERAE*

ABSTRACT

Vibrio cholerae NRT36S is a non-cholera toxin (CT) producing, non-O1 *V. cholerae* strain that causes diarrhea in volunteers. The genome of NRT36S was sequenced to a draft containing 174 contigs plus the superintegron region. Our analysis of the draft genome revealed several putative toxin genes and colonization factors. Besides confirming the existence of nonagglutinable heat-stable toxin (NAG-ST), we also identified the genes for a type three secretion system, a putative exotoxin, and two different RTX toxin genes. Four pili systems were also identified.

STUDIES AND RESULTS

V. cholerae is best known as the causative agent of cholera. Cholera epidemics are associated with O1 and O139 serogroup isolates that produce cholera toxin (CT). *V. cholerae* from the other more than 200 serogroups reported for this species may be isolated from sporadic small outbreaks and isolated cases of diarrhea. The mechanisms by which these latter strains cause human disease remain controversial. In this study, we sequenced the genome of a known pathogenic non-O1 *V. cholerae* clinical isolate and compared it to the fully sequenced genome of *V. cholerae* O1 El Tor N16961 (20-22). *V. cholerae* NRT36S was originally isolated from a Japanese patient with travelers' diarrhea. It is serogroup O31, CT negative and produces NAG-ST (144). When fed to volunteers this isolate caused diarrhea, including, in one patient, a 5.3-liter diarrhea purge

(41). Molecular studies have indicated that this isolate is not closely related to epidemic O1 and O139 isolates (155).

The *V. cholerae* NRT36S genome was sequenced by the company 454 Life Sciences. Genomic DNA was extracted with a QIAamp DNA minikit (Qiagen, Valencia, CA). 454 Life Sciences employed a different sequencing strategy to the conventional shotgun sequencing. The method is as described in (151). Genomic DNA is randomly sheared to small fragments and ligated to common adaptors. The fragments are then amplified by PCR and then sequenced. Sequencing is performed on a Genome Sequencer 20^{TM} System, which is based on pyrosequencing protocol. The sequencing runs generated 1,082,967 reads and output 104,531,256 bp of sequence. The estimated coverage depth was 26X. The draft genome consisted of 174 large contigs plus the superintegron regions, with total length of 4,079,433 bp. The average GC content for the draft genome was 47.5%. The genome was annotated by the National Microbial Pathogen Data Resource and is available from the link: (http://anno-2.nmpdr.org/umd/FIG/index.cgi).

We compared the genome of *V. cholerae* NRT36S to the published genome of *V. cholerae* N16961, of serogroup O1, using the program MUMMER 3.0 (settings of the program were as in default) (Figure 3.1) (156). The genome sizes were comparable, at approximately 4.1 megabases. About 3.5 megabases (89%) of the sequences were common to both genomes. Substantial differences between the genomes of *V. cholerae* NRT36S and N16961 were noted, especially in the genes for pathogenesis, surface polysaccharides, and in the superintegrons. The sequences identified in only one isolate by MUMMER (cutoff 70% nucleotide identity) were considered strain-specific.

We confirmed that the genes related to pathogenesis in *V. cholerae* N16961 were strain-specific. These genes made up 30% of the N16961 specific sequences. The CTX∅ prophage (VC1452-VC1478), which encodes CT in *V. cholerae* O1 (32), was absent from NRT36S. The genome of CTX∅ also encodes genes responsible for the phage morphogenesis and its insertion into the host genome. None of these genes were present in NRT36S. The toxin co-regulated pilus (TCP), the major colonization factor for *V. cholerae* O1 (33, 34), was also missing from NRT36S. TCP is encoded in a 39.5 kb region of the *V. cholerae* genome, designated *Vibrio* pathogenic island 1 (Karaolis, 1998). Together with TCP, the genes encoded in *Vibrio* pathogenic island 1 (VC0819-VC0845), including ToxR and other accessory colonization factors were absent from NRT36S. Another genomic island was described as *Vibrio* pathogenic island 2 in *V. cholerae* O1 (VC1758-1803). VC1759-1772 encode a restriction modification system; VC1773-1787 encode enzymes, including a neuraminidase, involve in the utilization of amino sugars (nan-nag region); VC1788-1803 encode phage proteins (Figure 3.2) (157). Neuraminidase was found to be an important virulent factor that can increase the human receptor for CT (158). Most of the *Vibrio* pathogenic island 2 (VC1759-1772 and VC1788-1803) was absent from NRT36S. But an internal section (VC1773-1787) the nan-nag region was present. This interesting mosaic structure suggests that this region of the genome is a hot spot for lateral gene transfer.

Two other genomic islands, named *Vibrio* seventh pandemic islands-I and II (VC0175-0185 and VC0490-0497) have been reported to be specific to *V. cholerae* O1 El Tor and related O139 strains (157). The functions of these two islands are not clear. Neither of these islands was present in NRT36S.

The genes related to pathogenesis in NRT36S were different to N16961. In a previous study (41), Morris et al. suggested that the heat stable enterotoxin NAG-ST and the ability to colonize were the major factors contributing to occurrence of disease in human volunteers ingesting NRT36S. NAG-ST causes fluid accumulation in suckling mouse model (41, 144, 159). NAG-ST was not present in N16961. Interestingly, the NAG-ST gene was located to the superintegron region of the NRT36S genome (described below).

The colonization factor(s) in NRT36S were not identified in the previous studies. We found four pili systems in NRT36S. First, a type 1 pili assembly system was identified in an 8 kb sequence specific to the genome of NRT36S. Second, a pili system in NRT36S related to a mating pili system in *Salmonella* typhi (160) was identified. Third, a type IV pili system, which was a mannose-sensitive hemagglutinin, described by Jonson et al. (161), was present in both NRT36S and N16961. The protein sequences of the mannose-sensitive hemagglutinin system shared 60-99% identity between the two isolates. Fourth, another type IV pili system, described previously (162), was also conserved in NRT36S and N16961. None of the pili systems identified in NRT36S was similar to TCP, which is a type IV pilus.

We also identified a type III secretion system (TTSS) in a 48 kb gene cluster specific to NRT36S. It is highly similar to the one described for another non-O1/non-O139 *V. cholerae* isolate AM-19226 (163), sharing 99% sequence similarity. The location of the TTSS in NRT36S was next to the homolog of VC1758, while in N16961, the genes next to VC1758 were designated as *Vibrio* pathogenic island 2. In NRT36S, we also found 2 additional potential toxin gene clusters that were not found in N16961.

The first was an exotoxin A precursor gene. This gene had a homolog in the other non-O1/non-O139 *V. cholerae* isolates AM-19226 (163) and V51 (164). The predicted protein product is related to an NAD-dependent ADP-ribosyltransferase of *Pseudomonas aeruginosa*. Second, there were two RTX toxin genes present in NRT36S. The first one is similar to *rtxA* of N16961, sharing 99% amino acid identity. The second RTX toxin gene is similar to the RTX toxin gene in *Aeromonas salmonicida* (E=0. Score=640) and was very divergent from the *rtxA* found in N16961.

We identified two additional phage-like gene clusters in NRT36S. One prophage-like gene cluster was 33 kb and located adjacent to genes on chromosome II in N16961. The other cluster was 6 kb and showed similarity to filamentous bacteriophages KSF-1phi and VGJphi. Whether these two prophages play a role in virulence remains unknown. Altogether, the putative virulence genes and phage sequences made up 33% of the strain-specific sequences in NRT36S.

Besides the genes related to toxin and colonization, *V. cholerae* N16961 and NRT36S differed in their genes associated with synthesis of surface polysaccharide, which might also contribute to their virulence and survival. These genes made up 9% and 13% of the strain-specific sequences, respectively, in N16961 and NRT36S. The genes for O and K antigens in the two isolates were entirely different. N16961 is serogroup O1 and has no capsule; its O antigen biogenesis region has been identified (48), occurring between *gmhD* and *rjg* in chromosome II. NRT36S is serogroup O31 and has a capsule in addition to the O31 antigen. The O-antigen and capsule shared the same genetic locus in NRT36S, also located between *gmhD* and *rjg* (Chen, Bystricky et al.,

submitted for publication). Nevertheless as expected, the gene contents of this locus in N16961 and NRT36S were very different.

A superintegron exists in all the *Vibrio* genomes examined so far (20-22), and NRT36S was no exception. The superintegron is a highly variable region. The 454 sequencing strategies generated only short reads that could not read through the *Vibrio cholerae* Repeats and gave problems for assembling the superintegron in NRT36S. Therefore our knowledge to the superintegron regions is not complete. Among the genes we could confirm to the superintegron region, most of them were hypothetical proteins, as expected. Several genes were homologous to the recognized genes in N16961 superintegron: a killer protein, 4 putative acetyl transferase and some hypothetical proteins were conserved in the superintegron of both NRT36S and N16961. There were also several other genes recognized to be specific to the NRT36S superintegron. As mentioned above, the superintegron encoded NAG-ST, which may be the major toxin of NRT36S. We also identified a quinone oxidoreductase, and a hydrolase in the NRT36S superintegron.

In addition to the genes for toxins, pilus, surface polysaccharide, and superintegron, the difference between N16961 and NRT36S extended to other genes in the other functional categories such as chemotaxis, transportation and metabolism, and transcription regulation. These genes made up 32% and 24% of the strain-specific sequences in N16961 and NRT36S respectively. A methyl accepting chemotaxis system in N16961 (VC0511-VC0515) was very divergent from its analogue in NRT36S and was identified by MUMMER as isolate specific. The two systems shared about 30% amino acid identity. A phosphenolpyruvate phosphotransferase system (PTS) was found in

N16961, which might be involved in the transport of chitobiose in *V. cholerae* (20). This PTS system was unique to N16961 and was not found in NRT36S. NRT36S had a unique ABC type oligopeptide transport system of 11 kb long. It contained an outer membrane receptor gene, two genes of ATPase components, a periplasmic component gene and two genes of permease component. There was also a 6 kb NRT36S specific sequence contained a lysine motif (LysM) repeat, it was next to the homolog of VC0018 in NRT36S. While in N16961, there are three hypothetical proteins (VCA0019-VCA0021) specific to N16961 next to VC0018. There were also 54 other genes unique to N16961 and about 20 recognizable proteins plus some hypothetical proteins specific to NRT36S; they were randomly distributed in the genome and were in clusters less than 3kb.

To gain insight of the variations between homologous genes in *V. cholerae*, we also compared the proteomes of the two isolates by BLASTP (settings for BLASTP: no filter, e value< 1e-10). All best-hit pairs were identified from the two genomes. Best-hit pair here was defined as one gene in one genome found the other in the other genome as its best match and vice versa. Genes with 85% or more amino acid identity (AAI) and that cover at least 70% of the full length were considered conserved. 84% of the genes were conserved between the two genomes, less than 1% of the genes had low AAI (between 21% to 84%), another 1% of the genes represented paralogs in the genome, and the rest of the 14% of genes from NRT36S did not match N16961. Genes that had less than 85% AAI were considered strain-specific.

Our genome analysis reveals extensive variations among pathogenic strains within *V. cholerae*. NRT36S and the O1 strain N169061 clearly have entirely different

sets of virulence-associated genes. The observed variations in their surface polysaccharides and membrane transportation systems may reflect the adaptation to different niches during their life cycle. The function of the superintegron remains cryptic. NAG-ST is the first functional virulence gene found to be in the superintegron, which strongly suggests that the superintegron is a mechanism by which members of this species can import exogenous genes and convert them for their own adaptation.

We thank Ross Overbeek and Michael Fonstein in the National Microbial Pathogen Data Resource for their kindly help in the annotation of the *V. cholerae* NRT36S genome, and Lutz Krause in the University of Bielefeld, Germany for allow us to use their program GISMO (165) in the initial gene calling. We also thank Dr. Shiladitya DasSarma and Beenish Bhatia (Center of Marine Biotechnology, Baltimore, MD) for their help in the initial analysis of the genome.

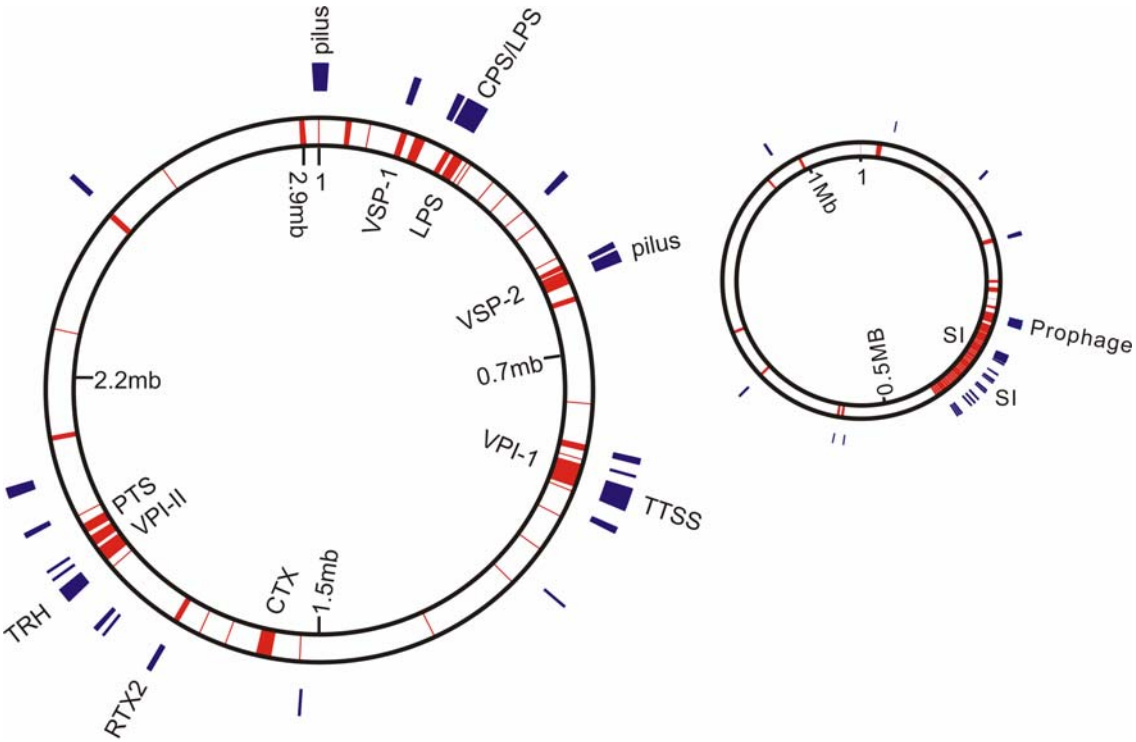


Figure 3.1. Comparison between *V. cholerae* genomes of N16961 and NRT36S. The two close rings represent two chromosomes of N16961. Chromosome I is on the left and chromosome II is on the right. The white areas in the rings are sequences conserved between the two genomes. Red areas are sequences specific to N16961, with labels inside the rings. Blue areas outside the rings are sequences specific to NRT36S, with labels outer most.

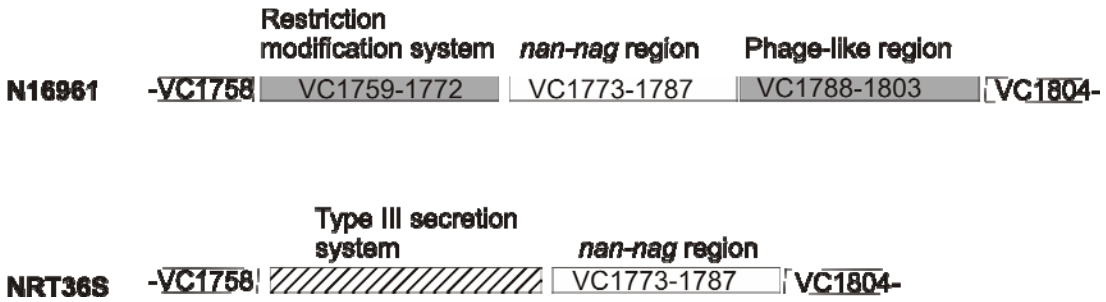


Figure 3.2. Mosaic pattern of *Vibrio* pathogenic island 2. The three regions in the *Vibrio* pathogenic island were described in (157). White boxes are conserved regions between N16961 and NRT36S. Shades boxes indicates genes specific to N16961. Box with stripe pattern indicates genes specific to NRT36S.

CHAPTER IV: GENETIC VARIATION OF CAPSULE/LIPOPOLYSACCHARIDE BIOGENESIS IN TWO SEROGROUP O31 *VIBRIO CHOLERAE* ISOLATES

ABSTRACT

NRT36S and A5 are both NAG-ST producing, serogroup O31 *Vibrio cholerae.* NRT36S is encapsulated and causes diarrhea when administered to volunteers; A5 is acapsular and does not colonize or cause illness in humans. The Capsule/Lipopolysaccharide (CPS/LPS) biogenesis regions in these two isolates were similar except that a 6.5 kb fragment in A5 replaced a 10 kb fragment in NRT36S in the middle of the LPS gene cluster. Our data highlight the apparent mobility within the CPS/LPS region that would provide a basis for the large number of observed *V. cholerae* serogroups and the emergence of novel epidemic strains.

STUDIES AND RESULTS

Vibrio cholerae has nearly 200 serogroups based on the O antigen (14). Antibodies specific for O1 and O139 antigens are critical for protective immunity against cholera. Most non-O1 serogroup isolates also have a capsule that is critical for virulence in extraintestinal infections (50, 73). The biogenesis of these polysaccharides in *V. cholerae* is not well understood. Although O1 is the predominant serogroup that causes pandemic cholera, new strains with epidemic potential continue to emerge by acquiring new O antigens and capsule polysaccharides. Understanding the genetic basis underlying the generation of new O antigens and capsules is critical for developing insight into the evolution of *V. cholerae*, and may influence the strategy for vaccine development.

The O antigen biogenesis regions of five *V. cholerae* isolates from different serogroups (O1, O22, O31, O37 & O139) have been resolved (15, 16, 18, 27, 47, 49, 166), all flanked by the genes *gmhD* and *rjg*. In serogroup O139 isolates, both the O antigen and the capsular polysaccharide (CPS) are constructed of the same subunit that is encoded in a single region. In the O31 *V. cholerae* strain NRT36S (isolated from a Japanese patient with travelers diarrhea (41, 159)), the CPS and O antigen share the same biogenesis region, which consists of 46 genes (166). Although there are immunologic differences in the CPS and the LPS, there is no immediately obvious organizational differentiation between the genes specific for CPS versus O antigen biosynthesis.

V. cholerae A5 (cultured from frozen shrimp (144)) is a naturally-occurring acapsular isolate closely related to NRT36S, as shown by genetic studies (155). In a multi locus sequencing typing (MLST) study, the two strains were identical for seven sequenced housekeeping genes, *cat*, *chi*, *dnaE*, *gyrB*, *lap*, *recA*, and *pgm* (unpublished data). They have the same serogroup, O31, and both produce NAG-ST (144, 159). However, a volunteer study showed that A5 failed to colonize and did not cause disease while NRT36S colonized volunteers and produced severe diarrhea (41). Given that both strains are toxigenic (producing NAG-ST, a well-recognized diarrheal toxin), it is reasonable to hypothesize that the difference in pathogenicity is a function of colonizing ability; this, in turn, may be related to capsule status. To assess the genetic basis for this difference, we sequenced the O antigen/capsule biosynthesis region of A5, and compared it with the previously reported sequence of NRT36S.

Origin of the A5 polysaccharide genes: Since both NRT36S and A5 are in serogroup O31, we anticipated that their O-antigen genes would be very similar and that

CPS gene(s) might have been deleted or mutated in A5. We designed primers to amplify different regions of the CPS/LPS gene cluster in NRT36S and A5, with the total outcome spanning the entire cluster from *gmhD* to *rjg*. The same size PCR products were amplified from NRT36S and A5 from all regions except one. Primers 1197 (TGAGCAATCCGGTATTGAAGTGAAA) and 1528 (AGGTTATCGTACAGTGCTTT) amplified a fragement of ~7 kb from A5 in contrast to a ~11 kb fragment from NRT36S. The PCR products amplified with primers 1197 and 1528 were sequenced (GenBank accession number EF076669). Open reading frames were annotated using the program Artemis (154) and putative function was assigned by similarity searching against GenBank with the program BlastP (152, 167). Sequence analysis from this region indicated that a 6.4 kb DNA sequence in A5 had replaced a 10.5 kb region in NRT36S, with breakpoints in the genes *wzc* and *ugd* (Figure 4.1A, 1B and 1C) suggesting a single mutational event. The 10.5 kb region specific for NRT36S encoded 11 genes, including a set of the 4 rhamnose pathway (*rmlBADC*) genes, a flippase (*wzx*), two O-acetyl transferase, a putative *wzy* gene, two glycosyltransferases, and a putative rhamnosyltransferase (orf43) (Figure 4.1A, Table 4.1).

The A5 specific sequences in the LPS region encoded six genes (Figure 4.1A, Table 4.1), including a *wzx* (O antigen flippase), a *wzy* (O antigen polymerase), an O-acetyl transferase (orf D), and three putative glycosyltransferases (orfs B, E and F). Between the genes specific for A5 and the ones specific for NRT36S, *wzx* was the only gene that had significant similarity to its counter part in the other isolate, sharing 28% amino acid identity (E=4e-35). Although there was little similarity between the strain specific regions from A5 and NRT36S, there was a striking similarity in the organization (Table

4.1). The unique region from each isolate consisted of a *wzx* homologue, glycosyltransferase, putative *wzy* homologue, acetyltransferase, and two glycosyltransferases in the same order in both isolates. The rhamnose genes in the NRT36S specific region are duplicates with *rml*A-D genes occurring in ORFs 1-4 in NRT36S.

Sufficient genes to synthesize O antigens that react with O31 typing sera are clearly present in the regions shared by NRT36S and A5. It is possible that the A5 specific genes may be able to substitute for the NRT36S genes they replace in O antigen biogenesis, or that the exchange of genes corresponds to a change in the O antigen structure that is not reflected by the specificity of the typing sera.

The region in NRT36S replaced in A5 contains at least part of the genes necessary to encode capsule biogenesis. In a previous study of NRT36S, orf43 was disrupted in two independent transposon mutagenesis experiments, and the disruption resulted in the loss of the capsule in NRT36S (166). The gene orf43 was among the ones replaced in the acapsular A5.

As noted above, the genes *gmhD* and *rjg* appear to form a cassette-like element containing the genes necessary for O antigen/CPS biogenesis. The emergence of *V. cholerae* O139 is consistent with homologous exchange within these flanking genes (15, 47, 48). Within this larger cassette, the region divergent between NRT36S and A5 was flanked by two genes, *wzc* and *ugd,* consistent with formation of a "sub-cassette" capable of undergoing lateral gene transfer. Interestingly, this substitution did not result in a change in O antigen or capsule type, but rather the loss of capsule expression. This contrasts with observations made with the polysaccharide biogenesis genes of *V. cholerae*

O139 and O22; these biosynthesis regions are very similar, but have a small region that differs, resulting in a change in serogroup. It is less clear that a single exchange could result in the differences between O139 and O22 strains, and it has been suggested that insertion elements were involved in the gene rearrangement in these two strains (49).

GC percentage pattern: Our results are consistent with the idea that there is a constant "mixing and matching" of different sub-cassettes within the O antigen/CPS biosynthesis region, providing a basis for the large number of observed *V. cholerae* serogroups. To further explore this concept, we examined the patterns of GC content in the O antigen/CPS region of *V. cholerae*. GC percentage plots were generated using a 120 bp widow size with the program Artemis (154). The O antigen/CPS of NRT36S displayed five distinct regions, with dramatically changes of GC content among regions (Figure 4.2). In region I, III, V, the GC contents were (46-47%), close to the GC content of the chromosomal backbone (48%). Region II and IV had lower GC content, 36% and 31% respectively, suggested these two regions were acquired recently. In A5, region IV was replaced by the A5 specific sequence and GC content of the new sequence was 29%. Interestingly, all genes in region I, III, V of NRT36S had homologs in *Vibrio*, suggesting these genes were acquired by the ancestral *V. cholerae*. In contrast, none of the genes in region II and IV had homologs in *Vibrio* in the current database (166). This observation was consistent with the GC patterns, and strongly suggested that there were at least two recombination events recently in the CPS region of NRT36S, acquiring region II and region IV respectively.

We also examined the GC pattern in the O antigen/CPS cluster in *Vibrio cholerae* O139 and O22. There were three distinguishable regions in either O139 or O22, and

region II had low GC content (32% and 31%, Figure 4.2). This region was found to be the only region to diverge significantly from O139 to O22 (49). The case of O139 and O22 appears to be very similar to that of NRT36S and A5, supporting the concept that homologous exchanges between sub-cassettes in the polysaccharide region alter the structure of the surface polysaccharide. Secondly, the %GC differences suggest that this exchange may involve acquisition of genes from other species.

Phylogeny: We noticed that GC content changed abruptly between different regions in the CPS/LPS region, as described above. Furthermore, genes in the region with low GC content did not have homologs in *Vibrio*. This suggested that original source of these low GC content regions might be outside of *Vibrionacea*. To further investigate that idea, we examined the phylogeny of housekkeping genes and compare it to two conserved polysaccharide genes within *V. cholerae*, *ugd* and *galE*. If there were extensive recombination within polysaccharide biosynthetic regions, one would expect the evolution of polysaccharide biosynthesis genes to follow an evolutionary tree that includes multiple independent branches compared to genes from genetic backbone where the tree will have single branch for *V. cholerae*. Homologous genes of *galE*, *ugd*, were retrieved from GenBank. The protein sequences were aligned with ClustalX (168) and the genetic trees were constructed with Treeview (169). Besides O31, Ugd was also present in *V. cholerae* serogroups O22, O139, O141 and in many other species of gram-negative bacteria. An unrooted tree (Figure 4.3) was constructed using the protein sequences of *ugd* genes. Within the cluster of *Vibrionacea*, the Ugd sequences of *V. cholerae* O139, O22 and O141 were closely related and clustered together (cluster A), while those of *ugd* from the O31 isolates (A5 and NRT36S) were far away from cluster

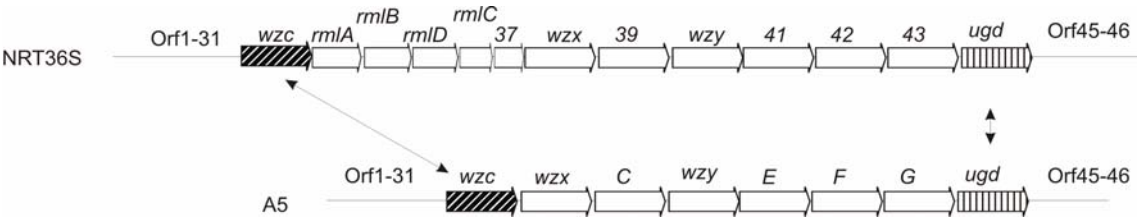
A. The Ugd of O31 and O139 shared 80% Amino Acid Identity (AAI). The distance from the O31 to cluster A was further than the distance of many other *Vibrio* sp. (other than *V. cholerae*) to cluster A, suggesting there were two different evolution lineages of *ugd* gene in *V. cholerae*.

galE was present in many *V. cholerae* strains. An unrooted tree was constructed based on their protein sequences (Figure 4.3). Similar to Ugd, *V. cholerae* separated into two clusters in the GalE tree (C and D in Figure 4.3), Cluster C consisted of *V. cholerae* O1 (El Tor and classical strain 0395), O37 (1322-69 and V52), O31 (NRT36S and A5), cluster D consisted of O139 (MO10 and AI1837) and O22. The AAI between the two clusters were about 50%, while the AAIs were 98-99% inside clusters. These results strongly suggest that there are different ancestral *galE* genes for *V. cholerae.*

Our analysis from the conserved polysaccharide biogenesis genes indicated that there were at least two evolutionary lineages of *V. cholerae* polysaccharide genes. Opposite to the evolution lineages of the genetic backbone, O1 and O139 were separated in two evolution lineages for the polysaccharide genes, each together with different serogroups of non-O1, non-O139 strains. In the evolution of polysaccharide genes, O1 was close to O31 and O37, while O139 was close to O22 and O141. There are many other species intervening between the two *V. cholerae* lineages, supporting our conclusion that homologous exchange between different species contributed to the vast variation in *V. cholerae* surface polysaccharides.

The polysaccharide biogenesis region is one of the most variable regions in the genome and lateral gene transfer has played an important role in its evolution. As a result, the serogroups in *V. cholerae* do not necessarily follow the evolution of its

chromosomal backbone, as reflected in phylogenetic trees constructed in other studies (23). For housekeeping genes such as those used for MLST, all of the genes are very closely related, indicate that vast majority of housekeeping genes have a single lineage or phylogenetic branch for *V. cholerae* (23, 25, 155). In contrast, *gmd*, a gene for O-antigen biogenesis in *V. cholerae* O139, has an allele that has greater similarity to *E. coli* than to *Vibrio*, suggests lateral gene transfer (25). The evolution of the CPS/LPS region hence seems to follow its own path, different from the lineage of the chromosomal backbone, with the frequent occurrence of lateral gene transfer, often across species boundaries. From our study, we found that genes switch from NRT36S to A5 were strikingly similar in three ways: a) same type of genes, both related to polysaccharide biogenesis; b) similar organization; c) similar GC pattern. All these evidences suggest that there is a controlled mechanism for the recognition of and recombination in the LPS/CPS region leading to increased mobility of genes in this region, and the surface polysaccharides of *V. cholerae* are highly plastic. The inclusion of LPS/CPS genes from other species into the pool greatly increases the chance for the emergence of new epidemic strains of *V. cholerae*. As a result, novel epidemic strains with new serogroups and CPS structures will continue to rapidly emerge.



A.

```
NRT36S
AACGCCGTTGAGAAGAAAGCATCAAGTACCTATGGCTATTACAACTACAGCTATGGTGATGCAAA
A5
AACGCCGTTGAGAAGAAAGCATCAAGTGCTTACGGGTATGGTTACTACAATTATAGTTACAGCGA
*************************** * ** ** ***    ******  *** ** *     *
                 breakpoint^
```

B.

```
NRT36S
CGGAGAAGTTACATGAATATAACAATAGCTGGTACAGGCTACGTCGGCCTGT
A5
AAGAAGGTTAATATGCAAATTGCTATTGCAGGTACAGGCTACGTCGGTCTGT
  **    * * *** * **  * ** ** ***************** ****
                  breakpoint^
```

C.

Figure 4.1. Divergent regions in CPS/LPS of NRT36S and A5. Arrows indicate the breakpoints. A. NRT36S specific region have 11 genes. A5 specific region have 6 different genes. The two regions are flanked by two conserved genes *wzc* and *ugd*. B. Alignment of the 3' end of *wzc* gene. C. Alignment of the beginning (5' end) of *ugd* gene.

Table 4.1. Strain specific genes.

NRT36S	A5
O antigen translocase wzx (*Pelodictyon luteolum* 30%)	O antigen translocase wzx (*Pelodictyon luteolum* 32%)
Glucosyltransferase (*Pseudoalteromonas tunicate* 34%)	Glycosyltransferase (*Methanosphaera stadtmanae* 29%)
Predicted saccharide polymerase wzy	Putative saccharide polymerase wzy (*Shewanella oneidensis* 23%)
Sugar acetyltransfease (*Cytophage hutchinsonii* 51%)	O-actetyl transferase (*Clostridium acetobutylicum* 26%)
Glycosyltransferase (*Cytophaga hutchinsonii* 29%)	Glycosyltransferase (*E. coli* 39%)
Rhamnosyltransfease (*Shewanella sp*. 66%)	Glycosyltransferase (*Vibrio fischeri* 51%)

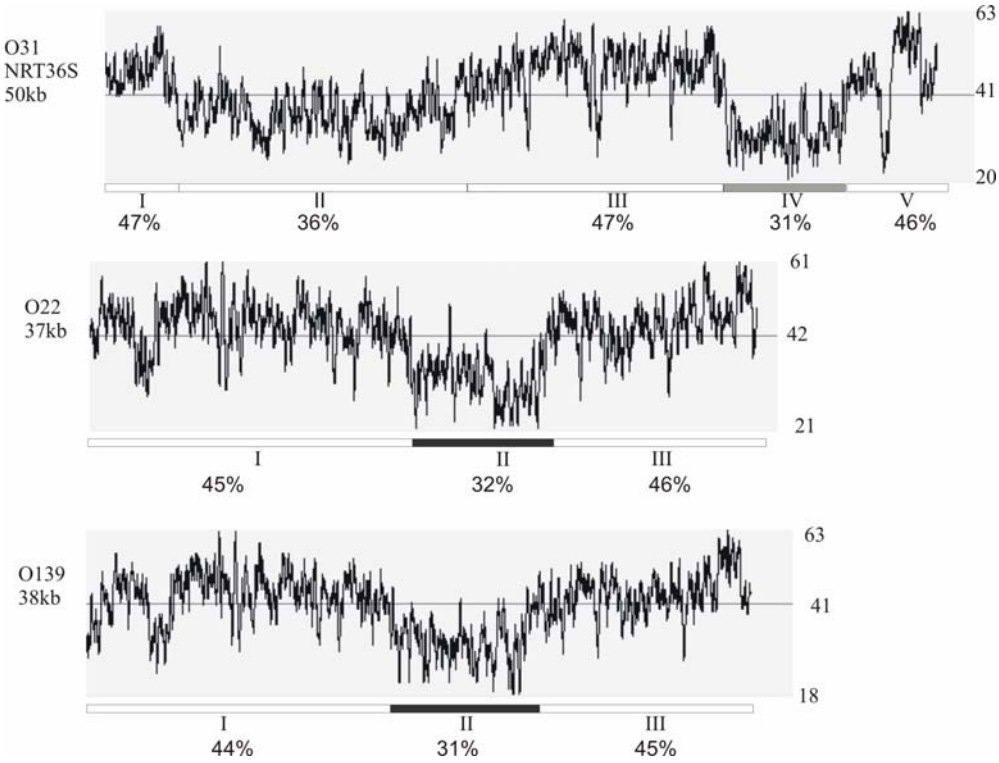


Figure 4.2. GC content of O antigen/CPS clusters in *V. cholerae*. Average GC content of each region is indicated under each plot. The numbers on the right are the range of GC content in percentage. The grey line of each plot indicated the average GC content of the entire cluster. Grey filled bar indicates the divergent region between NRT36S and A5. Black filled bars indicate the divergent regions between O139 and O1.

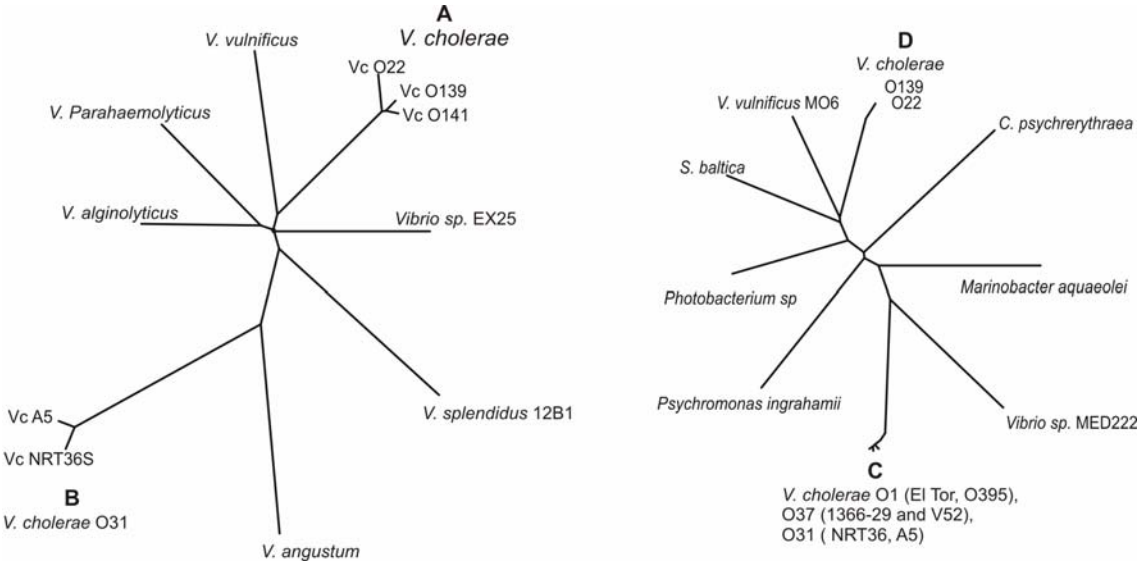


Figure 4.3. Unrooted protein trees of Ugd (left) and GalE (right). The length of branch is proportional to the genetic distance. Ugd of *V. cholerae* separate into two clusters (A and B). GalE of *V. cholerae* also sepatate into two clusters (C and D).

CHAPTER V: DETAILED EXPERIMENTAL METHODS

Culture conditions

Cultures were maintained in L broth with 15% glycerol at -70°C. Bacteria were cultured in the L broth or agar at 37°C unless otherwise stated. Appropriate antibiotics were added in concentrations: 50 mg/ml Kanamycin, 50 unit/μl polymycin B. To visualize the opacity of colonies, bacteria were grown on L agar containing 10g/L NaCl. A beaker of water was put inside the incubator to keep humility. I found these growing conditions produce better differentiation between opaque and translucent morphology.

Transposon mutagenesis

Conjugations were performed between *V. cholerae* NRT36S as the recipient strain and *E. coli* S17λpir/putKm2 as the donor strain. Ten μl overnight culture of *V. cholerae* NRT36S was spotted on LB agar and let dry, 10 μl overnight culture of S17λpir/putKm2 was then spotted on top. After overnight incubation at room temperature, the mixture was re-suspended in 1 ml of LB broth; 50 μl of the suspension was plated onto LB agar with kanamycin and polymycin B to select for *V. cholerae* mutants.

Analysis of DNA in NRT36S mutants

DNA flanking the transposon in the mutants was amplified and sequenced by a modified inverse polymerase chain reaction (PCR) protocol (147). Genomic DNA was isolated with PrepMan^{TM} (Applied Biosystems) according to the manufacturer's instruction. A small amount of bacteria grown on LB agar was picked into 100 μl of PrepMan^{TM} sample buffer, dispersed by gentle pipetting, boiled for 15 min. The samples

were cooled down and spun for 15 min. The supernatant was transferred to a fresh tube and ready for downstream application. Two µl genomic DNA of the above preparation was digested with 5 units of *Nla* III (New England Biolab) in a 20 µl reaction overnight followed by denaturing at 65°C for 15 minutes. Two µl of the digested DNA was self-ligated with 5 units of T4 DNA ligase (Invitrogen) in a 10 µl reaction. One micro liter of this solution was used as PCR template. Two primers were designed to anneal to the transposon mini-Km2, pointing outwards to amplify the flanking sequence of the mutant genomic DNA. The sequences of the primers are, L8 (reverse), GTACCGAGCTCGAATTCGGCCTAG; and L9 (forward), GGAGAAAACTCACCGAGGCAGTTC. PCR was performed in a 30 µl reaction containing 100 µM of each dNTP, 1.5 mM of $MgCl_2$, 1x PCR buffer (Invitrogen) and 1 unit of Taq DNA polymerase (Invitrogen). We used a PE9700 thermo cycler (Applied Biosystems) with the following program: An initial denaturing at 95ºC for 2min, followed by 35 cycles of: 95ºC for 30 second, 62ºC for 30 second, and 72ºC for 2min. PCR reaction was kept at 4ºC. PCR products were purified with the Multiscreen PCR plates (Millipore) and sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). The resulting fragments were separated and recorded in an ABI 3730x1 automatic sequencer (Applied Biosystems). DNA sequence was then analyzed by the PHRED and PHRAP software (148-150).

Serum Killing

Translucent colonies were challenged with complement in human serum as previously described (50). Thirty six µl of guinea pig compliment (Sigma) was added to

1 ml of pooled human serum to make serum mixture. Thirty five µl of bacteria culture, containing $0.5x10^7$ cells was mixed with 65 µl of serum mixture and incubated at 37ºC for 1 hour. Control aliquots were mixed with serum that had been heated to 56ºC for 30 min. Two concentrations, either 5 or 45 µl of this mixture were plated on L agar.

Size Exclusion Chromatography (SEC)

Capsule preps were analyzed by SEC using a Beckman Coulter 32 Karat HPLC, with TSK gel column (JOSOHAAS; G3000SWxL; 10um; 30cmx7.5mm), and detected at 200nm wavelength. Purified NRT36S capsule was the same sample that was used for NMR. Capsule preps were prepared as following: The amount of 10^9 cells were harvested into 0.5X PBS and shaken for 2 hours in a rotary shaker at 250rpm followed by centrifuge at 12000g for 20min. The supernatant was treated with 20 µl proteinase K (Qiagen). The supernatant was then extracted with phenol-chloroform and precipitated with ethanol. The pellet was re-suspended in water and 1/3 of the amount was loaded.

Immuno Blotting

Immuno blotting was performed as described (47). Circa $5x10^6$ bacterial cells were washed with 0.5X PBS and treated with proteinase K (Qiagen). Washed whole cell in PBS was mixed with 2X sample buffer and boiled for 5 min and was loaded on 16% SDS-polyacrylamide gel. The samples were separated for 1.5 hours at 150 volt in an Xcell II electrophoresis apparatus (Invitrogen). Samples were transferred to an Immuno-Blot PVDF membrane (BioRad, Hercules, CA) for 2 hours at 25 volt in an Xcell II transfer module. Blots were blocked in PBS contains 3% non-fat dry milk and then

incubated for 1h in 1:1000 rabbit antiserum specific for *V. cholerae* NRT36S. The blots were washed three times with PBS and incubated with alkaline phosphatase-conjugated goat anti-rabbit immunoglobin G (Sigma) at 1:10,000 in PBS for 1h. The blot was washed five times with PBS and developed with Western Blue colorimetric detection solution (Promega).

Transformation of *V. cholerae* by electroporation

V. cholerae was grown to log phase. The following steps were all performed at 4ºC. Twenty ml of culture was centrifuged for 2 min at 1000g; Pellets were washed twice with 10 ml of 137 mM sucrose pre-chilled to 4ºC, and re-suspended in 400 µl of 137 mM sucrose with 10% glycerol. The cells were aliquot into 100 µl each.

One µg plasmid DNA in 10 µl water was mixed with one aliquot of competent cells; chilled on ice for 15 min and transferred to a 0.2 cm Gene Pulsar cuvette (Biorad). Electroportation was performed in a Gene Pulsar Xcell electroporator (Biorad) with settings: 25 uF capacity, 2500 volt and 200 Ω. Cells were immediately transferred to 1 ml of SOC medium pre-warmed to 37ºC, and shaken for 1 hour at 37ºC. Fifty to 100 µl of culture was plated L agar with appropriate antibiotics.

Complementation of translucent mutant Tr3

A DNA fragment containing genes *galE*, *wbeW* and their promoter was amplified from the genomic DNA of NRT36S by PCR. Primers used are Stine1454: TGTCTAGAATTGCTTAGCAGCTCGCCTT, and Stine 1455: TTCTCGAGAATCCGAACCGCATCTGGAT. PCR was performed in a 30 µl reaction

containing 20 ng template DNA, 100 μM of each dNTP, 20 picomole of primers, 1.5 mM of $MgCl_2$, 1x PCR buffer and 1 unit of TaqGold^{TM} DNA polymerase (Applied Biosystems). The PCR program of thermal cycler was: Initial denaturing at 95ºC for 30 sec, followed by 30 cycles of 95ºC for 30 second, 55ºC for 30 second, and 72ºC for 1 min.

The PCR fragment was inserted into vector PCR2.1 using TA cloning kit (Invitrogen) and transformed into Top10 *E. coli* cells (Invitrogen). The plasmid constructed, i. e. *galE* and *wbeW* in PCR2.1 was named pYS101. The *galE* gene and its promoter were excised from pYS101 by restriction enzymes Hind III and Xho I; inserted into plasmid pACYC184, which was linearized by Hind III and Sal I. The resulted plasmid was named pYS104 and multiplied in *E. coli* OmniMAX (invitrogen). pYS104 was isolated and introduced into competent cells of Tr3 by electroporation (described previously).

Amplification of the LPS region in A5:

We designed primers to amplify different regions of the CPS/LPS gene cluster in NRT36S and A5, with the total outcome spanning the entire cluster from *gmhD* to *rjg*. The same size PCR products were amplified from NRT36S and A5 from all regions except one. Primers are listed in table 5.1. Unless stated otherwise, PCR conditions are as follow: 94ºC for 1min, followed by 35 cycles of 94ºC 30 sec, 55 ºC or 50ºC 30 sec and 72ºC 4 min. The elongation time for primer pair 10 is 1 in. Primers 1197 and 1528 amplified a fragment of ~7 kb from A5 in contrast to a ~11 kb fragment from NRT36S. PCR with primers 1528 and 1197 was performed with GeneAmp XL PCR kit (Applied

Biosystems) following the manufacture's instruction. The long range PCR program of thermal cycler was: 94ºC for 1min, followed by 30 cycles of 94ºC 15 sec, 55 ºC 30 sec and 68ºC 10 min. The PCR products amplified with primers 1197 and 1528 were sequenced.

Table 5.1 Primers to amplify the CPS/LPS region of NRT36S and A5. LR PCR, long range PCR. *, this primer anneals nonspecifically to *wzc* gene.

Pair	Primer	Position in NRT36S CPS	Sequence	Condition
1	J101 F	>1 (*gmhD*)	GCCATCCCACTCTGTGGTCGCA GAGCAAGCTCC	LR PCR
	1383 R	9584 (orf 8)	TAGAGTATTCATCATTCCCC	
2	J101 F	>1 (*gmhD*)	GCCATCCCACTCTGTGGTCGCAGA GCAAGCTCC	55°C
	1150 R	4526 (*rmlC* 1)	TGTCCGATAATAATAGCCCTTCCG	
3	1149 F	4920 (*wzm*)	CCATTAGAGATACTTCCTGTGGTA G	55°C
	1385 R	8097 (orf 7)	TCATTTCCGTCAATGTAGCA	
4	1384 F	9092 (orf 8)	ACTAACTTCAATATTGAGAGAG	55°C
	1541 R	13125 (orf10)	CTGTTAAACCAGCAGCACAATA	
5	1540 F	13105 (orf 10)	TATTGTGCTGCTGGTTTAACAG	50°C
	1543 R	17146 (orf 16)	AACAGCAAAAGCAAACACAGGGA	
6	1542 F	17122 (orf 16)	TCCCTGTGTTTGCTTTTGCTGTT	50°C
	1410 R	21552 (orf 22)	GGAACTTGATGTATGCTGTC	
7	1411 F	20950 (orf 21)	CCGCCCATTCAATTCTGATT	55°C
	1381 R	24596 (*wbeW*)	CTTACCGACTCGCTGTTGGC	
8	1382 F*	24063 (*galE*)	GATTCCCTTACCTCTTGGTGC	LR PCR
	J103 R	48630 (*rjg*)	CCCGTGACACTCGCCTTCCCTCCGT GATGAACC	
9	1197 F	34660 (*wzc*)	TGAGCAATCCGGTATTGAAGTGAA A	LR PCR
	1528 R	45712 (*ugd*)	AGGTTATCGTACAGTGCTTT	
10	1547 R	A5 *(wzy)*	AGTGAGTTTATTGTAAGATTTAG	55°C
	1549 F	A5 (orf C)	CATTAAAGTAAGCAACACGCA	
11	1190 F	34730 (*wzc*)	GCTATTACAACTACAGCTATGGTG	LR PCR
	J103 R	48630 (*rjg*)	CCCGTGACACTCGCCTTCCCTCCGT GATGAACC	

CHAPTER VI: CONCLUSIONS AND FUTURE PROSPECTS

SUMMARY

My first hypothesis was that there was a distinct region for the capsule biogenesis, different from the lipopolysaccharide (LPS) biogenesis region. I have identified the LPS region in the NRT36S genome based on our understanding of the conserved genetic structure of the LPS in *V. cholerae*. I also identified the genes for capsule biogenesis by constructing transposon mutants of *V. cholerae* NRT36S. Transposons mapped to the LPS region resulted in the loss of CPS, suggesting that it was a single region for both LPS and CPS biogenesis. When the entire genomic sequence of NRT36S was examined, only two polysaccharide biogenesis regions were identified, one is as the exopolysaccharide (rugose) biogenesis gene cluster (VC0917-VC0939), and the other was the LPS/CPS region. The lack of a third polysaccharide region supported the sharing of LPS and CPS genes. Therefore, in contrast to my original hypothesis, I concluded that the capsule genes were embedded among the O-antigen gene region. Although this is similar to the case in O139, there are significant differences. The capsule polysaccharide in O139 shares the same repeating unit as the O-antigen and the O-antigen has only one single repeating unit attached to the core. Although we have not determined the structure of the O-antigen in NRT36S, there are several lines of evidence suggesting that the capsule and O-antigen are two different entities in this strain. First, the capsule was unable to react to the antibodies raised against the whole cell, while the O-antigen did; second, the capsule contains rhamnose, while other studies suggested the O-antigen of this serogroup did not have rhamnose (142); third, the natural mutant A5 lost the capsule

yet retained its O-antigen suggesting that some genes required for capsule biogenesis are not needed for LPS formation.

In contrast to my original hypothesis, I drew our final conclusion that the capsule and LPS biogenesis region were combined in at least some strains of *V. cholerae*. This finding was novel and significantly different from *E. coli*. In all four groups of *E. coli*, there are different gene clusters for the capsule and O antigen. The integration of CPS and O-antigen genes, and the fact that they are flanked by two conserved homologous genes, suggests that *V. cholerae* can readily change its capsule and O-antigen simultaneously, just like changing its camouflage, yet maintaining its other virulence factors in the genome. This finding has important implications in vaccine development since the O antigen is the major protective antigen. Vaccine strains should also be acapsular since encapsulated *V. cholerae* can evade phagocytosis and become invasive, and cause septicemia in the immuno-compromised patients.

To further investigate the genetic basis underlying the generation of new O antigen/CPS, I compared the CPS/O-antigen biogenesis regions in two natural *V. cholerae* variants (A5 and NRT36S), with the same O serogroup. I found direct evidence that homologous exchange of part of the polysaccharide biogenesis region caused the variation in the surface polysaccharide structures. *V. cholerae* A5 replaced some of the CPS/O-antigen biogenesis genes of NRT36S with other new genes. Consequently, A5 lost the CPS but retained the same O-antigen as NRT36S. This evidence supported the hypothesis that lateral gene transfer by homologous recombination played an important role in the evolution of polysaccharide genes, and hence of *V. cholerae*.

The O antigen/CPS region in *V. cholerae* is flanked by two homologous genes *gmhD* and *rjg*. And inside this region, there are different genes conserved among different serogroups and can serve as the basis for homologous recombination. The genetic structure of O antigen/CPS in *V. cholerae* suggests that this region not only can undergo homologous recombination as one unit, but also can exchange its component sections as sub-cassettes. The concept of recombination between sub-cassettes is supported by the GC patterns of the O antigen/CPS regions. GC percentage changes abruptly between different sub-cassettes. The most variable sub-cassettes in NRT36S and A5, and in O139 and O22 have much lower GC content compared to their genetic backbone, suggesting recent acquisition by lateral gene transfer.

Examination of the phylogeny of the conserved genes in the O antigen/CPS of *V. cholerae* revealed that lateral gene transfer in this region not only involved the same species, but also involved different species, and even different genera. Therefore, the large gene pool plus the greatly increased mobility in the O antigen/CPS region may contribute to the large number of serogroups observed in *V. cholerae*. Furthermore, emergence of novel strains is expected to continue.

My study supports my second hypothesis that horizontal gene transfer plays a critical role in the evolution of *V. cholerae*, especially in generating the diversity of surface polysaccharides. Lateral gene transfer contributes to the emergence of new strains by acquiring new polysaccharide genes as well as virulence genes. In turn, it makes the genome of *V. cholerae* dynamic. I studied the genome of *V. cholerae* NRT36S, as discussed in chapter 3. This was the first study to directly compare two genomes of *V. cholerae*. This study gave a direct evidence of the types and extent of

variation in the genome of *V. cholerae*. Eleven percent of the sequences were specific to the specific *V. cholerae* genomes. Many of those sequences appeared as genomic islands that may have arisen from lateral gene transfer. Insertion of bacteria phage and homologous recombination may be major mechanisms of the lateral gene transfer responsible for the rise of the genomic islands.

Lateral gene transfer has resulted in two entirely different pathogenesis mechanisms in the epidemic and NAG-ST pathogenic *V. cholerae*. In epidemic *V. cholerae*, the major virulence factors are CT and TCP. CT and TCP, as well as other genes related to pathogenesis, including VPI-1, VPI-II, VSP-1 and VSP-II are in different gene clusters present specifically in epidemic strains and were not found in NRT36S. In contrast, NRT36S possessed multiple different potential virulence factors, at different locations in the genome. NAG-ST has been shown to be an important toxin in non-O1 *V. cholerae* (41, 144). We located the NAG-ST gene in the super-integron region (SI). This is the first functional virulence gene identified to the SI, strongly suggests the role of SI in acquiring exogenous genes. In addition to NAG-ST, we identified a type III secretion system, which could be a potential virulence factor. There are several other potential toxins, including an exotoxin A and two RTX toxins. One of the RTX toxins is present in both the epidemic strain and NRT36S, while the other is specific to NRT36S. The role of these potential toxins remains to be investigated. In NRT36S, I also identified 4 pili systems as potential colonization factors. A type 1 pilus and a mating pilus are present only in NRT36S; two type IV pili were conserved in both the epidemic strain and NRT36S. Altogether, these potential toxins and colonization factors suggest the pathogenic mechanism in non-O1 *V. cholerae* is distinct from that of the epidemic

strains. My analysis of the genome of NRT36S has advanced our understanding of pathogenic mechanism of non-O1 *V. cholerae*.

My analysis suggests that phages were tightly linked to pathogenesis. Most of the virulence genes specific to O1 N16961 were closely related to phage. In NRT36S, phage related genes and sequences made up 25% of its genome specific genes. None of the virulence factors found in the epidemic O1 and O139 strains was present in the O31 pathogenic strain NRT36S. Instead, I have identified new putative virulence factors in non-O1 *V. cholerae*. Both pathogenic *V. cholerae*, O1 El Tor and O31 NRT36S, share many biochemical and genetic properties, and both have multiple virulence factors, but surprisingly almost none of the virulence factors overlap. This suggests that the ancestral strain of *V. cholerae* was an environmental organism and that toxigenic strains arose from multiple independent acquisitions of virulence factors through horizontal gene transfer events.

In summary, studies presented in my dissertation have extended our knowledge in understanding the evolution of *V. cholerae*. The genome of *V. cholerae* is dynamic. It may acquire new pathogenic features, antigens and capsules by lateral gene transfer. As we showed with NRT36S and A5 and has been shown for O1 El Tor, O139 and O37, a virulent strain may acquire genes for a new or modified polysaccharide antigen in a single simple recombination event and become a new serogroup capable of causing disease even in those exposed to another toxigenic strain of *V. cholerae*. Lateral gene transfer has played a critical role in maintaining the polysaccharide diversity, as well as acquiring virulence factors, and hence the emergence of new epidemic strains. I would

expect novel epidemic strains, with a similar toxigenic background to the current ones, but different serogroups, continues to emerge.

CONCLUDING REMARKS AND FUTURE PROSPECTS

Lateral gene transfer plays a critical role in the evolution of *V. cholerae* and eventually leads this species to pathogenesis. My preliminary analysis of the pathogenesis genes provides new clues and directions for the future study into pathogenesis. By examining the genome of NRT36S, I identified several potential virulence factors, which might be critical in the pathogenic mechanism for non-O1 *V. cholerae*. Among them, NAG-ST was suggested to be an important toxin (41). But others were not described before for NRT36S. A TTSS system, an exotoxin A gene, and a RTX toxin were found. These could be important virulence factors, but functional studies will need to be done in order to be conclusive. Several pili systems were also identified which could be important for colonization of the human host. Although colonization was found to be required for the pathogenesis of NRT36S, no pilus genes were identified previously (41). To confirm a role of these pili in pathogenesis, functional studies including specific gene knockouts and animal models are required.

Bacteria often employ similar strategies for polysaccharide biogenesis. From the genes present in the capsule biogenesis region in *V. cholerae* NRT36S, I speculate that the capsule synthesis pathway in *V. cholerae* is similar to group 1 *E. coli*, although there are also significant differences. Capsule biogenesis in *E. coli* depends on either *wzx*/*wzy* in conjugation with *wza, wzb*, *wzc* pathway or *wzm*/*wzt* pathway. All of these genes are present in the capsule/LPS region of NRT36S, consistent with *V. cholerae* using similar

mechanism for capsule exportation. However, never there has been a case reported in the other species that all 7 of capsule exportation genes are present in a single isolate and some of these genes may not be involved in polysaccharide transport in *V. cholerae*. The capsule repeating subunit in NRT36S is a complex heteropolymer with four sugars. Polysaccharides consisting of heteropolymers in other species have been associated with the *wzx*/*wzy* pathway while polysaccharides consisting of homopolymers use the *wzm*/*wzt* pathways. The structure of NRT36S suggests it may depend on the *wzx*/*wzy* pathway. The presence of *wza* further suggests it is a group 1 type capsule. On the other hand, *wzm* seems to be involved in the capsule exportation in NRT36S since mutation in this gene decreased the capsule amount. Therefore, capsule in NRT36S may use an exportation pathway involving the combination of *wzx/wzy* and *wzm/wzt* pathways. *Wzm* is conserved in O-antigen or capsule biogenesis region of all the *V. cholerae* examined so far, including O139, which has both a capsule and O-antigen. The precise mechanism of understanding LPS/CPS exportation in *V. cholerae* NRT36S remains unclear.

The pandemic *V. cholerae* serogroup O1 does not have a capsule while most the environmental strains have capsules. O1 *V. cholerae* could have evolved from an acapsular ancestral strain; alternatively it could have evolved from an encapsulated strain, which later lost its capsule as it became a human pathogen. The fact that serogroup O1 has been abundant for centuries may suggest that a capsule is not beneficial for the spread of this pathogen via human host. *V. cholerae* O139 emerged in 1992 and caused epidemic in Bangladesh, but subsided for more than a decade and then re-emerged around 2004 (7). More recent isolates (2002) of O139 have less capsular material than the isolates from 1992-1993 (170). This evidence supports the idea that a capsule is not

beneficial for human adapted strains. Production of CPS requires energy; perhaps the acapsular variant could out-compete the encapsulated variants in human intestine because of reduced energy needs. However, this hypothesis will need to be tested.

In the natural environment, capsule may be more beneficial for the growth of *V. cholerae*. The majority of non-O1 *V. cholerae* isolates are encapsulated. Bacteria won't waste that much energy if a capsule is not advantageous to their survival.

The role of capsule in the natural aquatic environment needs more investigation. Capsule is a virulence factor for non-O1 *V. cholerae* infecting humans. However, infection of humans by non-O1 *V. cholerae* is an occasional event. Capsules may help bacteria attach to the surface of phytoplankton and zooplankton. But *V. cholerae* also lives in free planktonic form. Capsules retain a lot of water and prevent bacteria from desiccation. It may be possible that capsules provide a microenvironment, which serves as an osmotic buffering system for the cell and help cells to maintain their structure and regular metabolism. Capsules may also protect bacterial from bacteriophage infection or grazing by predators.

Many questions remain to be answered in the genesis and regulation of polysaccharide genes in *V. cholerae*. In both cases in which a capsule has been studied, i. e., O139 AND NRT36S, the CPS/O-antigen share the same biogenesis region. But these studies should be extended to include other serogroups to confirm our conclusion. Is there a generalized theme in the biogenesis of polysaccharide in *V. cholerae*, or do different mechanisms existing in this species, similar to *E. coli*? These questions remain to be answered by extending the study to include more serogroups.

One interesting question that arose while I was studying the surface polysaccharide is the relationships and the interactions between different polysaccharide forms. The interaction between capsule and exopolysaccharide is especially interesting because both of them are loose surface entities. The capsule structures in *V. cholerae* O139 and NRT36S are different. I anticipate there would be many capsule structures in *V. cholerae*, considering the nature of a capsule from the other species, and the mobility and diversity of capsule genes of *V. cholerae*. On the other hand, the exopolysaccharide gene cluster shares high similarity between O1 and NRT36S, suggesting their structures may be very similar. The exopolysaccharide expressed in the rugose phenotype and promotes biofilm formation. Biofilm is a survival strategy when bacteria encounter stress conditions. Between epidemic periods, O1 *V. cholerae* exists as either planktonic VBNC form or in biofilms. I would like to propose that non-O1 *V. cholerae* switches between encapsulated form and rugose form, one as a planktonic life style, one as surface attached or biofilm life style. In this hypothesis, capsule should not co-express with exopolysaccharide at the same time. Under this assumption, bacteria will need a regulation system to control the different expression of these polysaccharides.

The interesting finding of CPS/O-antigen biogenesis sharing the same gene cluster and even sharing some common genes also raised the question of regulation. Understanding of the regulatory mechanisms underlying polysaccharide biosynthesis will lead to better understanding of the general regulation of surface polysaccharide expression in *V. cholerae*. Functional studies with microarrays could be conducted to understand the regulation system. The CPS/O-antigen biogenesis was a complicated system in O31 and some genes appeared to be redundant. My examination of NRT36S

and A5 differentiate, to a limited extent, the genes for CPS and genes for O-antigen. But most genes remained to be clarified. Specific gene knockouts can be conducted to resolve the gene functions.

The importance and impact of surface polysaccharide diversity in *V. cholerae* has been underappreciated until this study. As a result, the study of polysaccharide biogenesis in *V. cholerae* is far from comprehensive compared to *E. coli*. My data suggests that surface polysaccharides, including exopolysaccharide, O-antigen and CPS are related to each other. My study provides the first insight to a non O-antigen capsule in *V. cholerae*. However, my study is limited to only one isolate and I can only speculate as to the general scheme of capsule biogenesis in this diverse species. Capsule typing is well established in many bacteria, including *E. coli*, *Streptococcus pneumoniae* and *Neisseria meningitidis*. In contrast, capsule typing has not been applied to *V. cholerae*. Currently, we do not have complete knowledge of the capsule diversity in this species. Our preliminary data open many doors to future research, which will ultimately lead us to understanding the function, biogenesis regulation of capsule in *V. cholerae*, and how they related to the capsules in other species, and furthermore to understanding the function and regulation of all forms of polysaccharides.

REFERENCES

1. Jiang, S. C., Louis, V., Choopun, N., Sharma, A., Huq, A. & Colwell, R. R. (2000) *Appl Environ Microbiol* **66,** 140-7.
2. Kaper, J., Lockman, H., Colwell, R. R. & Joseph, S. W. (1979) *Appl Environ Microbiol* **37,** 91-103.
3. Louis, V. R., Russek-Cohen, E., Choopun, N., Rivera, I. N., Gangle, B., Jiang, S. C., Rubin, A., Patz, J. A., Huq, A. & Colwell, R. R. (2003) *Appl Environ Microbiol* **69,** 2773-85.
4. Roszak, D. B. & Colwell, R. R. (1987) *Microbiol Rev* **51,** 365-79.
5. Glass, R. I., Becker, S., Huq, M. I., Stoll, B. J., Khan, M. U., Merson, M. H., Lee, J. V. & Black, R. E. (1982) *Am J Epidemiol* **116,** 959-70.
6. Huq, A., Colwell, R. R., Rahman, R., Ali, A., Chowdhury, M. A., Parveen, S., Sack, D. A. & Russek-Cohen, E. (1990) *Appl Environ Microbiol* **56,** 2370-3.
7. Alam, M., Sultana, M., Nair, G. B., Sack, R. B., Sack, D. A., Siddique, A. K., Ali, A., Huq, A. & Colwell, R. R. (2006) *Appl Environ Microbiol* **72,** 2849-55.
8. Colwell, R. R. & Huq, A. (1994) *Vibrios in the environment: viable but non-culturable Vibrio cholerae.* (American Society for Microbiology, Washington, D. C.).
9. Faruque, S. M., Biswas, K., Udden, S. M., Ahmad, Q. S., Sack, D. A., Nair, G. B. & Mekalanos, J. J. (2006) *Proc Natl Acad Sci U S A* **103,** 6350-5.
10. Huq, A., Small, E. B., West, P. A., Huq, M. I., Rahman, R. & Colwell, R. R. (1983) *Appl Environ Microbiol* **45,** 275-83.
11. Islam, M. S., Drasar, B. S. & Sack, R. B. (1994) *J Diarrhoeal Dis Res* **12,** 87-96.

12. Tamplin, M. L., Gauzens, A. L., Huq, A., Sack, D. A. & Colwell, R. R. (1990) *Appl Environ Microbiol* **56,** 1977-80.

13. Matz, C., McDougald, D., Moreno, A. M., Yung, P. Y., Yildiz, F. H. & Kjelleberg, S. (2005) *Proc Natl Acad Sci U S A* **102,** 16819-24.

14. Yamai, S., Okitsu, T., Shimada, T. & Katsube, Y. (1997) *Kansenshogaku Zasshi* **71,** 1037-45.

15. Bik, E. M., Bunschoten, A. E., Gouw, R. D. & Mooi, F. R. (1995) *Embo J* **14,** 209-16.

16. Stroeher, U. H. & Manning, P. A. (1999) *Genetics of Vibrio cholerae O1 and O139 surface polysaccharides* (CRC Press, Boca Raton, Fl.).

17. Johnson, J. A., Salles, C. A., Panigrahi, P., Albert, M. J., Wright, A. C., Johnson, R. J. & Morris, J. G., Jr. (1994) *Infect Immun* **62,** 2108-10.

18. Li, M., Shimada, T., Morris, J. G., Jr., Sulakvelidze, A. & Sozhamannan, S. (2002) *Infect Immun* **70,** 2441-53.

19. Kaper, J. B., Morris, J. G., Jr. & Levine, M. M. (1995) *Clin Microbiol Rev* **8,** 48-86.

20. Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Umayam, L., Gill, S. R., Nelson, K. E., Read, T. D., Tettelin, H., Richardson, D., Ermolaeva, M. D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R. D., Nierman, W. C., White, O., Salzberg, S. L., Smith, H. O., Colwell, R. R., Mekalanos, J. J., Venter, J. C. & Fraser, C. M. (2000) *Nature* **406,** 477-83.

21. Chen, C. Y., Wu, K. M., Chang, Y. C., Chang, C. H., Tsai, H. C., Liao, T. L., Liu, Y. M., Chen, H. J., Shen, A. B., Li, J. C., Su, T. L., Shao, C. P., Lee, C. T., Hor, L. I. & Tsai, S. F. (2003) *Genome Res* **13,** 2577-87.

22. Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, M., Nakano, M., Yamashita, A., Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H., Hattori, M. & Iida, T. (2003) *Lancet* **361,** 743-9.

23. Beltran, P., Delgado, G., Navarro, A., Trujillo, F., Selander, R. K. & Cravioto, A. (1999) *J Clin Microbiol* **37,** 581-90.

24. Byun, R., Elbourne, L. D., Lan, R. & Reeves, P. R. (1999) *Infect Immun* **67,** 1116-24.

25. Garg, P., Aydanian, A., Smith, D., J. Glenn M, J., Nair, G. B. & Stine, O. C. (2003) *Emerg Infect Dis* **9,** 810-4.

26. Popovic, T., Bopp, C., Olsvik, O. & Wachsmuth, K. (1993) *J Clin Microbiol* **31,** 2474-82.

27. Stroeher, U. H., Parasivam, G., Dredge, B. K. & Manning, P. A. (1997) *J Bacteriol* **179,** 2740-7.

28. Karaolis, D. K., Lan, R. & Reeves, P. R. (1995) *J Bacteriol* **177,** 3191-8.

29. Farfan, M., Minana-Galbis, D., Fuste, M. C. & Loren, J. G. (2002) *J Bacteriol* **184,** 1304-13.

30. Karaolis, D. K., Johnson, J. A., Bailey, C. C., Boedeker, E. C., Kaper, J. B. & Reeves, P. R. (1998) *Proc Natl Acad Sci U S A* **95,** 3134-9.

31. Karaolis, D. K., Lan, R., Kaper, J. B. & Reeves, P. R. (2001) *Infect Immun* **69,** 1947-52.

32. Waldor, M. K. & Mekalanos, J. J. (1996) *Science* **272,** 1910-4.

33. Herrington, D. A., Hall, R. H., Losonsky, G., Mekalanos, J. J., Taylor, R. K. & Levine, M. M. (1988) *J Exp Med* **168,** 1487-92.

34. Taylor, R. K., Miller, V. L., Furlong, D. B. & Mekalanos, J. J. (1987) *Proc Natl Acad Sci U S A* **84,** 2833-7.

35. Gill, D. M. (1976) *Biochemistry* **15,** 1242-8.

36. Lonnroth, I. & Holmgren, J. (1973) *J Gen Microbiol* **76,** 417-27.

37. Holmgren, J., Lonnroth, I. & Svennerholm, L. (1973) *Scand J Infect Dis* **5,** 77-8.

38. Holmgren, J. & Svennerholm, A. M. (1977) *J Infect Dis* **136 Suppl,** S105-12.

39. Van Heyningen, W. E., Carpenter, C. C., Pierce, N. F. & Greenough, W. B., 3rd (1971) *J Infect Dis* **124,** 415-8.

40. O'Shea, Y. A., Reen, F. J., Quirke, A. M. & Boyd, E. F. (2004) *J Clin Microbiol* **42,** 4657-71.

41. Morris, J. G., Jr., Takeda, T., Tall, B. D., Losonsky, G. A., Bhattacharya, S. K., Forrest, B. D., Kay, B. A. & Nishibuchi, M. (1990) *J Clin Invest* **85,** 697-705.

42. Chitnis, D. S., Sharma, K. D. & Kamat, R. S. (1982) *J Med Microbiol* **15,** 53-61.

43. Freter, R. & Jones, G. W. (1976) *Infect Immun* **14,** 246-56.

44. Cash, R. A., Music, S. I., Libonati, J. P., Snyder, M. J., Wenzel, R. P. & Hornick, R. B. (1974) *J Infect Dis* **129,** 45-52.

45. Levine, M. M., Nalin, D. R., Craig, J. P., Hoover, D., Bergquist, E. J., Waterman, D., Holley, H. P., Hornick, R. B., Pierce, N. P. & Libonati, J. P. (1979) *Trans R Soc Trop Med Hyg* **73,** 3-9.

46. Aydanian, A. (2004) (Univerisity of Maryland, Baltimore, Baltimore), Vol. Doctor of Philosophy.

47. Comstock, L. E., Johnson, J. A., Michalski, J. M., Morris, J. G., Jr. & Kaper, J. B. (1996) *Mol Microbiol* **19,** 815-26.

48. Manning, P. A., Heuzenroeder, M. W., Yeadon, J., Leavesley, D. I., Reeves, P. R. & Rowley, D. (1986) *Infect Immun* **53,** 272-7.

49. Yamasaki, S., Shimizu, T., Hoshino, K., Ho, S. T., Shimada, T., Nair, G. B. & Takeda, Y. (1999) *Gene* **237,** 321-32.

50. Johnson, J. A., Panigrahi, P. & Morris, J. G., Jr. (1992) *Infect Immun* **60,** 864-9.

51. Griffith, F. (1928) *Journal of Hygine* **27,** 113-159.

52. Powell, J. L. (1999) *Clin Lab Med* **19,** 537-52, vi.

53. Strom, M. S. & Paranjpye, R. N. (2000) *Microbes Infect* **2,** 177-88.

54. Wright, A. C., Simpson, L. M., Oliver, J. D. & Morris, J. G., Jr. (1990) *Infect Immun* **58,** 1769-73.

55. Yoshida, S., Ogawa, M. & Mizuguchi, Y. (1985) *Infect Immun* **47,** 446-51.

56. Jesudason, M. V., Cherian, A. M. & John, T. J. (1993) *Lancet* **342,** 431.

57. Whitfield, C. (2006) *Annu Rev Biochem* **75,** 39-68.

58. MacLachlan, P. R., Keenleyside, W. J., Dodgson, C. & Whitfield, C. (1993) *J Bacteriol* **175,** 7515-22.

59. Orskov, I., Orskov, F., Jann, B. & Jann, K. (1977) *Bacteriol Rev* **41,** 667-710.

60. Barrett, B., Ebah, L. & Roberts, I. S. (2002) *Curr Top Microbiol Immunol* **264,** 137-55.

61. Cieslewicz, M. & Vimr, E. (1996) *J Bacteriol* **178,** 3212-20.

62. Simpson, D. A., Hammarton, T. C. & Roberts, I. S. (1996) *J Bacteriol* **178,** 6466-74.

63. Whitfield, C. & Paiment, A. (2003) *Carbohydr Res* **338,** 2491-502.

64. Whitfield, C. & Roberts, I. S. (1999) *Mol Microbiol* **31,** 1307-19.

65. Henrichsen, J. (1995) *J Clin Microbiol* **33,** 2759-62.

66. Llull, D., Lopez, R. & Garcia, E. (2001) *Curr Mol Med* **1,** 475-91.

67. Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M. A., Samuel, G., Skovsted, I. C., Kaltoft, M. S., Barrell, B., Reeves, P. R., Parkhill, J. & Spratt, B. G. (2006) *PLoS Genet* **2,** e31.

68. White, P. B. (1938) *J. Pathol. Bacteriol* **46,** 1-6.

69. Frosch, M., Weisgerber, C. & Meyer, T. F. (1989) *Proc Natl Acad Sci U S A* **86,** 1669-73.

70. Roberts, I. S. (1996) *Annu Rev Microbiol* **50,** 285-315.

71. Wright, A. C., Powell, J. L., Kaper, J. B. & Morris, J. G., Jr. (2001) *Infect Immun* **69,** 6893-901.

72. Smith, A. B. & Siebeling, R. J. (2003) *Infect Immun* **71,** 1091-7.

73. Johnson, J. A., Joseph, A., Panigrahi, P. & Morris, J. G. (1992) in *American society for microbiology annual meeting*, New Orleans, Louisiana).

74. Waldor, M. K., Colwell, R. & Mekalanos, J. J. (1994) *Proc Natl Acad Sci U S A* **91,** 11388-92.

75. Preston, L. M., Xu, Q., Johnson, J. A., Joseph, A., Maneval, D. R., Jr., Husain, K., Reddy, G. P., Bush, C. A. & Morris, J. G., Jr. (1995) *J Bacteriol* **177,** 835-8.

76. Cox, A. D., Brisson, J. R., Varma, V. & Perry, M. B. (1996) *Carbohydr Res* **290,** 43-58.

77. Yildiz, F. H., Liu, X. S., Heydorn, A. & Schoolnik, G. K. (2004) *Mol Microbiol* **53,** 497-515.

78. Yildiz, F. H. & Schoolnik, G. K. (1999) *Proc Natl Acad Sci U S A* **96,** 4028-33.

79. Ali, A., Mahmud, Z. H., Morris, J. G., Jr., Sozhamannan, S. & Johnson, J. A. (2000) *Infect Immun* **68,** 6857-64.

80. McCarter, L. L. (1998) *J Bacteriol* **180,** 3166-73.

81. Guvener, Z. T. & McCarter, L. L. (2003) *J Bacteriol* **185,** 5431-41.

82. Hammerschmidt, S., Hilse, R., van Putten, J. P., Gerardy-Schahn, R., Unkmeir, A. & Frosch, M. (1996) *Embo J* **15,** 192-8.

83. Hammerschmidt, S., Muller, A., Sillmann, H., Muhlenhoff, M., Borrow, R., Fox, A., van Putten, J., Zollinger, W. D., Gerardy-Schahn, R. & Frosch, M. (1996) *Mol Microbiol* **20,** 1211-20.

84. Waite, R. D., Struthers, J. K. & Dowson, C. G. (2001) *Mol Microbiol* **42,** 1223-32.

85. Linton, D., Gilbert, M., Hitchen, P. G., Dell, A., Morris, H. R., Wakarchuk, W. W., Gregson, N. A. & Wren, B. W. (2000) *Mol Microbiol* **37,** 501-14.

86. Abraham, J. M., Freitag, C. S., Clements, J. R. & Eisenstein, B. I. (1985) *Proc Natl Acad Sci U S A* **82,** 5724-7.

87. Low, D., Braaten, B. A. & Vvan der Woude, M. (1996) *Fimbriae* (ASM press, Washington, D. C.).

88. Fussenegger, M., Rudel, T., Barten, R., Ryll, R. & Meyer, T. F. (1997) *Gene* **192,** 125-34.

89. Karlyshev, A. V., Linton, D., Gregson, N. A. & Wren, B. W. (2002) *Microbiology* **148,** 473-80.

90. Ottemann, K. M. & Lowenthal, A. C. (2002) *Infect Immun* **70,** 1984-90.

91. Dempsey, J. A., Litaker, W., Madhure, A., Snodgrass, T. L. & Cannon, J. G. (1991) *J Bacteriol* **173,** 5476-86.

92. Stern, A. & Meyer, T. F. (1987) *Mol Microbiol* **1,** 5-12.

93. Dworkin, J. & Blaser, M. J. (1997) *Mol Microbiol* **26,** 433-40.

94. van Ham, S. M., van Alphen, L., Mooi, F. R. & van Putten, J. P. (1993) *Cell* **73,** 1187-96.

95. de Vries, N., Duinsbergen, D., Kuipers, E. J., Pot, R. G., Wiesenekker, P., Penn, C. W., van Vliet, A. H., Vandenbroucke-Grauls, C. M. & Kusters, J. G. (2002) *J Bacteriol* **184,** 6615-23.

96. Banerjee, A., Wang, R., Supernavage, S. L., Ghosh, S. K., Parker, J., Ganesh, N. F., Wang, P. G., Gulati, S. & Rice, P. A. (2002) *J Exp Med* **196,** 147-62.

97. van der Ende, A., Hopman, C. T. & Dankert, J. (2000) *Infect Immun* **68,** 6685-90.

98. Willems, R., Paul, A., van der Heide, H. G., ter Avest, A. R. & Mooi, F. R. (1990) *Embo J* **9,** 2803-9.

99. Hagblom, P., Segal, E., Billyard, E. & So, M. (1985) *Nature* **315,** 156-8.

100. van der Woude, M. W. & Baumler, A. J. (2004) *Clin Microbiol Rev* **17,** 581-611, table of contents.

101. Klemm, K. W. (1986) *Clin Orthop Relat Res***,** 174-81.

102. Braaten, B. A., Nou, X., Kaltenbach, L. S. & Low, D. A. (1994) *Cell* **76,** 577-88.

103. Nicholson, B. & Low, D. (2000) *Mol Microbiol* **35,** 728-42.

104. Kotetishvili, M., Stine, O. C., Chen, Y., Kreger, A., Sulakvelidze, A., Sozhamannan, S. & Morris, J. G., Jr. (2003) *J Clin Microbiol* **41,** 2191-6.

105. Chowdhury, N. R., Stine, O. C., Morris, J. G. & Nair, G. B. (2004) *J Clin Microbiol* **42,** 1280-2.

106. Li, M., Kotetishvili, M., Chen, Y. & Sozhamannan, S. (2003) *Appl Environ Microbiol* **69,** 1728-38.

107. Bhaskaran, K. (1959) *Indian J Med Res* **47,** 253-60.

108. Hedges, R. W., Vialard, J. L., Pearson, N. J. & O'Grady, F. (1977) *Antimicrob Agents Chemother* **11,** 585-8.

109. Prescott, L. M., Datta, A. & Datta, G. C. (1968) *Bull World Health Organ* **39,** 971-3.

110. Redfield, R. J. (2001) *Nat Rev Genet* **2,** 634-9.

111. Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C. Y. & Schoolnik, G. K. (2005) *Science* **310,** 1824-7.

112. Stahl, F. W., Crasemann, J. M. & Stahl, M. M. (1975) *J Mol Biol* **94,** 203-12.

113. Taylor, A. F., Schultz, D. W., Ponticelli, A. S. & Smith, G. R. (1985) *Cell* **41,** 153-63.

114. Redfield, R. J., Findlay, W. A., Bosse, J., Kroll, J. S., Cameron, A. D. & Nash, J. H. (2006) *BMC Evol Biol* **6,** 82.

115. Sourice, S., Biaudet, V., El Karoui, M., Ehrlich, S. D. & Gruss, A. (1998) *Mol Microbiol* **27,** 1021-9.

116. Bakkali, M., Chen, T. Y., Lee, H. C. & Redfield, R. J. (2004) *Proc Natl Acad Sci U S A* **101,** 4513-8.

117. Goodman, S. D. & Scocca, J. J. (1988) *Proc Natl Acad Sci U S A* **85,** 6982-6.

118. Smith, H. O., Gwinn, M. L. & Salzberg, S. L. (1999) *Res Microbiol* **150,** 603-16.

119. Stokes, H. W. & Hall, R. M. (1989) *Mol Microbiol* **3,** 1669-83.

120. Mazel, D. (2006) *Nat Rev Microbiol* **4,** 608-20.

121. Hochhut, B., Lotfi, Y., Mazel, D., Faruque, S. M., Woodgate, R. & Waldor, M. K. (2001) *Antimicrob Agents Chemother* **45,** 2991-3000.

122. Mazel, D., Dychinco, B., Webb, V. A. & Davies, J. (1998) *Science* **280,** 605-8.

123. Rowe-Magnus, D. A., Guerout, A. M. & Mazel, D. (2002) *Mol Microbiol* **43,** 1657-69.

124. Sozhamannan, S., Deng, Y. K., Li, M., Sulakvelidze, A., Kaper, J. B., Johnson, J. A., Nair, G. B. & Morris, J. G., Jr. (1999) *Infect Immun* **67,** 5033-40.

125. Johnson, J. A., Joseeph, A., Panigrahi, P. & Morris, J. G. (1992) in *American society for microbiology annual meeting*, New Orleans, Louisiana).

126. Cox, A. D. & Perry, M. B. (1996) *Carbohydr Res* **290,** 59-65.

127. Knirel, Y. A., Widmalm, G., Senchenkova, S. N., Jansson, P. E. & Weintraub, A. (1997) *Eur J Biochem* **247,** 402-10.

128. Simpson, L. M., White, V. K., Zane, S. F. & Oliver, J. D. (1987) *Infect Immun* **55,** 269-72.

129. Chatzidaki-Livanis, M., Jones, M. K. & Wright, A. C. (2006) *J Bacteriol* **188,** 1987-98.

130. Bergstrom, N., Nair, G. B., Weintraub, A. & Jansson, P. E. (2002) *Carbohydr Res* **337,** 813-7.

131. Panigrahi, P. & J.G., M. (1993) in *93rd General Meeting, American Society for Microbiology*, Atlanta, Georgia).

132. Hobbs, M. & Reeves, P. R. (1994) *Mol Microbiol* **12,** 855-6.

133. Li, Q., Hobbs, M. & Reeves, P. R. (2003) *Microbiology* **149,** 2463-74.

134. Cserzo, M., Bernassau, J. M., Simon, I. & Maigret, B. (1994) *J Mol Biol* **243,** 388-96.

135. Reid, A. N. & Whitfield, C. (2005) *J Bacteriol* **187,** 5470-81.

136. Feng, L., Senchenkova, S. N., Yang, J., Shashkov, A. S., Tao, J., Guo, H., Cheng, J., Ren, Y., Knirel, Y. A., Reeves, P. R. & Wang, L. (2004) *J Bacteriol* **186,** 4510-9.

137. Dolan-Livengood, J. M., Miller, Y. K., Martin, L. E., Urwin, R. & Stephens, D. S. (2003) *J Infect Dis* **187,** 1616-28.

138. Hammerschmidt, S., Birkholz, C., Zahringer, U., Robertson, B. D., van Putten, J., Ebeling, O. & Frosch, M. (1994) *Mol Microbiol* **11,** 885-96.

139. Hashimoto, Y., Li, N., Yokoyama, H. & Ezaki, T. (1993) *J Bacteriol* **175,** 4456-65.

140. Kroll, J. S., Zamze, S., Loynds, B. & Moxon, E. R. (1989) *J Bacteriol* **171,** 3343-7.

141. Swartley, J. S., Ahn, J. H., Liu, L. J., Kahler, C. M. & Stephens, D. S. (1996) *J Bacteriol* **178,** 4052-9.

142. Kondo, S. & Hisatsune, K. (1989) *Microbiol Immunol* **33,** 641-8.

143. Reeves, P. P. & Wang, L. (2002) *Curr Top Microbiol Immunol* **264,** 109-35.

144. Arita, M., Takeda, T., Honda, T. & Miwatani, T. (1986) *Infect Immun* **52,** 45-9.

145. de Lorenzo, V., Herrero, M., Jakubzik, U. & Timmis, K. N. (1990) *J Bacteriol* **172,** 6568-72.

146. Herrero, M., de Lorenzo, V. & Timmis, K. N. (1990) *J Bacteriol* **172,** 6557-67.

147. Martin, V. J. & Mohn, W. W. (1999) *J Microbiol Methods* **35,** 163-6.

148. Ewing, B. & Green, P. (1998) *Genome Res* **8,** 186-94.

149. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998) *Genome Res* **8,** 175-85.

150. Gordon, D. (2004) *Viewing and Editing Assembled Sequences Using Consed* (John Wilery & Co.,, New York).

151. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro,

K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. (2005) *Nature* **437,** 376-80.

152. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J Mol Biol* **215,** 403-10.

153. Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. (1998) *Nucleic Acids Res* **26,** 544-8.

154. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000) *Bioinformatics* **16,** 944-5.

155. Stine, O. C., Sozhamannan, S., Gou, Q., Zheng, S., Morris, J. G., Jr. & Johnson, J. A. (2000) *Infect Immun* **68,** 7180-5.

156. Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L. (2004) *Genome Biol* **5,** R12.

157. Jermyn, W. S. & Boyd, E. F. (2002) *Microbiology* **148,** 3681-93.

158. Galen, J. E., Ketley, J. M., Fasano, A., Richardson, S. H., Wasserman, S. S. & Kaper, J. B. (1992) *Infect Immun* **60,** 406-15.

159. Ogawa, A., Kato, J., Watanabe, H., Nair, B. G. & Takeda, T. (1990) *Infect Immun* **58,** 3325-9.

160. Boyd, D., Peters, G. A., Cloeckaert, A., Boumedine, K. S., Chaslus-Dancla, E., Imberechts, H. & Mulvey, M. R. (2001) *J Bacteriol* **183,** 5725-32.

161. Jonson, G., Holmgren, J. & Svennerholm, A. M. (1991) *Microb Pathog* **11,** 433-41.

162. Fullner, K. J. & Mekalanos, J. J. (1999) *Infect Immun* **67,** 1393-404.

163. Dziejman, M., Serruto, D., Tam, V. C., Sturtevant, D., Diraphat, P., Faruque, S. M., Rahman, M. H., Heidelberg, J. F., Decker, J., Li, L., Montgomery, K. T., Grills, G., Kucherlapati, R. & Mekalanos, J. J. (2005) *Proc Natl Acad Sci U S A* **102,** 3465-70.

164. Project, N. M. G. A. (2005) (National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA.

165. Krause, L., McHardy, A. C., Puhler, A., Stoye, J. & Meyer, F. (2006) *Nucleic Acids Research*.

166. Chen, Y., Bystricky, P., Adeyeye, J., Panigrahi, P., Ali, A., Johnson, J. A., Bush, C. A., Morris, J. G. & Stine, O. C. (2006).

167. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res* **25,** 3389-402.

168. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003) *Nucleic Acids Res* **31,** 3497-500.

169. Page, R. D. (1996) *Comput Appl Biosci* **12,** 357-8.

170. Qadri, F., Svennerholm, A. M., Shamsuzzaman, S., Bhuiyan, T. R., Harris, J. B., Ghosh, A. N., Nair, G. B., Weintraub, A., Faruque, S. M., Ryan, E. T., Sack, D. A. & Calderwood, S. B. (2005) *Infect Immun* **73,** 6577-83.

CURRICULUM VITA

Name: Yuansha Chen

Address: 3426 Orange Grove Ct.

Ellicott City, MD 21043 USA

Phone: (410) 480-0138 (home), (410) 706-1609 (lab)

e-mail: ychen@epi.umaryland.edu

Education:

Doctor of Philosophy	2003-2006
	Marine, Estuarine and Environmental Science
	Area of specialization: Environmental Molecular Biology/Biotechnology
	University of Maryland
	College Park, Maryland
Graduate Student	2001-2003
	University of Maryland
Master of Science	1995–1998
	Department of Biochemistry and Molecular Biology
	Oklahoma State University
	Stillwater, Oklahoma 74075-3035
Bachelor of Science	1989-1993
	Department of Biochemistry
	School of Life Sciences
	Zhongshan University (Sun Yat-Sen University)

Guangzhou, P. R. China 510275

Work Experiences:

Senior Research Assistant 2003-present

Department of Epidemiology and Preventive Medicine

University of Maryland, Baltimore

Senior Research Assistant 2000-2003

Genomic Core Facility, School of Medicine

University of Maryland, Baltimore

Senior Research Assistant 1998–2000

Center of Genetics for Asthma and Complex Diseases

School of Medicine

University of Maryland, Baltimore

Research Assistant 1995-1998

Department of Biochemistry and Molecular Biology

Oklahoma State University

Research Assistant 1993-1995

Zhuxin Biotech and Pharmaceutical Company

Zhuhai City, China

Publications and manuscripts:

Chen, Y., Bystricky, P., Adeyeye, J., Panigrahi, P., Ali, A., Johnson, J.A., Bush, C.A., Morris, J.G., & Stine, O.C. (submitted 2006).

Chen, Y., Johnson, J.A., Pusch, D.G., Morris, J.G., & Stine, O.C. (submitted 2006).

Chen, Y., Stine, O.C., Morris, J.G., & Johnson, J.A. (in preparation 2006).

Kotetishvili, M., Stine, O.C., Chen, Y., Kreger, A., Sulakvelidze, A., Sozhamannan, S., & Morris, J.G., Jr. (2003) *J Clin Microbiol* **41:** 2191-2196.

Li, M., Kotetishvili, M., Chen, Y., & Sozhamannan, S. (2003) *Appl Environ Microbiol* **69:** 1728-1738.

Davis, E.M., Chen, Y., Essenberg, M., & Pierce, M.L. (1998) *Plant Physiol*. **116:** 1192.

Abstracts:

Chen, Y., Bystricky, P., Adeyeye, J., Panigrahi, P., Ali, A., Johnson, J.A., Bush, C.A., Morris, J.G., & Stine, O.C. (2006). in *FASEB summer research conferences: microbial polysaccharides of medical, agricultural, and industrial importance.* Tucson, Arizona).

Stine, O.C., Kotetishvili, M., Chen, Y., Snyder, A., Turabelidze, D., Scott, J., Gombas, D.,. Morris, J. G, & Sulakvelidze, A. (2001) in *101st American society for microbiology general meeting*, Orlando, Florida).

Li, M, Chen, Y, Shimada, T, Johnson, J.A., Morris, J.G., Stine, O.C., & Sozhamannan, S (2001) in *101st American society for microbiology general meeting,* Orlando, Florida).