

## ABSTRACT

Title of Dissertation: THE POPULATION STRUCTURE OF *VIBRIO*  
*CHOLERA*E IN CHESAPEAKE BAY

Nipa Choopun, Doctor of Philosophy, 2004

Dissertation directed by: Professor Rita R. Colwell  
Marine Estuarine Environmental Sciences  
University of Maryland, College Park

The population structure of *V. cholerae* in Chesapeake Bay, United States, was analyzed and a simple procedure that employed only two biochemical tests in an abbreviated identification scheme, i.e., arginine dihydrolase and esculin hydrolysis, was developed. After enrichment in alkaline peptone water and selective plating on thiosulfate-citrate-bile salts-sucrose agar, all of the sucrose fermenting colonies identified by the two tests were confirmed as *V. cholerae* by the polymerase chain reaction. A non-redundant collection of 224 *V. cholerae* strains collected from 1998 through 2000 from Chesapeake Bay was analyzed for phenotype, genotype, and genomic fingerprint. A long-range enterobacterial repetitive intergenic consensus (ERIC) PCR method that was developed for this study provided fingerprint patterns that proved useful in assessing relatedness among the strains. Cluster analysis was done using three different methods

and revealed three well separated, primary clusters: Cluster A, consisting of the majority of the isolates, including the toxigenic type strain for the species and luminescent strains of *V. cholerae*; a smaller Cluster B, with the noteworthy characteristic of low *toxR* gene homology; and Cluster M, consisting exclusively of *V. mimicus*. Another primary cluster, Cluster C, was also identified as a single clone of a sucrose-negative, luminescent, *toxR*-negative strain. Because *V. mimicus* formed a separate cluster with similar distance values as demonstrated by *V. cholerae* Cluster B, using both ERIC fingerprinting and DNA-DNA hybridization, and had phenotypic and genotypic traits and 16S rDNA sequences similar to *V. cholerae*, it is concluded that all of the primary clusters observed in this study, including *V. mimicus*, belong to a single species, *V. cholerae*. *V. mimicus* was judged to be the highest risk group of the non-toxigenic isolates, in terms of susceptibility to CTX $\Phi$  and possession of the heat-stable enterotoxin gene (*stn*). Approximately 11% of the *V. cholerae* strains that lacked the toxin-coregulated pili (TCP) and 50% of the *V. mimicus* strains were found to be susceptible to CTX $\Phi$ . In addition to the *V. cholerae* strains in the toxigenic subcluster, luminescent *V. cholerae* strains represented the next highest risk, since 14% of the luminescent strains were susceptible to CTX $\Phi$  and 33% were *stn* positive.

THE POPULATION STRUCTURE OF *VIBRIO CHOLERAE* IN CHESAPEAKE BAY

by

Nipa Choopun

Thesis 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  
2004

Advisory Committee:

Professor Rita R. Colwell, Chair  
Associate Professor Anwar Huq  
Professor Sam W. Joseph  
Professor James B. Kaper  
Professor Estelle Russek-Cohen

©Copyright by

Nipa Choopun

2004

## ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to my advisor, Prof. Rita R. Colwell, for providing me the opportunity to study and work in her laboratory, for her hard work and dedication, and for insight into the greatness of science. I would like to thank all committee members, Prof. Sam W. Joseph, Prof. James B. Kaper, Prof. Estelle Russek-Cohen, and Ass.Prof. Anwar Huq, for dedication of their valuable time to thoroughly review this work and give valuable suggestions.

I am very grateful for financial support from the Ministry of Science and Technology, Thailand. I also would like to express my gratitude to the Colwell-Huq Laboratory for additional financial support in the time of need.

I also would like to thank all members of the Colwell-Huq Laboratory, past and present, for their continuous support and great friendship. In particular, I would like to thank Dr. Irma Rivera for inspiration and providing me with the confidence to publish my first scientific paper. Special thanks to Dr. Valérie Louis, my sampling buddy, for a great time in sample collection, many long nights in the lab, free yoga lessons and continuous support throughout. I also thank Chris Grim, for providing valuable luminescent data and help at a time of need. I am in deepest gratitude for my friend, Young-Gun Zo, for his great insights, suggestions, and dedications to make this work possible.

Last but not least, I am in deepest gratitude to my family and my husband, for their continuous support, patient and understanding throughout my study.

## TABLE OF CONTENTS

|  |     |
|--|-----|
| List of Tables .....   | vi  |
| List of Figures.....   | vii |
| Chapter 1 Introduction.....  | 1   |
| 1.1 Background for cholera and <i>Vibrio cholerae</i> .....  | 1   |
| 1.1.1 Cholera disease .....  | 1   |
| 1.1.2 History of cholera .....   | 2   |
| 1.1.3 <i>Vibrio cholerae</i> , the causative agent.....  | 3   |
| 1.1.4 Dynamics of epidemic strains of <i>V. cholerae</i> serogroups O1 and O139 .....                                    | 3   |
| 1.1.5 Pathogenicity and epidemiology of <i>V. cholerae</i> non O1/ non O139 .....  | 4   |
| 1.2 Diversity of <i>V. cholerae</i> .....  | 5   |
| 1.3 Significance of studying <i>V. cholerae</i> population structure and clonal<br>determination .....                   | 7   |
| 1.4 Census survey as a strategy to understand <i>V. cholerae</i> population structure .....                              | 8   |
| 1.5 Objectives of this study.....  | 9   |
| Chapter 2 A Simple Procedure for the Rapid Identification of <i>Vibrio cholerae</i> from the<br>Aquatic Environment..... | 10  |
| 2.1 Introduction.....  | 10  |
| 2.2 Materials and Methods.....   | 11  |
| 2.2.1 Isolation of presumptive <i>Vibrio cholerae</i> .....  | 11  |
| 2.2.2 Confirmation of the identities of presumptive <i>V. cholerae</i> isolates.....                                     | 13  |
| 2.2.3 Selection of the biochemical test .....  | 14  |
| 2.3 Results and Discussion .....   | 15  |
| 2.3.1 Selection of the effective tests .....   | 15  |
| 2.3.2 Evaluation of the simple biochemical test procedure .....  | 17  |
| 2.4 Conclusion .....   | 24  |
| Chapter 3 Isolation and Characterization of <i>V. cholerae</i> from the Chesapeake Bay ....                              | 25  |
| 3.1 Introduction.....  | 25  |
| 3.1.1 Direct detection of <i>V. cholerae</i> by PCR (Non-culture based method) .....                                     | 26  |
| 3.1.2 Isolation of <i>V. cholerae</i> from the aquatic environment .....   | 27  |
| 3.1.3 Phenotypic characterization of environmental isolates of <i>V. cholerae</i> .....                                  | 29  |
| 3.1.4 Genotypic characterization using gene probes .....   | 30  |
| 3.1.5 Objectives of this study.....  | 33  |
| 3.2 Materials and Methods.....   | 34  |
| 3.2.1 Sampling sites .....   | 34  |
| 3.2.2 Sample collection.....   | 38  |
| 3.2.3 Direct detection by PCR (Non-culture based method): .....  | 38  |
| 3.2.4 Isolation of <i>V. cholerae</i> .....  | 39  |
| 3.2.5 Confirmation of <i>V. cholerae</i> isolates .....  | 41  |
| 3.2.6 Characterization of <i>V. cholerae</i> isolates .....  | 42  |

|           |  |     |
|-----------|--|-----|
| 3.2.7     | Statistical analysis.....  | 46  |
| 3.3       | Results.....   | 48  |
| 3.3.1     | Direct detection of <i>V. cholerae</i> in water samples by PCR.....  | 48  |
| 3.3.2     | Isolation of <i>V. cholerae</i> from Chesapeake Bay .....  | 48  |
| 3.3.3     | Characterization of isolates.....  | 57  |
| 3.4       | Discussion.....  | 72  |
| 3.4.1     | Seasonality and annual variability .....   | 72  |
| 3.4.2     | Comparison of enrichment and direct plating methods.....   | 73  |
| 3.4.3     | Distribution of traits and their function .....  | 77  |
| 3.5       | Conclusions.....   | 82  |
| Chapter 4 | Determination of Clonality and Relatedness of <i>Vibrio cholerae</i> by Genomic Fingerprinting, using Long-Range Enterobacterial Repetitive Intergenic Consensus (ERIC) PCR..... | 84  |
| 4.1       | Introduction.....  | 84  |
| 4.1.1     | Importance of determining bacterial clonality.....   | 84  |
| 4.1.2     | Available methods for determining bacterial clonality.....   | 85  |
| 4.1.3     | A new method based on ERIC PCR fingerprinting.....   | 89  |
| 4.1.4     | Objectives of this study.....  | 90  |
| 4.2       | Materials and Methods.....   | 91  |
| 4.2.1     | Bacterial strains.....   | 91  |
| 4.2.2     | DNA preparation.....   | 91  |
| 4.2.3     | Long-range ERIC PCR conditions.....  | 94  |
| 4.2.4     | Gel electrophoresis.....   | 94  |
| 4.2.5     | Image analysis.....  | 95  |
| 4.2.6     | Verification of genome relatedness estimation by ERIC PCR fingerprinting.....  | 96  |
| 4.3       | Results.....   | 104 |
| 4.3.1     | Optimization of the ERIC PCR fingerprinting protocol.....  | 104 |
| 4.3.2     | Interpretation of fingerprint .....  | 108 |
| 4.3.3     | Reproducibility of the long-range ERIC PCR fingerprinting.....   | 111 |
| 4.3.4     | Identification of artificial clones.....   | 116 |
| 4.3.5     | Resolution power of the long-range ERIC PCR fingerprinting.....  | 119 |
| 4.3.6     | Comparison of genome relatedness estimation with DNA-DNA hybridization .....   | 123 |
| 4.4       | Discussion.....  | 129 |
| 4.4.1     | Selection of fingerprinting protocol.....  | 129 |
| 4.4.2     | Improvement in fingerprint banding patterns by adjustment of the electrophoresis setting.....  | 130 |
| 4.4.3     | Interpretation of fingerprint patterns using the band-based algorithm .....  | 132 |
| 4.4.4     | Reproducibility and resolution of the long-range ERIC PCR fingerprinting protocol .....  | 133 |
| 4.4.5     | Estimation of genomic relatedness .....  | 134 |
| 4.5       | Conclusion .....   | 136 |

|            |  |     |
|------------|--|-----|
| Chapter 5  | Analysis of the population structure of <i>V. cholerae</i> in Chesapeake Bay....       | 137 |
| 5.1        | Introduction.....  | 137 |
| 5.2        | Materials and Methods.....   | 139 |
| 5.2.1      | Bacterial strains.....   | 139 |
| 5.2.2      | Determination of operational taxonomic units (OTUs).....                               | 140 |
| 5.2.3      | Cluster analysis.....  | 141 |
| 5.3        | Results and discussion.....  | 142 |
| 5.3.1      | Structure of the <i>V. cholerae</i> population in Chesapeake Bay.....                  | 142 |
| 5.3.2      | Structure of <i>V. cholerae</i> populations outside Chesapeake Bay.....                | 161 |
| 5.3.3      | Inference of population structure from DNA-DNA hybridization.....                      | 170 |
| 5.3.4      | Identification of Clusters B and C and <i>V. mimicus</i> .....                         | 176 |
| 5.3.5      | Reassessment of the relationship between <i>V. mimicus</i> and <i>V. cholerae</i> .... | 182 |
| 5.4        | Conclusion.....  | 184 |
| Chapter 6  | The pathogenic risk of non-toxigenic <i>V. cholerae</i> .....                          | 186 |
| 6.1        | Introduction.....  | 186 |
| 6.1.1      | Objectives of this study.....  | 189 |
| 6.2        | Materials and Methods.....   | 190 |
| 6.2.1      | Bacterial strains employed in this study.....  | 190 |
| 6.2.2      | Preparation of the CTX-Km $\Phi$ particles.....  | 191 |
| 6.2.3      | Assay for CTX-Km $\Phi$ infection.....   | 191 |
| 6.2.4      | Analysis of CTX-Km $\Phi$ infected cells.....  | 192 |
| 6.3        | Results.....   | 193 |
| 6.3.1      | Production of CTX-Km $\Phi$ and its stability.....                                     | 193 |
| 6.3.2      | Natural occurrence of kanamycin resistance in recipient cells.....                     | 194 |
| 6.3.3      | Analysis of colonies containing CTX-Km $\Phi$ .....                                    | 194 |
| 6.3.4      | Distribution and efficiency of isolates capable of CTX $\Phi$ infection.....           | 195 |
| 6.3.5      | Assessment of pathogenic risk by cluster membership and other toxin-related genes..... | 199 |
| 6.4        | Discussion.....  | 206 |
| 6.4.1      | Occurrence of spontaneous kanamycin resistance in recipient cells.....                 | 206 |
| 6.4.2      | Mechanism of CTX-Km $\Phi$ infection in TCP-negative isolates.....                     | 207 |
| 6.4.3      | Other toxin-related genes related to pathogenic risk.....                              | 208 |
| 6.4.4      | Levels of pathogenic risk of non-toxigenic <i>V. cholerae</i> .....                    | 211 |
| 6.4.5      | Habitat and serotype of isolates susceptible to CTX $\Phi$ infection.....              | 213 |
| 6.4.6      | <i>V. mimicus</i> , a reservoir for toxin genes.....                                   | 213 |
| 6.5        | Conclusions.....   | 214 |
| Chapter 7  | Summary.....   | 216 |
| References | .....  | 222 |



## LIST OF TABLES

|  |     |
|--|-----|
| Table 2.1. Thirteen biochemical tests used to screen for <i>V. cholerae</i> .....                                  | 16  |
| Table 2.2. Results of the simple two-test procedure compared with the results of<br>molecular analysis.....        | 23  |
| Table 3.1. Sampling sites, location, and salinity range.....   | 36  |
| Table 3.2. Primer sequences employed for gene probe amplification.....   | 45  |
| Table 3.3. Number of positive samples by direct detection using PCR <sup>a</sup> .....                             | 49  |
| Table 3.4. Number of isolates and relevant information for each sample <sup>a</sup> .....                          | 50  |
| Table 3.5. Summary of phenotypic characteristics of the <i>V. cholerae</i> from Chesapeake<br>Bay.....             | 58  |
| Table 3.6. Genotypic characteristics of the <i>V. cholerae</i> isolated from Chesapeake Bay                        | 60  |
| Table 4.1. Bacterial strains tested by long-range ERIC-PCR fingerprinting.....                                     | 93  |
| Table 4.2. Estimates of variance components from ANOVA on transformed RBR and<br><i>r<sub>ERIC</sub></i> .....     | 125 |
| Table 4.3. Estimates of regression analysis of the three probes by ANCOVA.....                                     | 127 |
| Table 5.1. Percent positive reactions in the phenotypic tests of strains in each cluster.                          | 149 |
| Table 5.2. Genotypic traits of strains in each cluster.....  | 150 |
| Table 5.3. Frequency of isolates in each cluster by habitat and chronology (time) .....                            | 160 |
| Table 5.4. Percentage Relative Binding Ratios (RBR) from DNA-DNA Hybridizations<br>among the primary clusters..... | 178 |
| Table 6.1. Bacterial strains and efficiency of infection by CTX-Km $\Phi$ .....                                    | 197 |
| Table 6.2. Pathogenic risk of the isolates according to cluster membership.....                                    | 200 |
| Table 6.3. Estimation of etiological risk of non-toxigenic <i>V. cholerae</i> and <i>V. mimicus</i>                | 203 |
| Table 6.4. Environmental parameters and serogroups of strains susceptible to CTX $\Phi$ .                          | 205 |

## LIST OF FIGURES

|   |     |
|---|-----|
| Figure 2.1. Esculin hydrolysis reaction.....  | 18  |
| Figure 2.2. Results of esculin hydrolysis reaction after 10 days of incubation. ....  | 19  |
| Figure 2.3. Example of esculin hydrolysis reaction at different incubation times. ....  | 20  |
| Figure 3.1. Map of locations of the sampling sites.....   | 37  |
| Figure 3.2. Distribution of isolates among sampling sites.....  | 53  |
| Figure 3.3. Seasonal and annual variation in the occurrence of <i>V. cholerae</i> .....   | 54  |
| Figure 3.4. Number of isolates obtained from each sample by different isolation methods.....  | 56  |
| Figure 3.5. Bubble graph showing characteristics of <i>V. cholerae</i> from different samples.....  | 64  |
| Figure 3.6. Column graphs plotted between number of samples yielding luminous (lum+) and non-luminous (lum-) isolates and environmental parameters.....   | 69  |
| Figure 3.7. Column graphs plotted between number of samples yielding <i>stn</i> positive ( <i>stn</i> +) and <i>stn</i> negative ( <i>stn</i> -) isolates and environmental parameters.....   | 70  |
| Figure 3.8. Column graphs plotted between number of samples yielding low <i>toxR</i> isolates (signal level 0-2) and high <i>toxR</i> (level 3-4) isolates and environmental parameters.....  | 71  |
| Figure 4.1. Three dimensional scatter graph of principle component analysis (PCA) from ERIC PCR fingerprints of representative <i>V. cholerae</i> and <i>V. mimicus</i> strains.....  | 98  |
| Figure 4.2. Schematic flow diagram of statistical analyses for comparison of DNA-DNA hybridization and ERIC PCR fingerprinting in estimating genome relatedness between a <i>V. cholerae</i> strain and a probe strain (RC145)..... | 100 |
| Figure 4.3. Effects of comb thickness and running buffer on the banding patterns in ERIC PCR fingerprints.....  | 107 |
| Figure 4.4. Example of curve-based and band-based calculation for obtaining similarity value between pair of ERIC-PCR fingerprint.....  | 110 |
| Figure 4.5. Variability in fingerprint similarity caused by gel electrophoresis.....  | 112 |
| Figure 4.6. Variability in fingerprint similarity caused by different PCR reactions.....  | 115 |
| Figure 4.7. Cluster analysis of ERIC PCR fingerprints for identification of artificial clones.....  | 117 |
| Figure 4.8. Example of fingerprints and UPGMA dendograms obtained for different bacterial strains.....  | 120 |
| Figure 4.9. Genomic fingerprints of toxigenic <i>V. cholerae</i> .....  | 121 |

|  |     |
|--|-----|
| Figure 4.10. Scatter plot showing distribution of the transformed means of RBR and $r_{ERIC}$ .  | 124 |
| Figure 4.11. Regression analysis of transformed RBR and $r_{ERIC}$ .   | 128 |
| Figure 5.1. Dendrogram showing the overall population structure of the <i>V. cholerae</i> Chesapeake Bay isolates using the Neighbor-Joining (NJ) clustering method.                                 | 146 |
| Figure 5.2. Unrooted radial tree of <i>V. cholerae</i> Chesapeake Bay isolates obtained using the unweighted pair group method with arithmetic mean (UPGMA) and Fitch-Margoliash (Fitch) clustering. | 148 |
| Figure 5.3. NJ-Dendrogram of Cluster A1  | 153 |
| Figure 5.4. NJ-Dendrogram of Cluster B.  | 157 |
| Figure 5.5. Population structure of the extended set of <i>V. cholerae</i> strain collection by Neighbor-Joining clustering.   | 162 |
| Figure 5.6. Unrooted radial tree of the <i>V. cholerae</i> strain collection based on the unweighted pair group method with arithmetic mean (UPGMA) and Fitch-Margoliash (Fitch) clustering.         | 164 |
| Figure 5.7. Distribution of strains by DNA-DNA hybridization.  | 173 |
| Figure 5.8. Distribution of RBR values for strains in intermediate clusters against two Cluster A probes (RC466 and RC145).  | 174 |
| Figure 5.9. Distribution of RBR values for strains in subclusters against probes RC145 and RC466 (panel A) or RC145 and RC395 (panel B).   | 175 |
| Figure 5.10. NJ tree based on the number of differences in the 16S rDNA sequences for <i>V. cholerae</i> and <i>V. mimicus</i> .   | 180 |
| Figure 5.11. The region of 16S rDNA sequences in <i>V. cholerae</i> and <i>V. mimicus</i> where RC518 and RC586 showed significant deviation.  | 181 |

## **Chapter 1 Introduction**

### **1.1 Background for cholera and *Vibrio cholerae***

#### **1.1.1 Cholera disease**

Cholera is a long standing and devastating disease extensively recorded in human history since ancient times. The word “cholera” has been used for over 2,500 years to describe disease with diarrhea and vomiting. The most remarkable symptom of cholera is the large volume of watery diarrhea causing dehydration and hypotensive shock that can result in death if proper treatment is not administered in time. The disease affects large numbers of people each year, especially in areas where public health and sanitary measures are compromised. In year 1995, the estimated annual number of deaths during cholera outbreaks was 120,000 worldwide (194). Although with higher success rate in preventing and treating cholera, outbreaks continue to be a major public health concern. In the most recent report from the World Health Organization, for the year 2003, an emergency from cholera outbreaks was declared and it was designated as one of the health crises following the civil war in Liberia (196). Within a four month period, June-September 2003, more than 17,000 people were affected and 110 fatalities reported, with more than 2,000 new cases reported each week (195).

### **1.1.2 History of cholera**

Before the eighteenth century, cholera was mostly confined locally and was reported to occur in Asia, including China and Europe. However, after 1770, for the first time in its known history, the infection showed a marked tendency to spread further to neighboring countries. From its origin in India, the spread was recorded as far as Burma (Myanmar). Since the 1817 first pandemic (1817-1823), cholera showed extraordinary virulence, becoming increasingly fatal and running its course generally in a few hours and sometimes in a few minutes. It spread throughout Asia, but stopped before reaching Europe, perhaps blocked by the severe winter of 1823-1824, rather than from preventive measures since none was undertaken in many areas. The second pandemic (1829-1851) was believed to have originated from recrudescence of infection which had persisted from the time of the first pandemic. However, some believed it spread from the Indian subcontinent, going to Asia, Europe, the Americas, Arabia, and East and North Africa. It reached the United States in 1832, starting in New York, Philadelphia, Baltimore, Washington, D. C. and spreading throughout New Orleans inward to the west coast. The invasion of cholera continued until 1835, moving across the south and up to Charleston, South Carolina, mysteriously disappearing for nearly fifteen years, until another outbreak in 1849 which persisted for seven years until 1855. The following four pandemics, the third (1852-1859), fourth (1863-1879), fifth (1881-1896), and sixth (1899-1923), were believed to be the combined result of recrudescence and importation of the disease, starting from the Indian subcontinent and affecting mostly those continents located in the southern hemisphere, as well as North America and Europe. The seventh pandemic started in 1961 and is still going on. It originated in Indonesia, and then spread to the

Indian subcontinent and the Middle East, then to Africa in the 1970s, finally reaching South America in the early 1990s. Details of cholera pandemic history has been extensively covered in many reviews (10, 17, 97, 139, 156).

### **1.1.3 *Vibrio cholerae*, the causative agent**

During the fifth pandemic in 1883-1884, Robert Koch demonstrated that cholera was caused by a comma-shaped organism, the “comma bacillus” (105), and it was subsequently named *Vibrio comma* because of its shape. However, the cholera vibrio was first described in 1854 in Italy by Pacini, who named it *Vibrio cholera*. After the pioneering work of Pacini was recognized, the bacterial name was then changed to *Vibrio cholerae*.

*Vibrio cholerae* is a motile, gram negative, curved rod and is a facultative anaerobic bacterial species in the genus *Vibrio*, family *Vibrionaceae*. It is autochthonous to the aquatic environment, especially in brackish and estuarine systems (34, 35, 96). *V. cholerae* is a heterogeneous species and at least 206 serogroups have been identified (201), but only a few are known to cause disease in humans. There are only two serogroups, O1 and O139 that so far have been recognized as having the potential to cause epidemic cholera.

### **1.1.4 Dynamics of epidemic strains of *V. cholerae* serogroups O1 and O139**

Toxigenic *Vibrio cholerae* has the ability to infect and colonize the human intestinal tract and produce cholera toxin (CT), the potent enterotoxin responsible for onset of the watery diarrhea characteristic of cholera. Until 1992, all cholera epidemics were thought to be caused by *V. cholerae* serogroup O1. Surveillance of cholera in

Bangladesh has shown that *V. cholerae* biotypes and genotypes isolated from different epidemics change with time. Until 1970, more than 90% of cholera cases were caused by the *V. cholerae* O1 classical Inaba strains. By 1972, 85% of the cases were due to the classical Ogawa strain (58). Between 1969-1973, the El Tor biotype emerged and replaced the classical biotype. In 1982, the classical biotype re-emerged and coexisted with the El Tor until 1992. In 1992, a second etiologic agent of cholera first appeared in Madras, India then appeared in the coastal region of southern Bangladesh in the Bay of Bengal (26, 146). The new serogroup, *V. cholerae* O139 synonym Bengal caused explosive epidemics throughout Bangladesh, India, and neighboring countries during 1992-1993. However, during 1994-1995, in most northern and central areas of Bangladesh, the O139 serogroup was replaced by a new clone of *V. cholerae* O1 El Tor. However, *V. cholerae* O139 still exists in the southern region (56, 57). After a brief, transient displacement by a new clone of El Tor, the O139 reemerged in 1995 and co-exists with the El Tor biotype.

#### **1.1.5 Pathogenicity and epidemiology of *V. cholerae* non O1/ non O139**

The other serogroups of *Vibrio cholerae* (non O1/non O139) are readily isolated from natural waters and are mostly regarded as non-toxigenic because a majority of these strains do not produce cholera toxin. However, it has been reported to cause sporadic cases, usually associated with consumption of shellfish, and has been isolated from a variety of extraintestinal infections, including wounds, ear, sputum, urine, and cerebrospinal fluid (97). Inability of the non O1/ non O139 *V. cholerae* to cause epidemics was challenged when several epidemics caused by several serogroups were reported. An upsurge of *V. cholerae* non O1/ non O139 that cause cholera-like disease

was reported in India in 1996 (168). None of the 15 non O1/non O139 serogroups isolated from patients possessed major virulence factor *ctx* or *tcp* genes. Another epidemic in Peru in 1994 was shown to have been caused by serogroup O10 and O12 and all 58 isolates that were tested lacked the ability to produce CT and TCP (40). Therefore, these non O1/non O139 *V. cholerae* are concluded to be capable of causing epidemic diarrhea, but by an unknown mechanism quite different from that of toxigenic O1 and O139.

Strains of *V. cholerae* non O1/ non O139 possessing the major virulence genes, CTX and TCP, have been reported, such as serogroup O141 that cause sporadic cases (41), and O37 that caused a large outbreak of cholera in Sudan in 1968 (95).

## **1.2 Diversity of *V. cholerae***

As more advanced research using molecular tools became available, there is accumulating evidence showing that *V. cholerae*, as a whole, is highly diverse and active in genetic exchanges. Phenotypic markers used to identify toxigenic and epidemic capability by the serogroup O1 were no longer valid after serogroup O139 emerged. Other serogroups also showed the ability to cause epidemics and are equally devastating to public health.

*Antigenic shift of toxigenic strains.* The genes responsible for expression of O1 antigen are located in the *rfb* gene cluster, later named the *wbe* cluster, following a new scheme for bacterial polysaccharide gene nomenclature (BPGN) (147). Several non-O1 strains, including O139, O27, O37, O53, and O65, have been shown to possess the genetic backbone of O1 strains (determined by the genetic fingerprint, AFLP) and were proposed to have arisen from homologous recombination-mediated exchange of the genes



responsible for biosynthesis of O-antigen, *wb\** gene clusters (i.e., *wbf* for O139-, and *wbe* for O1-biosynthesis) (109). Extensive analysis of the nature of the new epidemic O139 serogroup suggests that it originated from a seventh-pandemic O1 El Tor strain that underwent antigenic shift by homologous recombination, replacing the *wbe* region of the O1 antigen with the O139 *wbf* region (126, 179, 180). The donor of the O139 *wbf* gene cluster has been proposed to be O22 serogroup, because of extensive homology with the O139 *wbf* region (203).

*Mobile nature of toxigenic determinants.* Besides changes in serogroups of the toxigenic clone, non-toxigenic clones may become toxigenic by acquiring toxigenic determinants by horizontal gene transfer. The major virulence factor, the cholera toxin gene, was found to be encoded by a single strand DNA filamentous phage, CTX $\Phi$  (188) and the colonization factor, toxin-coregulated pilus (TCP), is part of the vibrio pathogenicity island (VPI), which was proposed to be another filamentous phage (99). Although the status of the VPI as a self-transmissible phage has not been confirmed (64), another means of mobilization of the VPI genes mediated by a generalized transduction phage, CP-T1, among strains of *V. cholerae* O1 has been demonstrated (133).

In contrast to VPI, horizontal gene transfer of CTX $\Phi$  is very active. Capability of the CTX lysogen to produce infectious phage, either spontaneously or by induction by DNA damaging agent such as mitomycin C or sunlight has been demonstrated (59, 60, 188). After transduction, the CTX $\Phi$  can form a stable lysogen by integration into the recipient genome at the attRS site or by maintaining itself extrachromosomally as a replicative form (RF) (188). Although expression of toxin coregulated pili (TCP) has been shown to be necessary for the CTX $\Phi$  infection, since the phage use TCP as a

receptor (188), other mechanisms of transduction exist. Faruque *et al.* (1998) demonstrated transduction of the genetically marked derivative of the phage CTX-km $\Phi$  to a nontoxigenic environmental strain of *V. cholerae* lacking TCP genes, as well as a closely related species, the TCP negative *Vibrio mimicus* (61, 62).

With genetic exchange of large gene cassettes actively on going in the environment, yielding new and distinctly different clones, there is also microevolution within the same clone that does not cause detectable change in strain characteristics. Diversity within toxigenic strains of *V. cholerae* has been demonstrated by multilocus enzyme electrophoresis (MLEE) (14, 53), DNA sequence analysis of housekeeping genes (21, 178) and genomic fingerprinting by ERIC-PCR (209).

### **1.3 Significance of studying *V. cholerae* population structure and clonal determination**

As genetic exchanges give rise to a highly diverse population of *V. cholerae*, there is constant risk of emergence of new toxigenic clones. The above examples of biotype shift, antigenic shift, and mobility of the virulence determinant emphasize the need for monitoring *V. cholerae* population as a whole, not only known toxigenic clones.

In order to understand the bacterial population structure, identification of the clonality of the isolates is critical. Many molecular tools have been successfully used to determine bacterial clonality, with variable resolution power. Advantages and disadvantages of each method have been discussed and a rapid, simple, high resolution method employing long-range ERIC-PCR was developed in this study.

## **1.4 Census survey as a strategy to understand *V. cholerae* population structure**

Current information about the clonal nature of *V. cholerae* has mainly emphasized clinical isolates (32, 58, 63) or used an arbitrary selection of strains. Thus, the range of the species being covered was not known in the latter case. In this study, a systematic collection of *V. cholerae* strains was made and a detailed survey of the species in a single geographic area for a specific time frame (census survey) was accomplished. This provided a framework for understanding the distribution and the changes within the species at a given locale. Since *V. cholerae* is indigenous to the aquatic environment and has been isolated from many parts of the world, the information obtained from thorough analysis of the population at one locale can be useful for the species population as a whole, since different aquatic systems share common characteristics, in terms of microhabitat. A broader implication of its population structure that has been constructed from a census survey of *V. cholerae* is inferred by selection of isolates from other habitats, both clinical and environmental.

The Chesapeake Bay, Maryland, was chosen for sampling and for collecting isolates collection because it is one of the most productive and extensively studied estuaries in the east coast of the United States. It has been shown to harbor natural populations of *V. cholerae* in various publications since the 1970s (96). Because it is a geographic location free from cholera, it is an ideal site for studying natural populations of *V. cholerae* in the environment, with no interference from clinical cases of cholera.

## 1.5 Objectives of this study

The major objective of this study was to describe the population structure of *V. cholerae* by census collection of *V. cholerae* strains isolated from the Chesapeake Bay. In order to achieve this objective, several experimental approaches were developed. First, a simple procedure for rapid identification of *V. cholerae* strains isolated from large scale environmental sampling was developed. Second, high throughput genomic fingerprinting protocol was designed to define and resolve clonal lineages of *V. cholerae* isolates in the collection. The isolates were characterized phenotypically and genotypically, including toxin-related genes. The etiological potential, prevalence, and distribution of each lineage were analyzed. Isolates from other locales, as well as clinical isolates, were included to provide broader inference of *V. cholerae* population structure.

## **Chapter 2 A Simple Procedure for the Rapid Identification of *Vibrio cholerae* from the Aquatic Environment**

The majority of this chapter was published in Choopun *et al.*, 2002 (27). It is reprinted with the permission of American Society of Microbiology. Additional Figures are provided for clarity.

### **2.1 Introduction**

Since *Vibrio cholerae* is autochthonous to the aquatic environment (35, 192, 200), monitoring the bacterium in water sources is important for control of cholera outbreaks. A selective medium, such as thiosulfate-citrate-bile salts-sucrose (TCBS) agar, eliminates most nontarget bacteria in clinical samples but is not satisfactory for environmental samples because many bacteria present in natural water sources can produce colonies on TCBS agar whose appearance is similar to that of *V. cholerae* colonies. Furthermore, the series of biochemical tests commonly used to identify *V. cholerae* (13, 55, 100, 191) was originally designed for clinical samples, i.e., to detect pathogenic vibrios. Molecular methods, including PCR and DNA/DNA hybridization performed with probes specific for *V. cholerae*, provide more reliable identification (28, 129), but have limitations because of cost and the facilities required for analysis; these limitations are particularly significant for field studies involving large numbers of samples.

We report here the results of a study in which we used 13 generally accepted biochemical tests for identifying clinical strains of *Vibrio* spp. to determine the minimum number of such tests that is required to identify *V. cholerae* in environmental samples.

## **2.2 Materials and Methods**

### **2.2.1 Isolation of presumptive *Vibrio cholerae***

Water and plankton samples were collected from Chesapeake Bay in Maryland over a 7-month period (January through July 1998). Samples of water (250 ml) and plankton (25 ml) were collected and filtered by using 0.22- $\mu$ m-pore-size Supor-200 membrane filters (Gelman Laboratories, Ann Arbor, Mich.); this was followed by incubation of the membranes in an enrichment medium consisting of 100 ml of alkaline peptone water (APW: 1% (wt/vol) bacto-peptone (Difco Laboratories, Detroit, Mich.); 1% (wt/vol) NaCl; pH 8.6) overnight at 30°C. The incubation temperature was lower than the temperature normally used for clinical samples in order to accommodate and reduce the stress for environmental isolates that inhabit lower temperature environments. Two loopfuls of the culture broth, taken from the top layer of the APW, were streaked onto TCBS agar (Oxoid Ltd., Basingstoke, England) in triplicate and incubated overnight at 37°C. Six *V. cholerae*-like colonies, i.e., yellow, flat, 1-3 mm diameter colonies were selected from each sample and subcultured on Luria-Bertani (LB) agar (Difco Laboratories) supplemented with 2% NaCl. A total of 844 colonies were purified and stored for further testing.

**Biochemical tests:** Thirteen biochemical tests were performed for each isolate. All bacteriological media and chemicals are from Difco Laboratories, Detroit, MI, unless specified.

Kligler iron agar (KIA) slants were prepared as recommended by the manufacturer, and 5-ml aliquots were dispensed into 13 x 100 mm tubes. The tubes were inoculated by stabbing. Results were read after incubation at 37°C for 24 h.

For the oxidase test, bacterial colonies were picked with a platinum wire and streaked on filter paper saturated with 0.5% tetramethyl-*p*-phenylenediamine hydrochloride (BBL Co.). Rapid appearance of a dark purple color was considered a positive reaction.

The method used for the arginine dihydrolase test was based on Thornley's method (183). The medium used was LB broth containing 1% (wt/vol) L-arginine, pH 6.8. Phenol red powder was added as an indicator. After inoculation, the medium was covered with sterile mineral oil and incubated at 37°C for 24 h. Appearance of a red color was considered a positive reaction.

Lysine and ornithine decarboxylase assays were performed by using Moeller decarboxylase base medium amended with 1% (wt/vol) of the indicated L-amino acids and adjusted to pH 6.8. The base medium, without an amino acid added, was inoculated as a control. After inoculation, the medium was covered with mineral oil and incubated at 37°C for 24 h. Positive reactions were indicated by a dark purple color compared to the color obtained with the base medium without an amino acid.

Cells grown in 0, 6, and 8% (wt/vol) NaCl in nutrient broth were used to determine the requirement for Na<sup>+</sup>. The medium was lightly inoculated and incubated at

30 or 37°C for up to 7 days, and positive results were determined by examining the turbidity.

Acid production from mannitol or arabinose was determined by using purple broth base amended with 1% (wt/vol) of the tested sugar, and adjusted to pH 6.8. The results were read after no more than 24 h of incubation at 37°C. A change in the color of the medium from purple to yellow indicated a positive reaction.

The methyl red reaction was tested by using MR-VP medium incubated at 37°C for 48 h after inoculation. A few drops of the methyl red reagent (0.1 g Methyl red, 300 ml 95% ethanol, adjusted to 500 ml with distilled water) were added to the culture broth. A bright red color appeared immediately after addition of the reagent were recorded as positive result (108).

The Voges Proskauer test assay was performed by using a culture grown in MR-VP medium at 37°C for 48 h. The Barritt method was used to detect a positive reaction. Two reagents, solution A (5 g  $\alpha$ -naphthol in 100 ml absolute ethanol), 0.6 ml, and solution B (40% wt/vol potassium hydroxide in water), 0.2 ml, were mixed with 1 ml of the culture broth, shake well after each addition, and the positive results were recorded when red coloration developed within 5 min after the reagents addition(108).

Esculin hydrolysis was tested by using heart infusion agar containing 0.1% esculin and 0.05% ferric chloride (108). Blackening of the medium after incubation at 30 or 37°C for up to 3 days indicated that there was a positive reaction.

### **2.2.2 Confirmation of the identities of presumptive *V. cholerae* isolates**

For development of the method, the identities of presumptive isolates of *V. cholerae* were confirmed by PCR. The PCR reaction was performed as described by



Chun *et al.*(28). The PCR primer set used, prVC-F (5' TTAAGCSTTTTCRCTGAGA ATG-3') (renamed as pVC-F2) and prVCM-R (5'-AGTCACTTAACCATAACAACCCG-3') (renamed as pVCM-R1), amplifies a portion of the intergenic spacer region between the 16S and 23S rRNA genes and is specific for *V. cholerae* (>100 strains tested)(28). In our evaluation of the simple biochemical test procedure, due to the vast number of isolates, a colony blot hybridization technique was used instead of PCR. The probe was based on a modification of PCR primer pVC-F2. The oligonucleotide probe, pVC-ITS-1 (5'-GCSTTTTCRCTGAGAATG-3'), was end labeled with <sup>32</sup>P radioisotope. This probe hybridized with all 201 *V. cholerae* strains tested and did not hybridize with 30 other *Vibrio* species and 10 non-*Vibrio* species (C. J. Grim, I. N. G. Rivera, N. Choopun, J. Chun, A. Huq, and R. R. Colwell, Abstr. 100<sup>th</sup> Gen. Meet., Amer. Soc. Microbiol., abstr. Q33, 2000). To evaluate the simple test, bacterial colonies were transferred to no. 541 filter paper (Whatman, Inc., Clifton, N.J.) and processed as previously described (199). The hybridization reaction was performed using sodium chloride/sodium citrate (SSC) method as described in a standard protocol (51).

### **2.2.3 Selection of the biochemical test**

The effectiveness of each biochemical test was evaluated based on sensitivity and specificity, and 100% sensitivity was sought in order to eliminate false negatives.

Sensitivity and specificity were calculated as follows:

$$\text{Sensitivity} = [(\text{number of isolates positive as determined by biochemical tests and PCR}) / (\text{total number of isolates positive as determined by PCR})] \times 100$$
$$\text{Specificity} = [(\text{number of isolates negative as determined by biochemical tests and PCR}) / (\text{total number of isolates negative as determined by PCR})] \times 100$$

## **2.3 Results and Discussion**

### **2.3.1 Selection of the effective tests**

The tests that provided 100% sensitivity included the arginine dihydrolase and esculin hydrolysis tests, which had specificities of 72% and 26%, respectively (Table 2.1), and these tests were shown to be very effective in screening for *V. cholerae*. Of 844 bacterial colonies (yellow colonies on TCBS agar) tested, 241 were identified as presumptive *V. cholerae* isolates by using only these two tests, and all were confirmed to be *V. cholerae* isolates by PCR. None of 13 isolates randomly selected from the isolates not identified as *V. cholerae* by this pair of tests were found to be a *V. cholerae* when PCR was used.

Table 2.1. Thirteen biochemical tests used to screen for *V. cholerae*<sup>a</sup>

| Biochemical test               | <i>V. cholerae</i> reaction <sup>b</sup> | % of isolates with same reaction as <i>V. cholerae</i> | Sensitivity (%) | Specificity (%) |
|--------------------------------|--|--|-----------------|-----------------|
| Kligler iron agar <sup>c</sup> | K/A                                      | 98   | 100             | 2               |
| Oxidase <sup>c</sup>           | +  | 99   | 100             | 1               |
| Arginine                       | -  | 49   | 100             | 72              |
| Lysine                         | +  | 75   | 98              | 30              |
| Ornithine                      | +  | 61   | 97              | 53              |
| NaCl (0 %)                     | +  | 57   | 90              | 56              |
| NaCl (6 %)                     | V  | NA <sup>d</sup>  | NA              | NA              |
| NaCl (8 %)                     | -  | 97   | 96              | 3               |
| Mannitol                       | +  | 68   | 71              | 30              |
| Arabinose                      | -  | 85   | 100             | 20              |
| Methyl red                     | V  | NA   | NA              | NA              |
| Voges-Proskauer                | +  | 56   | 80              | 53              |
| Esculin                        | -  | 81   | 100             | 26              |

<sup>a</sup> Total of 844 bacterial isolates from Chesapeake Bay that showed *V. cholerae*-like characteristics, i.e., produced yellow colonies on TCBS agar, were included in the study. See text for the methods used to calculate sensitivity and specificity.

<sup>b</sup> K/A, alkaline at the top and acid at the bottom; +, 90 to 100% of the isolates were positive; -, 0 to 10% of the isolates were positive; V, variable reaction.

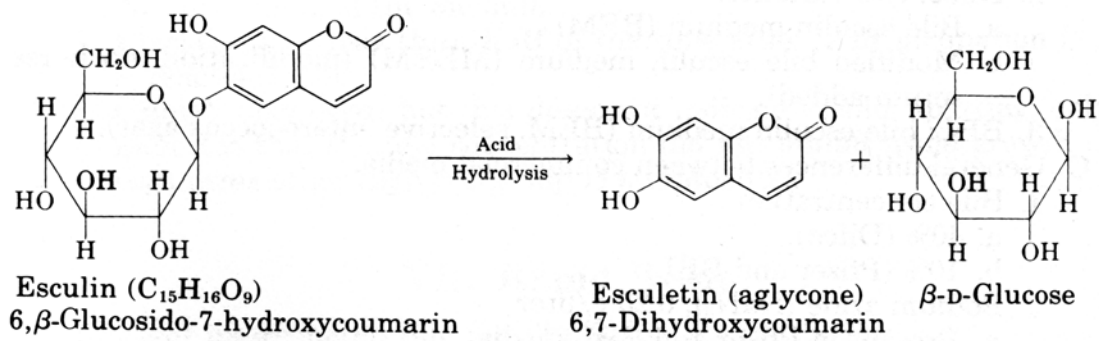
<sup>c</sup> The Kligler iron agar and oxidase test results are the results obtained with 336 and 330 isolates, respectively.

<sup>d</sup> NA, not applicable.

### 2.3.2 Evaluation of the simple biochemical test procedure

After the test procedure was developed, additional water samples were collected from the Chesapeake Bay from June through August 2000. A total of 481 bacterial colonies were picked and subjected to both the simple test procedure and colony blot hybridization. Sixteen isolates were identified as *V. cholerae* by the simple test. The results were in complete agreement with the hybridization results.

In addition to the Chesapeake Bay samples, 67 presumptive *V. cholerae* isolates from Peru, isolated from environmental samples collected by using the procedures described above, were subjected to the arginine dihydrolase and esculin hydrolysis tests. Of these 67 isolates from Peru, 25 turned the esculin medium dark but were identified as *V. cholerae* by PCR. Normally, blackening of the medium results from the reaction of esculin, a product of esculin hydrolysis, with ferric ion in the medium, which results in a black phenolic compound-iron complex (Fig. 2.1). The hydrolysis also results in a loss of the characteristic fluorescence of esculin that can be measured at 354 nm (long wavelength-UV light) (113). These 25 isolates from Peru darkened the medium, but there was no loss of fluorescence, even up to 10 days after inoculation (Fig. 2.2), demonstrating that these strains did not hydrolyze esculin. After stab inoculation, all the isolates that hydrolyzed esculin completely blackened the medium by day 2 and there was a total loss of fluorescence by day 3 (Fig. 2.3).



Fluorescence

Black complex

Figure 2.1. Esculin hydrolysis reaction.

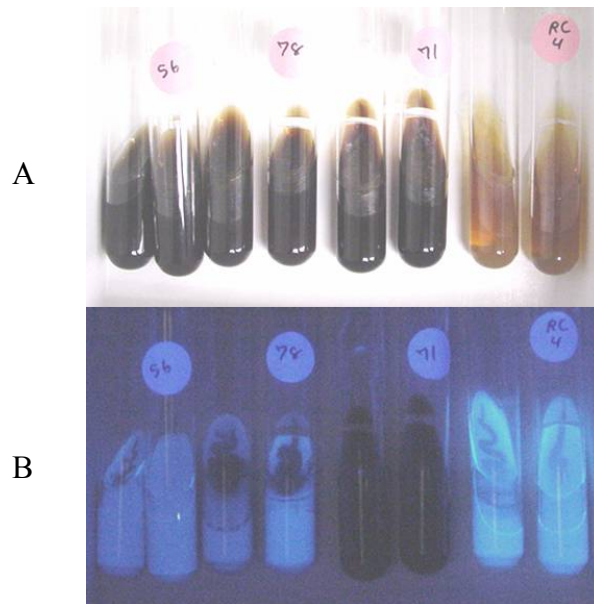


Figure 2.2. Results of esculin hydrolysis reaction after 10 days of incubation. Panel A, the picture was taken in visible light; panel B, the picture was taken when the tubes were illuminated by UV light (354 nm). Four strains were inoculated; each strain was inoculated into two tubes, using different inoculating methods. The first tube was inoculated by stabbing the medium and the second tube by streaking on the slant surface. The first and second pair of tubes represent *V. cholerae* isolates from Peru (56 and 78), the third an esculin hydrolysis positive strain (non-*V. cholerae*) (71), and the fourth a control, *V. cholerae* (RC4).

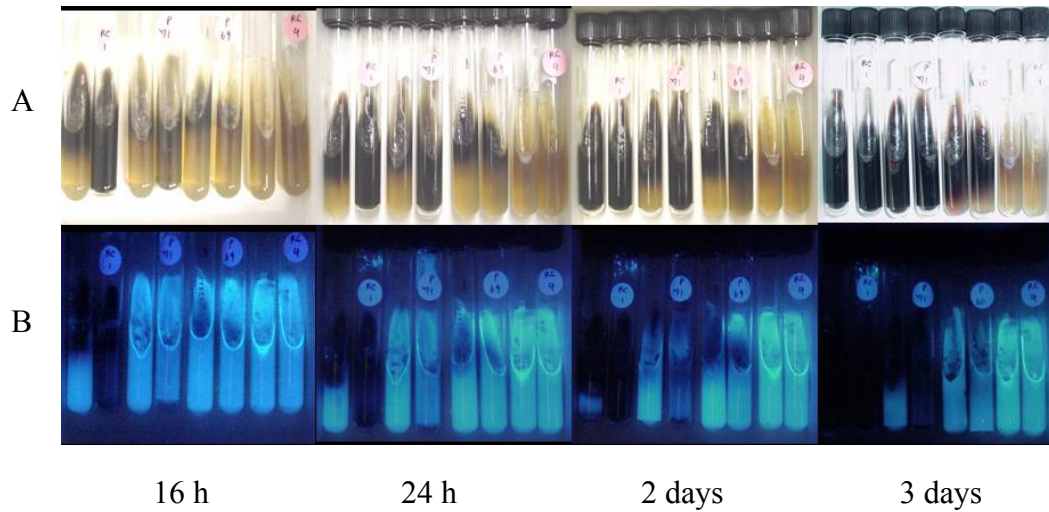


Figure 2.3. Example of esculin hydrolysis reaction at different incubation times. Panel A, the picture was taken in visible light; panel B, the picture was taken when the tubes were illuminated by UV light (354 nm). Each strain was inoculated into two tubes by different inoculating methods, the first tube was inoculated by streaking on the slant surface and the second by stabbing. The first pair represents the control for esculin hydrolysis positive reaction and is *Klebsiella pneumoniae* (RC1). The second pair is a non-*V. cholerae* esculin positive isolate from Peru (P71). The third is *V. cholerae* isolated from Peru (P69), and the fourth is a control, *V. cholerae* (RC4).

The difficulty in interpreting the results of the esculin hydrolysis test was overcome by testing for fluorescence of the esculin instead of relying only on darkening of the medium. Blackening of the test medium by some *V. cholerae* strains, without esculin hydrolysis, can result from production of other compounds (e.g., melanin) (36). However, to produce melanin, the bacterium must be grown under highly aerobic conditions (36, 90, 106). Addition of 20 mM L-proline, which has been shown to inhibit melanin production in *V. cholerae* (36), did not prevent blackening of the test medium by the isolates from Peru. When stab inoculation is used, esculin hydrolysis is indicated by total blackening of the medium within 48 h and/or by a total loss of fluorescence by 72 h. When the latter parameter was used, 60 of the 67 *V. cholerae* isolates from Peru were identified as *V. cholerae* by using only the two biochemical tests, and 57 of the 60 strains were subsequently confirmed to be *V. cholerae* strains by PCR. The three false positives were then identified by 16S ribosomal DNA sequencing and by comparing the sequences obtained with sequences in the GenBank database. Two of these three strains were identified as *Vibrio metschnikovii* and the third was identified as *Vibrio natriegens*. Although some strains of other *Vibrio* species might give false positive results with this simple test, it is relatively easy to distinguish these strains from *V. cholerae* by a simple Na<sup>+</sup> requirement test. None of the other sucrose-fermenting *Vibrio* species can grow in nutrient broth without added Na<sup>+</sup>. *V. metschnikovii*, which has low Na<sup>+</sup> requirement, is oxidase negative (13), hence it can be easily detected by using the oxidase test.

Table 2.2 shows a comparison of the results of the simple biochemical tests with the results obtained by using molecular methods. The simple pair of tests provided 100% sensitivity for both the Chesapeake Bay and Peru isolates; and 100 and 70% specificities



for the Chesapeake Bay isolates and Peru isolates respectively. The 100% sensitivity of this simple biochemical test procedure was meaningful because no false-negative results were obtained; therefore, all *V. cholerae* isolates were detected. Further field experiments, especially in endemic areas, will be useful to assess the effectiveness of this simple procedure in specific areas. Two more tests, the Na<sup>+</sup> requirement and oxidase tests, may be included in the screening process in areas where there are high number of false-positive results.

Other selective media, such as cellobiose-polymyxin B-colistin (CPC) agar and its modified formulas, mCPC and CC (cellobiose-colistin) agar, have been reported to be superior to TCBS agar for isolation of *Vibrio vulnificus* and *V. cholerae* (108, 119). These media may be useful for increasing the probability of isolation of *V. cholerae* from aquatic environments. However, whether these media can be used to isolate strains that are more susceptible to polymyxin B, such as the *V. cholerae* O1 classical biovar (13) has not been determined. In contrast to TCBS agar, which is routinely used in many laboratories, is commercially available, and requires only a boiling step, these media require more preparation steps, including pH adjustment, autoclaving, and filtration sterilization of the carbohydrates and antibiotics, and they must be freshly prepared before each use because they become more inhibitory as they age (119). These steps limit their use to larger laboratories with more resources.

Table 2.2. Results of the simple two-test procedure compared with the results of molecular analysis.

| Simple test <sup>a</sup> | Molecular Analysis <sup>b</sup> |    |       |              |     |       |              |    |       |             |     |       |
|--------------------------|---------------------------------|----|-------|--------------|-----|-------|--------------|----|-------|-------------|-----|-------|
|                          | 1998 Samples                    |    |       | 2000 Samples |     |       | Peru Samples |    |       | All Samples |     |       |
|                          | +                               | -  | Total | +            | -   | Total | +            | -  | Total | +           | -   | Total |
| +                        | 241                             | 0  | 241   | 16           | 0   | 16    | 57           | 3  | 60    | 314         | 3   | 317   |
| -                        | 0                               | 13 | 13    | 0            | 465 | 465   | 0            | 7  | 7     | 0           | 485 | 485   |
| Total                    | 241                             | 13 | 254   | 16           | 465 | 481   | 57           | 10 | 67    | 314         | 488 | 802   |
| Sensitivity              | 100 %                           |    |       | 100 %        |     |       | 100%         |    |       | 100 %       |     |       |
| Specificity              | 100 %                           |    |       | 100 %        |     |       | 70 %         |    |       | 99.3 %      |     |       |

<sup>a</sup> +, positive for identification as *V. cholerae* as determined by the two-test method (i.e., both arginine and esculin tests negative); -, negative for identification as *V. cholerae*.

<sup>b</sup> Molecular analysis was conducted by PCR for 1998 and Peru samples; by hybridization for 2000 samples.

## 2.4 Conclusion

In summary, aquatic samples can be screened rapidly for the presence of *V. cholerae* by enriching samples in APW, selecting yellow colonies on TCBS agar, and performing the arginine dihydrolase and esculin hydrolysis tests with the purified isolates. Isolates giving negative reactions in both of these tests (i.e., no color change in arginine medium and showing fluorescence in esculin medium) can presumptively be considered *V. cholerae* isolates.

The simple two-test procedure described here has additional useful features in that the results are easy to read and do not change with time, allowing more flexible operation. Furthermore, no additional steps are required to obtain the results. The simple two-test procedure is useful for field work, especially in developing countries where *V. cholerae* is endemic and facilities are limited.

## **Chapter 3 Isolation and Characterization of *V. cholerae* from the Chesapeake Bay**

### **3.1 Introduction**

To determine the population structure of a species, it is necessary to establish a collection of representative strains that is evenly and randomly distributed geographically. However, when a species is ubiquitous in the aquatic environment, this approach is difficult, if not impossible. Alternatively, one can select an ecosystem and undertake a census, i.e., isolate and characterize all isolates obtained from a single ecosystem to determine population structure. By establishing spatial and temporal scales for a survey, a global perspective then becomes possible. When sufficient data are obtained in a census for a specific ecosystem, the population structure of the microbial community can be determined.

In this chapter, the Chesapeake Bay, Maryland, was selected to carry out for the complete microbial survey of a defined geographical area, and a culture collection was established by isolating *V. cholerae* from selected stations in the Bay over a two year period. Described below is the rationale for selecting each of the sampling stations, a description of the detection and isolation methods employed, and characteristics used to describe strains included in the collection.

Chesapeake Bay, Maryland, one of the most important estuaries on the East Coast of the United States produces a majority of the seafood supply in the region. It was

selected as the location for a microbial survey for several reasons, the most obvious being that the United States is free of cholera, hence reducing complications arising from potential contamination by clinical isolates entering the environment. Based on previous studies carried out in the Chesapeake Bay, the number of *V. cholerae* present in water and sediment showed strong correlation with temperature and salinity, with highest numbers of *V. cholerae* being found where the salinity is 2-17 ppt. and the temperature 19-30°C (30, 96, 111). Therefore, the sampling sites employed were concentrated in the upper part of the Chesapeake Bay, where the salinity ranges from 0-15 ppt. The presence of *V. cholerae* in water samples and its abundance and characteristic traits provides insight into its environmental role, as well as any risk associated with this bacterium in the natural environment.

### **3.1.1 Direct detection of *V. cholerae* by PCR (Non-culture based method)**

Direct detection of the target bacterium by molecular techniques provides a very useful and practical alternative to culture or microscope methods, especially for environmental samples. PCR is a powerful and widely used technique, providing high sensitivity and speed in the detection of microorganisms. It has been used for direct detection of specific target species in environmental samples in many studies, including analysis of *V. cholerae* and its toxigenic traits, i.e. O1/O139 serotype and CTX genes (110). PCR primers specific for *V. cholerae* (28) allow direct detection of the entire population, including viable but nonculturable (VBNC) cells which cannot be detected by culture-based methods. Although the technique does not yield isolates, it provides useful information concerning occurrence and distribution of *V. cholerae* in the environment.

### **3.1.2 Isolation of *V. cholerae* from the aquatic environment**

In order to obtain *V. cholerae* isolates, an enrichment method was employed. However, in the second year of the study, a new method employing direct plating was performed in parallel because it promised rapid and quantitative enumeration of the bacterium. A brief description of the methodology and background of each method is provided as follows.

#### **A Enrichment method**

Standard practice for isolation of *V. cholerae* usually involves an enrichment step, followed by selection using a selective medium (11, 97). In this study, the media used to isolate *Vibrios*, alkaline peptone water (APW) and thiosulfate citrate bile salts sucrose (TCBS) agar, were used as enrichment and selective medium, respectively. *Vibrios*, including *V. cholerae* grow well in the peptone and alkaline pH of APW, while other bacteria cannot grow or grow at a slower rate. Therefore, this enrichment step increases the chance of detecting *V. cholerae* in samples, especially when the number of *V. cholerae* is low. After enrichment, characteristic colonies of the bacterium were picked from the selective medium, TCBS. Growth of other non-vibrio bacteria is inhibited by the bile salts in the medium. The medium also discriminates between sucrose fermenting and non-sucrose fermenting bacteria. *V. cholerae* is sucrose fermenting and grows as yellow colonies on TCBS.

## **B Direct plating method**

Direct detection and isolation by direct plating, combined with DNA hybridization, was explored as an alternative to enrichment. Advantages of this approach are rapid and quantitative measurement of the target bacteria.

Several variations in protocols and media have been used by many investigators. Jieng and Fu (2001) (92) enumerated *V. cholerae* in water samples by filtration, concentrating the bacteria on nylon membranes and incubating the membranes on TCBS, followed by DNA hybridization. This method has some drawbacks, in that estimation of cell concentration must be pre-determined in order to select the appropriate volume of the sample to be filtered. Furthermore, the inhibitory action of TCBS against some strains of *V. cholerae* and/or the physiological health of the target vibrios has been recognized by several investigators (72, 193), and in our own research. Gooch *et al.* (2001) used DNA hybridization to detect *V. parahaemolyticus* in oysters, after enrichment, as a mechanism to improve enumeration by MPN, and compared results with those obtained by direct enumeration using both TCBS and non-selective media. The results showed high correlation among all the methods, with nonselective media yielding higher counts than TCBS. Direct detection methods were, therefore, applied successfully in subsequent applications (49, 52, 71). Wright *et al.* (198, 199) were successful using Whatman no. 541 filter paper as an alternative to the nylon membrane for colony hybridization. This cheaper alternative provided significant advantage for environmental studies where large numbers of samples must be tested and would be cost-prohibitive otherwise.

Although direct detection has a lower sensitivity than the enrichment method and determining proper dilution is problematic, especially for environmental samples where

the number of target cells is not known, it has advantages in yielding isolates that verify presence in the samples. This is not true when enrichment is used, because the isolates may be an artifact arising from the enrichment step (termed “artificial clone” in this study) and may give rise to an over-estimation of selected clones.

In this study, we explored the use of direct plating, coupled with colony hybridization to obtain non-artificial clone isolates. Using the rationale that alkaline peptone has been used with great success as an enrichment medium for *Vibrio* spp., the medium in an agar form (alkaline peptone agar) was selected as a direct plating medium to provide a favorable (alkaline) environment for *V. cholerae*, without the presence of toxic selective agents. A similar medium, modified alkaline nutrient agar (mANA), was used and found to favor growth of *V. cholerae* over TCBS and also to favor direct detection by colony hybridization in the subsequent step (153). A lower incubation temperature of 15°C was used to provide less stress for cells coming from a lower temperature aquatic environment. Filter paper no. 541 was used for colony hybridization with *V. cholerae*-specific oligonucleotide probe.

### **3.1.3 Phenotypic characterization of environmental isolates of *V. cholerae***

Phenotypic characteristics of *V. cholerae* are well documented for clinical isolates and were used as criteria for screening and identifying *V. cholerae* in general (13, 55, 191). However, when environmental isolates, which may have a higher diversity, are of interest, these criteria may need to be re-assessed. With the help of molecular tools (specific DNA probe/primers) to confirm identity of the isolates, atypical strains (phenotypic) can be identified without having to employ exhaustive sets of biochemical tests to confirm identity of isolates.



The tests selected for this study included those tests widely employed to identify *V. cholerae*, such as requirement for NaCl, fermentation of sugars, degradation of amino acids, and growth at specific temperatures. The results provided an estimate of variability among environmental isolates from the Chesapeake Bay. Tests for ability to degrade macromolecules, such as chitin, lipid, and starch, and bioluminescence property provides insight, as well, into the role of vibrios in the environment. Tests for the serotypes O1 and O139, the characteristics associated with epidemic *V. cholerae*, were included to assess distribution of these important phenotypes in Chesapeake Bay.

#### **3.1.4 Genotypic characterization using gene probes**

Many genes contributing to the pathogenicity of *V. cholerae* have been identified. Possession of toxin-related genes provides useful information for health risk assessment. Many of the toxin genes have been shown to be present in environmental isolates, that is, in the genome of natural inhabitants of the aquatic environment (24, 151). In this study, nine gene probes were used, most of which are toxin-related, to provide information about the distribution of these traits among *V. cholerae* in Chesapeake Bay. These genes included *ctxA*, *zot*, *tcpA*, *toxR*, *hlyA*, *stn*, *ompU*, *ompW*, and *luxA*.

The *ctxA* is a gene encoding for the A subunit of cholera toxin, the major toxin responsible for severe watery diarrhea that is characteristic of cholera. The *zot* gene was first described as an additional toxin produced by *V. cholerae*. The encoding protein, zonular occludens toxin, increases permeability of the small intestinal mucosa by affecting the structure of the intercellular tight junction, or zonula occludens (12, 65). This gene was later found to be part of the CTX element, which has since been shown to be the genome of the filamentous bacteriophage, CTXΦ (188). The *zot* gene function

was determined to be essential for assembly of the phage particle. Together with *ctxA*, it gives evidence of the presence of CTX $\Phi$  in the isolates.

The *tcpA* is the gene coding for the major protein subunit of toxin-coregulated pili (TCP). TCP contributes to pathogenicity of *V. cholerae* in two ways; first, it is an essential colonization factor inside the host intestine (7, 82), the first step in pathogenicity. Second, it has been shown to be the receptor site for CTX $\Phi$  (188), hence it increases the chance of the bacterium to acquire the CTX gene.

The transmembrane protein, ToxR, encoded by the *toxR* gene, is a central regulator, coordinating with environmental signals and regulating expression of several genes (137), including the *ctx*, *tcp*, and *ompU* genes. ToxR has been shown to be essential for the virulence of *V. cholerae* (50, 82).

Hemolysin encoded by *hlyA* was initially purified and shown to be cytolytic for a variety of erythrocytes and mammalian cells (84). The hemolysin, sometimes called El Tor hemolysin because it was described and used in the hemolysis of sheep erythrocytes test to distinguish between the El Tor (positive) and classical (negative) biotypes. However, more recent El Tor isolates have been found to be poorly hemolytic (9). The identical hemolysin is also produced by non-O1 *V. cholerae*, regardless of hemolytic ability (202). The lack of hemolysin in the classical biotype was shown to result from an eleven bp deletion in the gene which then produces a truncated protein (2).

A heat-stable enterotoxin similar to ST enterotoxin in *Escherichia coli*, produced by *V. cholerae* non-O1 (NAG-ST), was first purified and characterized by Arita *et al.* (1986) (6). The NAG-ST 17 amino acid toxin is encoded by the *stn* gene in *V. cholerae* non-O1 (sometimes referred to as the NAG-ST gene) (130). This toxin has been

proposed as one of the toxins responsible for clinical manifestations of non-O1 *V. cholerae* gastroenteritis and has been demonstrated to be the virulence factor in a human volunteer study (127). Although NAG-ST producing strains were reported to be primarily non-O1 and CTX negative, a similar toxin, with an almost identical gene sequence (*sto* gene), was reported to be present in one CTX positive, O1 serotype isolate (182).

Two of the outer membrane protein, OmpU and OmpW, genes were included in this study. The 38 kDa porin OmpU is a major outer membrane protein and comprises 30-60% of the total outer membrane protein of *V. cholerae*. It is regulated directly by ToxR. OmpU has been shown to be critical for *V. cholerae* bile resistance (141, 142), intestinal colonization (173, 174), and also plays a role in the organic acid tolerance response (ATR) (123). These properties implicate the importance of the *ompU* gene in providing a degree of fitness for *V. cholerae* in surviving harsh environments, including the environment inside the host.

Another membrane protein, OmpW, is a 22 kDa, highly immunogenic protein, expressed in minor amounts under normal laboratory conditions (116). The function of this protein is currently unknown but it has been suggested from its properties to be a porin-like protein (116) and has a high sequence similarity to a colicin S4 receptor, OmpW in *E. coli* (138). Parts of the OmpW amino acid sequence were shown to be identical to a partial sequence of a 20-kDa pilus protein produced by a diarrheogenic strain of non-O1/non-O139 *Vibrio cholerae* (165). This pilus protein was shown to have a function in intestinal colonization by the strain. Besides its suggested role in

toxigenicity, the *ompW* gene has been shown to be a promising gene target for specific detection of both clinical and environmental *V. cholerae* (129).

Bioluminescence is one of the interesting properties possessed by many marine bacteria, especially those in the family *Vibrionaceae*, including subpopulations of environmental isolates of *V. cholerae* (135, 190). Luminescence is controlled by the lux operon and is regulated in a cell-density dependent manner called quorum sensing, a response in target gene expression when extra cellular signal molecules, called autoinducers, reach a critical concentration. Similar quorum-sensing systems and many genes homologous to the lux operon have been described and shown to have a function in controlling a variety of cell-dependent responses, including virulence and biofilm formation in *V. cholerae* (76, 206, 207). The *luxA* is a gene encoding for a subunit of luciferase, an enzyme responsible for the luminescent reaction. The *luxA* sequence is highly conserved among *V. cholerae* (135, 145). Some non-luminous strains also were found to possess the *luxA* gene, with an identical sequence to luminous strains. Prevalence of this characteristic and its association with specific environmental factors can provide additional information and insight into the role of these isolates in the natural habitat.

### **3.1.5 Objectives of this study**

The primary objective of this study was to undertake a census of *V. cholerae* and obtain environmental isolates of this species from the Chesapeake Bay. While doing that, two isolation methods, enrichment and direct plating, were compared with respect to comparative efficiency of the methods and diversity of the isolates obtained. The

dynamics of the isolated *V. cholerae*, in terms of abundance, prevalence of specific traits, and preferred habitats were analyzed.

## **3.2 Materials and Methods**

The sampling sites and the sample collection and enrichment protocols have been described elsewhere (27, 111). The methods are described briefly in this chapter to aid in understanding the additional protocols employed, i.e., direct detection by PCR and isolation by direct plating.

### **3.2.1 Sampling sites**

Isolates from two sets of sampling sites were studied: the shore sampling sites were sites where sampling is done regularly by our laboratory. Samples were collected from the dock at the site. Cruise sampling sites provided additional sites and sampling was conducted aboard a research vessel in the summer, covering the entire salinity range of the Chesapeake Bay. The additional sampling provided by the cruise yielded additional *V. cholerae* isolates from more diverse habitats in the Bay.

**Shore sampling:** Samples were collected at four sampling sites (F, H, K, S) located on the shore along the northern part of the Chesapeake Bay, Maryland, beginning in January, 1998 and running until February, 2000. The fifth site B, located in the inner harbor, Baltimore, was added to the study in January, 1999. Sampling was done once every month and twice during the summer (June, July, and August), comprising a total of 15 sampling events each year.

**Cruise sampling:** An additional nine sampling locations were included in the study and covered the entire length of the Chesapeake Bay, from the upper part to the

mouth of the bay. Samples were processed aboard ship on research cruises of the RV/ *Cape Henlopen*. The sampling sites are listed by abbreviation of the site latitude, i.e., 707 to 908, corresponding to sites at latitudes of 37°07'N to 39°08'N, respectively. In each of the two years of the study, two rounds of sampling cruises were completed in the summer months when the number of *V. cholerae* was expected to be at a peak (August and September, 1999, and June and August, 2000).

Name, location, and salinity of each sampling site are listed in Table 3.1. The corresponding map is shown in Fig. 3.1.

Table 3.1. Sampling sites, location, and salinity range

| Site | Name   | latitude   | longitude  | Salinity |         |
|------|--|------------|------------|----------|---------|
|      |  |            |            | Range    | Average |
| F    | Susquehanna River Flat                           | 39°33.13'N | 76°02.20'W | 0        | 0.00    |
| B    | Inner Harbor, Baltimore                          | 39°17.00'N | 76°36.32'W | 3-12     | 7.88    |
| K    | Kent Island                                      | 38°58.84'N | 76°20.13'W | 2-13     | 8.98    |
| S    | Smithsonian Environmental Research Center (SERC) | 38°53.20'N | 76°32.51'W | 2-15     | 9.05    |
| H    | Horn Point Laboratory                            | 38°35.59'N | 76°07.80'W | 5-15     | 9.25    |
| 908  | Cape Henlopen research cruise                    | 39°08'N    | 76°20'W    | 5-14     | 9.27    |
| 858  | Cape Henlopen research cruise                    | 38°58'N    | 76°23'W    | 6-16     | 11.49   |
| 845  | Cape Henlopen research cruise                    | 38°45'N    | 76°26'W    | 9-17     | 13.66   |
| 834  | Cape Henlopen research cruise                    | 38°34'N    | 76°26'W    | 10-17    | 14.15   |
| 818  | Cape Henlopen research cruise                    | 38°18'N    | 76°17'W    | 12-18    | 15.21   |
| 804  | Cape Henlopen research cruise                    | 38°04'N    | 76°13'W    | 13-19    | 16.02   |
| 744  | Cape Henlopen research cruise                    | 37°44'N    | 76°11'W    | 14-21    | 17.15   |
| 724  | Cape Henlopen research cruise                    | 37°24'N    | 76°05'W    | 18-24    | 20.62   |
| 707  | Cape Henlopen research cruise                    | 37°07'N    | 76°07'W    | 21-27    | 24.26   |

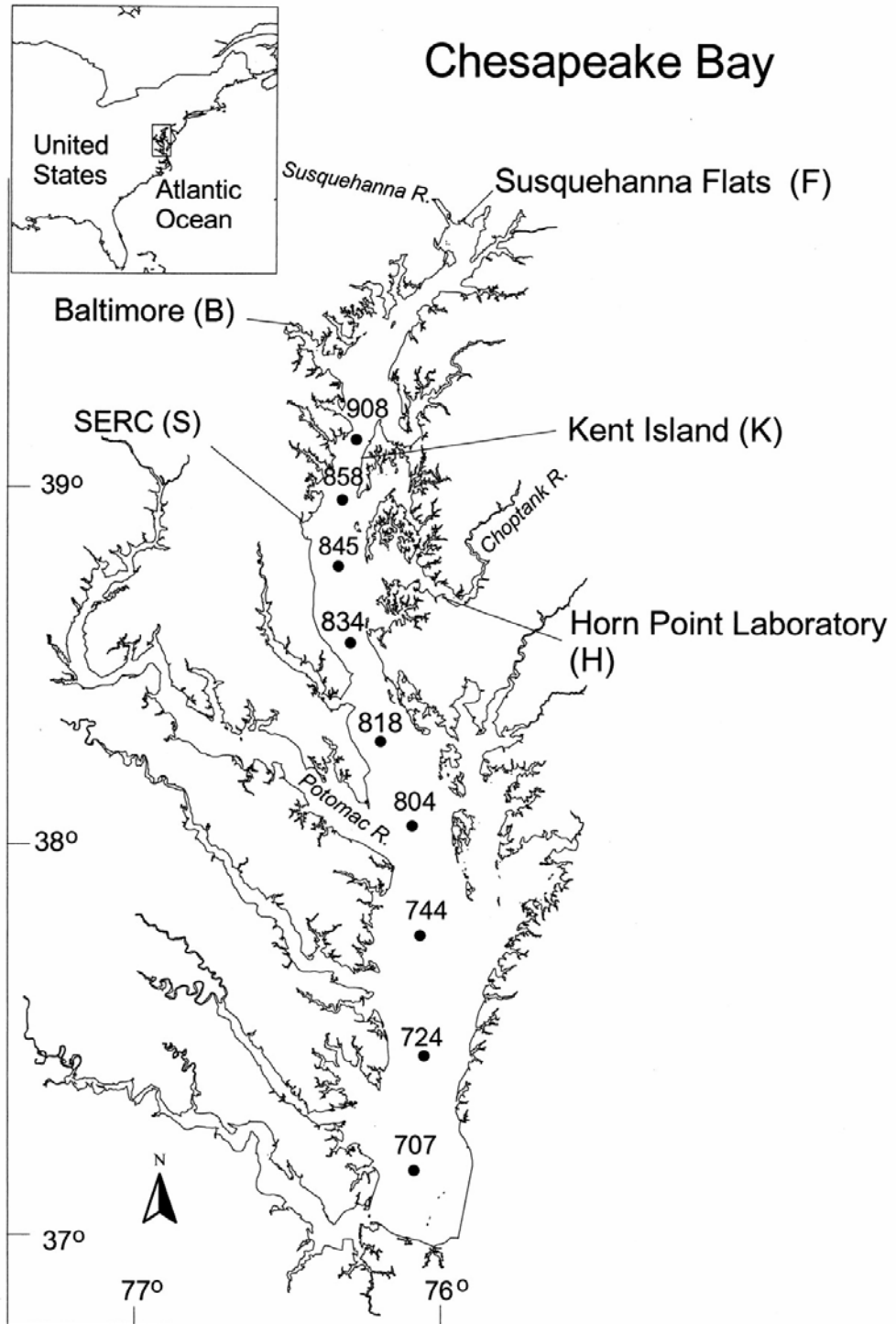


Figure 3.1. Map of locations of the sampling sites. (reprinted with permission of American Society of Microbiology and the authors, Louis, *et al.*, 2003 (111).



### 3.2.2 Sample collection

Water, small planktons, and large planktons for each sample were collected. Water comprised surface water collected using sterile 8-liter carboys. Planktons were collected using a plankton net that was either 20 or 64  $\mu\text{m}$  mesh size. Plankton collection was accomplished either by towing or pumping water through the nets for 5 min (ca.500 liters). Small planktons were collected by filter the sample through the 64  $\mu\text{m}$  net and captured on the 20  $\mu\text{m}$  plankton net and large planktons were collected by filter the sample through the 64  $\mu\text{m}$  net. Environmental parameters, i.e., pH, temperature, and salinity of each sample were recorded. All samples were transported at ambient temperature to the Center of Marine Biotechnology, Baltimore (COMB), for further processing, which was accomplished within six hours of collection.

### 3.2.3 Direct detection by PCR (Non-culture based method):

The method was employed in analyzing water samples from January, 1998, to March, 1999.

**DNA extraction:** Water samples (25 ml) were concentrated onto 0.2  $\mu\text{m}$ -polycarbonate membranes (Osmonics, Inc., CA) on the day of sampling and stored at -20°C in 2 ml of either SET or PBS buffer until processed. Total DNA was extracted from the membranes using CTAB/phenol/chloroform extraction (152). The DNA was dissolved in 100  $\mu\text{l}$  TE buffer adjusted to pH 8.0.

**VC-PCR reaction:** One microliter of the DNA extract was used as template for the PCR reaction. *V. cholerae* specific PCR (VC-PCR) was done essentially as described below, except that the MasterTaq kit (Eppendorf, Brinkmann Instruments, Inc., NY) was

used with 0.5 x TaqMaster PCR enhancer added to enhance the PCR reaction and compensate for any inhibitor effect that might occur in the DNA extracts.

**16S-PCR reaction:** The 16S PCR reaction, using universal primers amplifying 16S ribosomal DNA of bacteria, was done in parallel with VC-PCR to serve as a positive control. The results were used to assess the ability of the extracted DNA to be amplified, since bacteria were present in the samples at all times (determined by bacterial direct count by DAPI-DNA staining). The universal 16S-PCR primers employed were p16S-A (CAG CMG CCG CGG TAA TWC) and 16S-C (ACG GGC GGT GTG TRC) (3). The PCR conditions were the same as for VC-PCR, with 1x TaqMaster PCR enhancer added. The expected amplification product size of 888 bp indicated good quality, amplifiable DNA template.

#### **3.2.4 Isolation of *V. cholerae***

Two isolation methods used to obtain the *V. cholerae* isolates, enrichment and direct plating, as described below.

##### **A Enrichment method**

The detailed procedure was as described above (Chapter 2). In brief, the samples (250 ml for water and 25 ml for plankton) were filtered onto 0.22 µm membrane filters, incubated in APW enrichment medium, and streaked onto the selective medium, TCBS. Six bacterial colonies showing *V. cholerae*-like characteristics were selected from each sample. The presumptive *V. cholerae* isolates were screened by the simple procedure for rapid identification scheme (Chapter 2) and confirmed by using either PCR or colony hybridization with the *V. cholerae* specific probe (pVC-ITS-1). Artificial clones, defined

here as the bacterium that amplified in the enrichment step from the same original cell in the sample. These artificial clones gave bias to the true abundance of the traits studied, therefore were removed from the isolate collection. The artificial clones were identified as isolates with an identical genomic fingerprint (see Chapter 4), identical phenotypic and genotypic traits, and present in the same enrichment flask.

## **B Direct plating method**

Isolation of *V. cholerae* by direct plating was conducted for one full year, from January to December, 1999, at the shore sampling sites. Plankton fractions were homogenized with a glass tissue grinder before further processing. Dilutions of water and homogenized plankton were spread onto alkaline peptone agar (APA: 1% wt/vol bacto-peptone; 1% wt/vol NaCl; pH 8.6; 1.5% agar), and incubated for 3-5 days at 15 °C. Plates containing 30-300 colonies were blotted, using sterile Whatman filter paper no. 541, and hybridized with *V. cholerae* specific probe, as described in the confirmation of *V. cholerae* isolates section below. Additional filter paper with colonies of the type strains *V. cholerae* and *V. mimicus* were processed in parallel with the environmental samples to serve as positive and negative controls, respectively.

After the positive colonies were identified by dark spots on x-ray film, *V. cholerae* isolates were picked from the original plates by scraping the colonies with an inoculating loop and streaking on TCBS and 2% NaCl-LB plates. The isolates were subsequently confirmed by PCR.

### **3.2.5 Confirmation of *V. cholerae* isolates**

#### **A VC-PCR**

The PCR primers used for specific detection of *V. cholerae* were pVC-F2 (5' TTAAGCSTTTTCRCTGAGAATG 3') and pVCM-R1 (5' AGTCACTTAACCATAACAACCCG 3') based on 16S-23S rRNA intergenic spacer regions of *V. cholerae* developed by Chun *et al.* (28). The PCR templates were prepared by boiling the cell suspensions in 500 µl double distilled water in a boiling water bath for 10 minutes. The boiled suspensions were briefly centrifuged and diluted 1:1000 fold in sterile double distilled water. Then 5 µl of the dilution was used as PCR template. The PCR reaction was done in 25 µl reaction volume, consisting of 1X PCR buffer (50 mM KCl, 10 mM Tris-HCl (pH 9.0 at 25°C), 1.5 mM MgCl<sub>2</sub>, and 0.1% Triton<sup>®</sup> X-100), 200 µM of each dNTP, 800nM of each primer, and 0.625 units Taq Polymerase (Promega, Madison, WI). The amplification reaction was performed in a thermal cycler (MJ Research PTC-200 Peltier Thermal cycler) with calculated control option. The cycle condition was initial denaturation at 94°C for 2 min, followed by 35 cycles of 94°C for 30 sec, 60°C for 1 min and 72°C for 1 min, and one final extension step at 72°C for 5 minutes. Ten microliters of the amplified product were separated by electrophoresis using 1% agarose gel containing 3 ng/ml ethidium bromide and visualized under UV light. Isolates that yield the expected amplification product size of 295-310 bp were identified as *V. cholerae*.

#### **B Colony Hybridization**

This technique was used as alternative confirmation tool to PCR when large numbers of isolates were required to be tested. The bacteria were dot-inoculated, using

the tip of an inoculating needle, onto 2% NaCl-LB agar plates (up to 50 isolates can be inoculated onto the same agar plate). Type cultures of *V. cholerae* and *V. mimicus* were inoculated on the same plates serving as positive and negative controls respectively. After overnight incubation at 30 °C, the colonies were transferred to Whatman filter paper no. 541 (Whatman International Ltd., England) and processed by a method modified from that of Wright, *et al.* (1993) (199). In brief, the filter paper was overlaid on colonies on the plates for 1 hr., transferred colony side-up onto blotting paper soaked with alkaline buffer (1.5 M NaCl, 0.5 M NaOH) and microwaved for 2 minutes to release and denature the bacterial DNA. The filters were neutralized in 2 M ammonium acetate for 5 min, washed twice in 1X SSC buffer adjusted to pH 7.0 for 10 min each, allowed to air dry, and stored at -20°C until further processed. A modified pVCF2, termed pVC-ITS1 (GCSTTTTCRCTGAGAATG), end labeled with <sup>32</sup>P radioisotope, was used as the *V. cholerae* specific probe. Prior to hybridization, the blots were washed twice with 3X SSC-0.1% SDS at room temperature for 10 min, and once at 65°C for 4-16 hr. The hybridization reaction was performed using the sodium chloride/sodium citrate method described as standard protocol by Duby *et al.* (1988) (51). Blots were pre-hybridized at 37°C for 1 hr, hybridized at 42 °C overnight, and washed 3 times, 15 min each, with ample volume of 6x SSC-0.05% sodium pyrophosphate at room temperature, followed by stringent wash at 55°C for 30 min. The positive colonies were identified by a dark spot after autoradiography at -70°C overnight.

### **3.2.6 Characterization of *V. cholerae* isolates**

Phenotypic and genotypic traits of the *V. cholerae* isolates were assessed using a selected set of biochemical tests (see below) and dot blot hybridization with gene probes.

Luminescence and possession of *luxA* gene were detected in a collaborative study with C. Grim (73). The primer sequences for the *luxA* gene probe were included for completeness. The subsequent protocols for probe generation and hybridization were carried out essentially the same way as with other gene probes.

## **A Phenotypic characterization**

A list of phenotypic characteristics employed in this study included O1/O139 serotyping, luminescence, arginine dihydrolase, esculin hydrolysis, growth in nutrient broth containing different concentrations of NaCl, acid production from sucrose, arabinose, mannose, mannitol, lysine and ornithine decarboxylase, methyl red, Voges-Proskauer, oxidase, gelatinase, amylase, lipase (corn oil), chitinase, sensitivity to vibriostatin agent O/129, sensitivity to polymyxin B, and growth at 42 °C (see Table 3.5 in the Results section).

## **B Genotypic characterization**

Presence of nine selected genes (six toxin-related genes, two membrane protein coding genes, and one luminescence coding gene) in each isolate was determined by dot blot hybridization, as described below.

**DNA extraction:** Genomic DNA from each isolate was extracted using the DNeasy<sup>®</sup> Tissue Kit (Qiagen Inc., Valencia, CA) and eluted in 200 µl elution buffer AE (10 mM Tris-HCl pH 9.0, 0.5 mM EDTA). DNA concentrations were determined spectrophotometrically from absorbance at 260 nm ( $A_{260}$ ) (158). DNA was then diluted to 20 ng/µl in AE buffer and stored at -20 °C until analyzed.

**Dot blot:** Genomic DNA (500 ng) was denatured in 0.4 M NaOH, 10 mM EDTA and heated to 100 °C for 10 min to ensure complete denaturation. The DNA was cooled on ice and mixed with an equal volume of 6x SSC prior to being dot blotted onto a MagnaCharge Nylon membrane (MSI, Micron separations Inc., Massachusetts, USA) pre-wet with distilled water and soaked in 6x SSC before use (MSI manufacture's protocol). The Bio-Dot microfiltration apparatus (Bio-Rad Laboratories, Hercules, CA) was used following the manufacturer's instructions. After assembling the apparatus, the membranes were washed with 500 µl of 6x SSC before and after applying DNA samples. Slow vacuum was applied in all steps. After blotting, the DNA was immobilized on the nylon membrane by UV cross-linking (UV crosslinker, Fisher Scientific, Pittsburgh, PA) at an optimal cross-link setting (120 mJ/cm<sup>2</sup>, 30 sec). Each membrane was then rinsed briefly with distilled water, air dried, and placed between two sheets of dry filter paper, sealed in a plastic bag, and stored at -20 °C until analyzed. Genomic DNA from *V. cholerae* (RC2 classical O1, RC4 El Tor O1, and RC66 non-O1 *stn* positive), *V. mimicus* (RC5), and an archaeal bacterial DNA (*Aeropyrum pernix*) were included on the blots as positive and negative controls.

**Gene probe preparation:** The probes were generated using the PCR Dig Probe Synthesis Kit (Roche Diagnostics GmbH, Mannheim, Germany). Primer sequences for amplification of each gene probe are listed in Table 3.2. All probes were amplified using 40 ng of the template genomic DNA in 50 µl reaction volumes. A reference strain, RC145 (ATCC 39315, *V. cholerae* O1 El Tor), for which full genome has been sequenced (78), was used as the template for generation of all gene probes, except for the *stn* and *luxA* probes, in which case the non-O1 environmental isolate (RC66), and *V.*

*cholerae* biotype *albensis* (RC 293) were used, respectively. The PCR conditions for the labeling reaction were initial denaturation at 94 °C for 4 min, followed by 30 cycles of denaturation at 94 °C for 45 sec, annealing at 60 °C for 45 sec, extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The amplified products were analyzed using 1% agarose gel electrophoresis and the corresponding bands were cut from the agarose gel and purified using the QIAquick® Gel Extraction kit (Qiagen Inc.).

Table 3.2. Primer sequences employed for gene probe amplification

| Gene                                 | Primer | Sequence (5'-3')                   | Amplicon size (bp) | References |
|--------------------------------------|--------|------------------------------------|--------------------|------------|
| <i>ctxA</i>                          | 94F    | CGG GCA GAT TCT AGA CCT CCT G      | 564                | (67)       |
|                                      | 614R   | CGA TGA TCT TGG AGC ATT CCC AC     |                    |            |
| <i>zot</i>                           | 225F   | TCG CTT AAC GAT GGC GCG TTT T      | 947                | (151)      |
|                                      | 1129R  | AAC CCC GTT TCA CTT CTA CCC A      |                    |            |
| <i>tcpA</i><br>(ElTor/<br>O139)      | 72F    | CAC GAT AAG AAA ACC GGT CAA<br>GAG | 451                | (101)      |
|                                      | 477R   | CGA AAG CAC CTT CTT TCA CGT TG     |                    |            |
| <i>toxR</i>                          | 101F   | CCT TCG ATC CCC TAA GCA ATA C      | 779                | (151)      |
|                                      | 837 R  | AGG GTT AGC AAC GAT GCG TAA G      |                    |            |
| <i>hlyA</i><br>(ElTor/<br>classical) | 489F   | GGC AAA CAG CGA AAC AAA TAC C      | 738/727            | (166)      |
|                                      | 1184R  | CTC AGC GGG CTA ATA CGG TTT A      |                    |            |
| <i>stn</i>                           | 67F    | TCG CAT TTA GCC AAA CAG TAG<br>AAA | 172                | (151)      |
|                                      | 194R   | GCT GGA TTG CAA CAT ATT TCG C      |                    |            |
| <i>ompU</i>                          | 80F    | ACG CTG ACG GAA TCA ACC AAA G      | 869                | (151)      |
|                                      | 906R   | GCG GAA GTT TGG CTT GAA GTA G      |                    |            |
| <i>ompW</i>                          | F-1    | CAC CAA GAA GGT GAC TTT ATT<br>GTG | 588                | (129)      |
|                                      | R-2    | GAA CTT ATA ACC ACC CGC G          |                    |            |
| <i>luxA</i>                          | F-1    | CGA AGC GGT TTG GTT GCT A          | 650                | (73)       |
|                                      | R-1    | CGG GTA GCA TTG ACG TAG GA         |                    |            |



**Hybridization and detection:** Hybridization and detection were done using Dig hybridization and detection for chemiluminescence with the CSPD protocol (Roche Diagnostics GmbH). In brief, the membranes were pre-hybridized at 42 °C in Dig-EasyHyb solution for 30 min and hybridized in the presence of each gene probe (1 µl/ml) at 42 °C overnight. For *luxA* detection, the pre-hybridization and hybridization were done at 45 °C. The membranes were then subjected to a low stringency wash with SSC-0.1% SDS for 5 min, twice at room temperature, and high stringency wash with 0.5x SSC-0.1% SDS for 15 min, twice at 65°C. The signals were detected using the CSPD chemiluminescent substrate and autoradiography was performed at room temperature for 5-30 min. The signal intensities were recorded semi-quantitatively on a scale of 0-4.

### 3.2.7 Statistical analysis

Phenotypic traits were coded as binary data (positive/negative response). The signal intensities from the dot-blot hybridizations were recoded into five classes, 0-4 (0 for negative and 1-4 for low to high positive, respectively) as nominal data. When a probe generated only two responses i.e., strong intensity or no signal, the data were coded as binary (absence/presence).

A series of nonparametric analyses were performed on the phenotypic and genotypic trait data by appropriate procedures using statistical software, SPSS version 11.5 (SPSS Inc., Chicago, Illinois) with exact method for calculation. Chi-square test was used to test even distribution of the traits among the strain set. To determine correlation among pairs of traits, the Spearman rank correlation was used. To test association between a pair of binary traits or multi-nomial traits, the McNemar test or marginal homogeneity test was used, respectively (172). Binary response were

considered as negative (response level = 0) or strong positive (response level = 4) when compared with multi-nomial response. For all the tests, unless specified, significance was declared at type I error level of  $<0.05$ .

Simpson's diversity index, widely used in ecological research to measure richness and evenness of species distribution in a community (160) was used to calculate the diversity of isolates obtained by each isolation method. Isolates with the same characteristics (by phenotype, genotype, and genomic fingerprint) were considered to belong to the same clone. The diversity index was calculated using the equation  $DI_m = 1 - \sum (n_i/n_t)^2$ ; where  $DI_m$  is the Simpson's diversity index of method  $m$ ,  $n_i$  is number of isolates belonging to the  $i^{\text{th}}$  clone, and  $n_{tm}$  is total number of isolates from the method  $m$ .

### **3.3 Results**

#### **3.3.1 Direct detection of *V. cholerae* in water samples by PCR**

During the time period when direct detection was done (January, 1998, to March, 1999), DNA extracted from water samples gave positive signals, indicating presence of *V. cholerae* at least at one sampling site from January to November 1998, and February and March 1999. No positive signals were detected from any of the sampling sites in December 1998 and January 1999 (Table 3.3). This result shows that *V. cholerae* in water samples can be detected by this method most of the year, except in the coldest months of December and January. Results for three of the samples were unable to be interpreted because of negative results when universal 16S-PCR was employed, which indicated that the quality of the DNA template was not suitable for PCR amplification.

#### **3.3.2 Isolation of *V. cholerae* from Chesapeake Bay**

A set of isolates, the total of which was 278, was obtained by the end of the study. Among these isolates, there were 54 artificial clones from the enrichment step and these were removed from the collection. The final set of 224 *V. cholerae* was subjected to analysis as described below. Sources and environmental factors associated with the isolates are provided in Table 3.4.

Table 3.3. Number of positive samples by direct detection using PCR <sup>a</sup>

| Month  | Sampling site <sup>b</sup> |    |    |    |   | Total Positive |
|--------|----------------------------|----|----|----|---|----------------|
|        | B                          | F  | H  | K  | S |                |
| Jan-98 |                            | ND | 1  | 1  | 1 | 3              |
| Feb-98 |                            | 1  | 0  | 1  | 0 | 2              |
| Mar-98 |                            | 1  | 1  | 1  | 1 | 4              |
| Apr-98 |                            | 0  | 1  | 1  | 1 | 3              |
| May-98 |                            | 1  | 1  | 1  | 1 | 4              |
| Jun-98 |                            | 1  | 1  | 1  | 1 | 4              |
| Jul-98 |                            | 1  | 1  | 1  | 1 | 4              |
| Aug-98 |                            | 1  | 1  | 1  | 1 | 4              |
| Sep-98 |                            | 1  | 1  | 1  | 1 | 4              |
| Oct-98 |                            | ND | 1  | 1  | 1 | 3              |
| Nov-98 |                            | 1  | 0  | 0  | 0 | 1              |
| Dec-98 |                            | ND | 0  | 0  | 0 | 0              |
| Jan-99 | 0                          | 0  | 0  | 0  | 0 | 0              |
| Feb-99 | 0                          | 0  | 0  | 1  | 0 | 1              |
| Mar-99 | 0                          | 0  | 1  |    |   | 1              |
| Total  | 0                          | 8  | 10 | 11 | 9 | 38             |

<sup>a</sup> 1 = positive for presence of *V. cholerae* (VC-PCR positive), 0 = negative for presence of *V. cholerae* (16S-PCR positive, VC-PCR negative), ND = presence of *V. cholerae* cannot be determined (both 16S-PCR and VC-PCR negative).

<sup>b</sup> Sampling site: B = Baltimore harbor, F = Susquehanna River Flat, H = Horn Point Laboratory, K = Kent Island, S = Smithsonian Environmental Research Center (SERC).

Table 3.4. Number of isolates and relevant information for each sample<sup>a</sup>

| Sample ID | Site | Fraction | Method | pH  | Temp (°C) | Salinity (ppt) | Date      | No. of Isolates |
|-----------|------|----------|--------|-----|-----------|----------------|-----------|-----------------|
| F8414PE   | F    | P        | E      | 7.2 | 16.0      | 0.0            | 4/14/1998 | 2               |
| H8416WE   | H    | W        | E      | 8.0 | 17.0      | 6.0            | 4/16/1998 | 4               |
| S8416ZE   | S    | Z        | E      | 8.0 | 18.0      | 3.0            | 4/16/1998 | 1               |
| H8528WE   | H    | W        | E      | 8.0 | 24.0      | 5.0            | 5/28/1998 | 4               |
| H8528PE   | H    | P        | E      | 8.0 | 24.0      | 5.0            | 5/28/1998 | 2               |
| H8528ZE   | H    | Z        | E      | 8.0 | 24.0      | 5.0            | 5/28/1998 | 2               |
| K8528PE   | K    | P        | E      | 7.9 | 21.5      | 2.0            | 5/28/1998 | 1               |
| S8528WE   | S    | W        | E      | 8.1 | 21.0      | 2.0            | 5/28/1998 | 2               |
| S8528PE   | S    | P        | E      | 8.1 | 21.0      | 2.0            | 5/28/1998 | 1               |
| H8608WE   | H    | W        | E      | 7.8 | 22.0      | 5.0            | 6/8/1998  | 1               |
| H8608PE   | H    | P        | E      | 7.8 | 22.0      | 5.0            | 6/8/1998  | 1               |
| H8608ZE   | H    | Z        | E      | 7.8 | 22.0      | 5.0            | 6/8/1998  | 1               |
| K8609WE   | K    | W        | E      | 8.0 | 19.5      | 5.0            | 6/9/1998  | 2               |
| K8609PE   | K    | P        | E      | 8.0 | 19.5      | 5.0            | 6/9/1998  | 1               |
| K8609ZE   | K    | Z        | E      | 8.0 | 19.5      | 5.0            | 6/9/1998  | 4               |
| S8609WE   | S    | W        | E      | 8.1 | 21.0      | 5.0            | 6/9/1998  | 3               |
| S8609ZE   | S    | Z        | E      | 8.1 | 21.0      | 5.0            | 6/9/1998  | 1               |
| H8623WE   | H    | W        | E      | 7.6 | 28.0      | 5.0            | 6/23/1998 | 6               |
| H8623PE   | H    | P        | E      | 7.6 | 28.0      | 5.0            | 6/23/1998 | 4               |
| H8623ZE   | H    | Z        | E      | 7.6 | 28.0      | 5.0            | 6/23/1998 | 2               |
| K8623WE   | K    | W        | E      |     | 25.0      | 7.0            | 6/23/1998 | 4               |
| K8623PE   | K    | P        | E      |     | 25.0      | 7.0            | 6/23/1998 | 3               |
| K8623ZE   | K    | Z        | E      |     | 25.0      | 7.0            | 6/23/1998 | 3               |
| S8623WE   | S    | W        | E      | 8.1 | 27.0      | 6.0            | 6/23/1998 | 5               |
| S8623PE   | S    | P        | E      | 8.1 | 27.0      | 6.0            | 6/23/1998 | 4               |
| S8623ZE   | S    | Z        | E      | 8.1 | 27.0      | 6.0            | 6/23/1998 | 1               |
| H8623WE   | H    | W        | E      | 7.6 | 28.0      | 5.0            | 6/23/1998 | 3               |
| H8623PE   | H    | P        | E      | 7.6 | 28.0      | 5.0            | 6/23/1998 | 2               |
| H8623ZE   | H    | Z        | E      | 7.6 | 28.0      | 5.0            | 6/23/1998 | 1               |
| K8623WE   | K    | W        | E      |     | 25.0      | 7.0            | 6/23/1998 | 5               |
| K8623ZE   | K    | Z        | E      |     | 25.0      | 7.0            | 6/23/1998 | 5               |
| S8623WE   | S    | W        | E      | 8.1 | 27.0      | 6.0            | 6/23/1998 | 4               |
| S8623PE   | S    | P        | E      | 8.1 | 27.0      | 6.0            | 6/23/1998 | 2               |
| S8623ZE   | S    | Z        | E      | 8.1 | 27.0      | 6.0            | 6/23/1998 | 1               |
| F8624PE   | F    | P        | E      | 8.3 | 26.0      | 0.0            | 6/24/1998 | 2               |
| H8709WE   | H    | W        | E      | 7.8 | 25.0      | 6.0            | 7/9/1998  | 7               |
| H8709PE   | H    | P        | E      | 7.8 | 25.0      | 6.0            | 7/9/1998  | 3               |
| H8709ZE   | H    | Z        | E      | 7.8 | 25.0      | 6.0            | 7/9/1998  | 1               |
| K8709WE   | K    | W        | E      | 8.0 | 25.0      | 6.0            | 7/9/1998  | 6               |
| K8709PE   | K    | P        | E      | 8.0 | 25.0      | 6.0            | 7/9/1998  | 6               |
| K8709ZE   | K    | Z        | E      | 8.0 | 25.0      | 6.0            | 7/9/1998  | 5               |
| S8709WE   | S    | W        | E      | 8.0 | 26.0      | 6.0            | 7/9/1998  | 5               |
| S8709PE   | S    | P        | E      | 8.0 | 26.0      | 6.0            | 7/9/1998  | 5               |
| S8709ZE   | S    | Z        | E      | 8.0 | 26.0      | 6.0            | 7/9/1998  | 2               |
| K8721WE   | K    | W        | E      | 8.0 | 30.0      | 6.2            | 7/21/1998 | 1               |
| K8721PE   | K    | P        | E      | 8.0 | 30.0      | 6.2            | 7/21/1998 | 1               |
| K8721ZE   | K    | Z        | E      | 8.0 | 30.0      | 6.2            | 7/21/1998 | 3               |
| S8721WE   | S    | W        | E      | 8.0 | 31.0      | 8.0            | 7/21/1998 | 1               |
| S8721ZE   | S    | Z        | E      | 8.0 | 31.0      | 8.0            | 7/21/1998 | 1               |
| H8805WE   | H    | W        | E      | 7.3 | 28.0      | 9.5            | 8/5/1998  | 2               |

Table 3.4 (continue)

| Sample ID                       | Site | Fraction | Method | pH  | Temp (°C) | Salinity (ppt) | Date      | No. of Isolates |
|---------------------------------|------|----------|--------|-----|-----------|----------------|-----------|-----------------|
| H8805PE                         | H    | P        | E      | 7.3 | 28.0      | 9.5            | 8/5/1998  | 4               |
| H8805ZE                         | H    | Z        | E      | 7.3 | 28.0      | 9.5            | 8/5/1998  | 1               |
| K8805WE                         | K    | W        | E      | 7.6 | 26.0      | 10.0           | 8/5/1998  | 1               |
| K8805ZE                         | K    | Z        | E      | 7.6 | 26.0      | 10.0           | 8/5/1998  | 1               |
| S8805WE                         | S    | W        | E      | 7.7 | 28.0      | 9.0            | 8/5/1998  | 1               |
| S8805PE                         | S    | P        | E      | 7.7 | 28.0      | 9.0            | 8/5/1998  | 1               |
| S8805ZE                         | S    | Z        | E      | 7.7 | 28.0      | 9.0            | 8/5/1998  | 1               |
| B9422WE                         | B    | W        | E      | 7.9 | 12.5      | 5.0            | 4/22/1999 | 1               |
| B9422PE                         | B    | P        | E      | 7.9 | 12.5      | 5.0            | 4/22/1999 | 1               |
| B9607WE                         | B    | W        | E      | 7.7 | 26.5      | 8.0            | 6/7/1999  | 2               |
| S9629WH                         | S    | W        | H      | 7.9 | 29.5      | 10.5           | 6/29/1999 | 5               |
| S9629PH                         | S    | P        | H      | 7.9 | 29.5      | 10.5           | 6/29/1999 | 3               |
| S9629WE                         | S    | W        | E      | 7.9 | 29.5      | 10.5           | 6/29/1999 | 2               |
| S9629ZE                         | S    | Z        | E      | 7.9 | 29.5      | 10.5           | 6/29/1999 | 1               |
| H9708PH                         | H    | P        | H      | 7.8 | 28.5      | 11.0           | 7/8/1999  | 3               |
| H9708ZH                         | H    | Z        | H      | 7.8 | 28.5      | 11.0           | 7/8/1999  | 4               |
| K9708PH                         | K    | P        | H      | 8.0 | 27.0      | 10.5           | 7/8/1999  | 3               |
| K9708ZH                         | K    | Z        | H      | 8.0 | 27.0      | 10.5           | 7/8/1999  | 4               |
| S9708PH                         | S    | P        | H      | 7.7 | 29.0      | 11.0           | 7/8/1999  | 7               |
| S9708ZH                         | S    | Z        | H      | 7.7 | 29.0      | 11.0           | 7/8/1999  | 2               |
| B9720PE                         | B    | P        | E      | 7.2 | 27.0      | 9.5            | 7/20/1999 | 1               |
| F9720WE                         | F    | W        | E      | 8.1 | 28.5      | 0.0            | 7/20/1999 | 1               |
| H9809WE                         | H    | W        | E      | 8.1 | 25.0      | 11.5           | 8/9/1999  | 2               |
| K9809WE                         | K    | W        | E      | 8.2 | 25.0      | 12.0           | 8/9/1999  | 3               |
| K9809PE                         | K    | P        | E      | 8.2 | 25.0      | 12.0           | 8/9/1999  | 2               |
| K9809ZE                         | K    | Z        | E      | 8.2 | 25.0      | 12.0           | 8/9/1999  | 2               |
| F9810WE                         | F    | W        | E      | 9.1 | 26.0      | 0.0            | 8/10/1999 | 2               |
| F9810ZE                         | F    | Z        | E      | 9.1 | 26.0      | 0.0            | 8/10/1999 | 1               |
| S9830WH                         | S    | W        | H      | 8.0 | 24.0      | 13.0           | 8/30/1999 | 3               |
| 59920WE                         | 858  | W        | E      |     | 22.3      | 14.2           | 9/20/1999 | 1               |
| 49920WE                         | 845  | W        | E      |     | 22.7      | 16.1           | 9/20/1999 | 1               |
| 39920WE                         | 834  | W        | E      |     | 22.9      | 16.4           | 9/20/1999 | 1               |
| 39920ZE                         | 834  | Z        | E      |     | 22.9      | 16.4           | 9/20/1999 | 1               |
| 19920WE                         | 818  | W        | E      |     | 23.0      | 17.7           | 9/20/1999 | 1               |
| B9927WE                         | B    | W        | E      | 7.7 | 21.5      | 10.0           | 9/27/1999 | 1               |
| K9928WE                         | K    | W        | E      | 8.3 | 21.0      | 11.5           | 9/28/1999 | 1               |
| 90619WE                         | 908  | W        | E      |     | 23.7      | 4.7            | 6/19/2000 | 1               |
| 90619ZE                         | 908  | Z        | E      |     | 23.7      | 4.7            | 6/19/2000 | 1               |
| 50619WE                         | 858  | W        | E      |     | 23.9      | 5.8            | 6/19/2000 | 2               |
| 40619ZE                         | 845  | Z        | E      |     | 23.9      | 9.4            | 6/19/2000 | 1               |
| 00620WE                         | 804  | W        | E      |     | 24.0      | 12.8           | 6/20/2000 | 1               |
| 90819WE                         | 908  | W        | E      |     | 23.6      | 8.7            | 8/19/2000 | 2               |
| 50819WE                         | 858  | W        | E      |     | 24.0      | 10.0           | 8/19/2000 | 1               |
| 30819WE                         | 834  | W        | E      |     | 24.6      | 12.7           | 8/19/2000 | 2               |
| 90819WE                         | 908  | W        | E      |     | 23.6      | 8.7            | 8/19/2000 | 1               |
| 70820ZE                         | 707  | Z        | E      |     | 24.7      | 21.2           | 8/20/2000 | 1               |
| <b>Total number of isolates</b> |      |          |        |     |           |                |           | <b>224</b>      |

<sup>a</sup> Site, see Table 3.1 for full site description; Fraction, W = water, P = small

plankton, Z = large plankton; Method, E = enrichment, H = direct plating followed with colony hybridization.

## **A Distribution of isolates**

More isolates of *V. cholerae* from Chesapeake Bay were obtained from water than from the two plankton fractions. Distribution of the isolates among the sample fractions was 46% from water, 29 % from small plankton, and 25 % from large plankton. However, when the plankton samples were combined, approximately half of the isolates were from plankton. The majority of the isolates were from three brackish shore sampling sites (S, K, and H). Each site contributed a similar number of isolates ( $64 \pm 4$ ) (Fig. 3.2). Notably, F, the only fresh water site included in this study, yielded a very low number of isolates (i.e., 8 isolates), even though the number of samplings was equal to that of the other three sites. There were fewer isolates from the B site, which was included later during the year 1999, and from cruises with four sampling collections. Nevertheless, *V. cholerae* was isolated from all Chesapeake Bay sites, from the northernmost point at the Susquehanna River (site F) to the mouth of the Bay (site 707).

## **B Environmental conditions where *V. cholerae* was isolated**

Environmental parameters of sites where *V. cholerae* was isolated: pH 7.2-9.1, with a mean of  $7.9 \pm 0.48$ ; temperature 12.5-31°C, with a mean of  $25.5 \pm 3.2^\circ\text{C}$ ; and salinity 0-21.2 ppt, with a mean of  $7.4 \pm 3.27$ . Thus *V. cholerae* was isolated under a wide range of environmental conditions, but conditions correlated with a larger number of isolates were a narrow range of slightly alkaline pH, warm temperature, and mildly saline water.

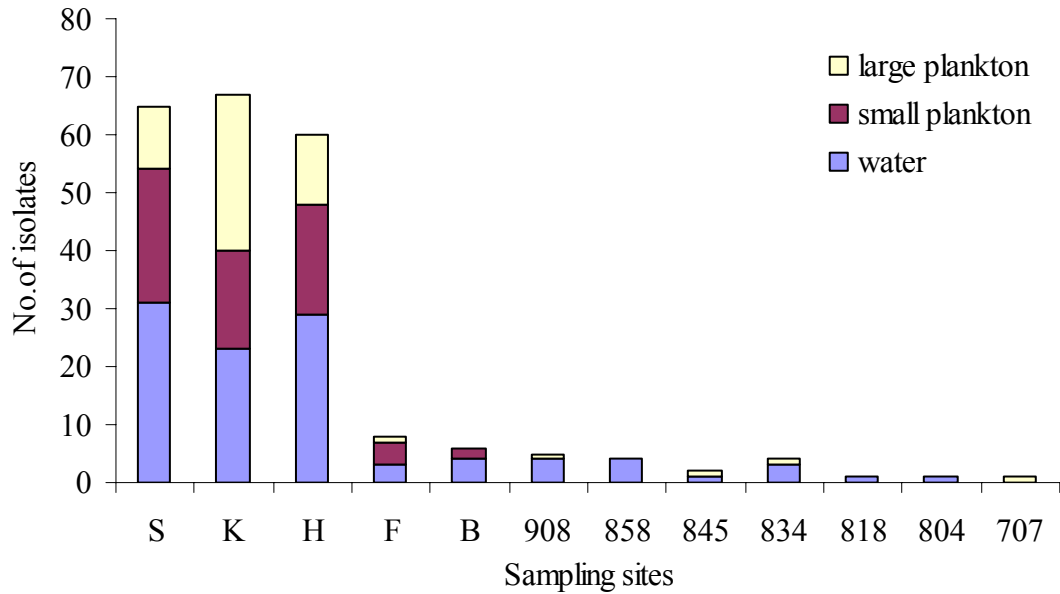


Figure 3.2. Distribution of isolates among sampling sites.

### C Seasonal and annual variation in the occurrence of *V. cholerae*

Results shown in Fig. 3.3 clearly demonstrate seasonal and annual variation in the occurrence of *V. cholerae* in Chesapeake Bay. *V. cholerae* was readily isolated in the spring, beginning in April and peaking in the summer (June, July, and August). Isolates were not obtained from October to March. In 1998, the total number of isolates was approximately 5 fold greater (149 versus 28) than in 1999. Notably, the salinity average in 1998 was lower than in 1999. The average salinity for all samples yielding isolates was 5.9 ppt in 1998, and 10.3 ppt in 1999. Other environmental factors i.e., pH and temperature were similar for both years (pH 7.9, 25.2°C in 1998 and pH 8.0, 25.0°C in 1999).



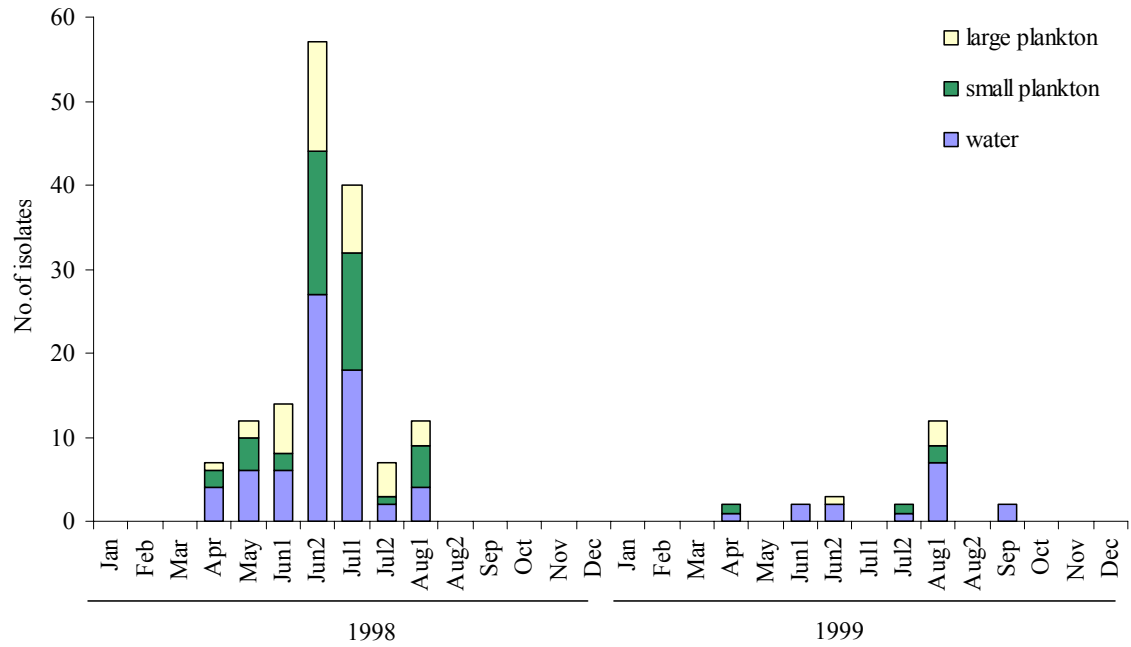


Figure 3.3. Seasonal and annual variation in the occurrence of *V. cholerae*. Number of isolates obtained from shore sampling sites by enrichment was plotted against time of sampling.

## **D Comparison of isolation methods**

Enrichment and direct plating methods were compared from the results obtained during the period that both methods were employed (1999). Simpson's diversity index indicated that enrichment yielded more diverse isolates, i.e., isolates with different characteristics, with a diversity index (DI) value of 0.949 compared with direct plating with a lower DI of 0.718. The number of isolates and their distribution in each sample are shown in Fig. 3.4.

It was noted that the majority of positive samples yielded isolates by one method but not by the other. Among these samples, 14 were positive only by enrichment, and 8 were positive only by direct plating. Although the detection limit of both methods was understood to differ, this result was interesting. Among the samples that were positive only by direct plating, the majority of the isolates (26/29) were from plankton. The remaining three isolates isolated from water were sucrose negative.

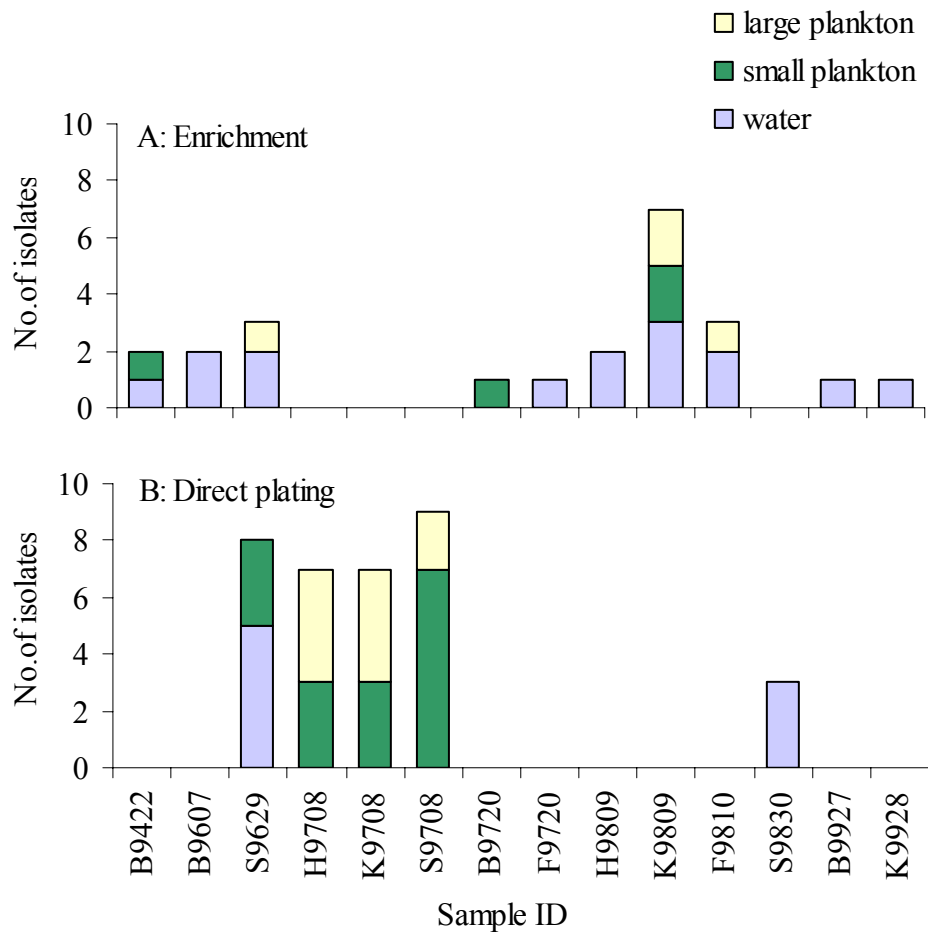


Figure 3.4. Number of isolates obtained from each sample by different isolation methods. Panel A shows number of isolates obtained from enrichment, panel B shows number of isolates from direct plating. Same position on x-axis indicates that the isolate was obtained from the same sampling site and time, different section in each column indicates different fractions of sample the strain was isolated from.

### **3.3.3 Characterization of isolates**

Results from phenotypic and genotypic characterization of the isolates were as follows:

#### **A Phenotypic characteristics**

None of the 224 isolates were serotype O1 or O139 by the monoclonal antibody agglutination test. Characteristics common to all isolates were negative responses to the arginine dihydrolase test, esculin hydrolysis, growth in 8% NaCl, and acid production from arabinose; and the positive responses were growth in 0, 1 and 3% NaCl, oxidase activity, and gelatin degradation (Table 3.5).

Table 3.5. Summary of phenotypic characteristics of the *V. cholerae* from Chesapeake Bay

| Test Name  | [abbreviation] | Number. Positive | % positive | Reference <sup>1</sup> | Reference <sup>2</sup> |
|--|----------------|------------------|------------|------------------------|------------------------|
| Serotype O1                                      |                | 0                | 0          | NA                     | NA                     |
| Serotype O139                                    |                | 0                | 0          | NA                     | NA                     |
| Luminescence                                     | [Lum]          | 117              | 52         | NA                     | 10                     |
| Arginine dihydrolase                             | [Arg]          | 0                | 0          | 0                      | 0                      |
| Esculin hydrolysis                               | [Esc]          | 0                | 0          | 0-2                    | 0-6                    |
| Growth in NaCl 0 %                               |                | 224              | 100        | 100                    | 90-99                  |
| Growth in NaCl 1 %                               |                | 224              | 100        | 100                    | 100                    |
| Growth in NaCl 3 %                               |                | 224              | 100        | NA                     | 100                    |
| Growth in NaCl 6 %                               | [NaCl6%]       | 200              | 89         | 53                     | 20-80                  |
| Growth in NaCl 8 %                               |                | 0                | 0          | 1                      | 0                      |
| Acid from sucrose                                | [Suc]          | 221              | 99         | 99                     | 75-100                 |
| Acid from arabinose                              | [Ara]          | 0                | 0          | 0                      | 0-3                    |
| Acid from mannose                                | [Mns]          | 139              | 62         | 78                     | 70-100                 |
| Acid from mannitol                               | [Mnt]          | 115              | 51         | 99                     | 91-100                 |
| Lysine decarboxylase                             | [Lys]          | 222              | 99         | 99                     | 99-100                 |
| Ornithine decarboxylase                          | [Orn]          | 222              | 99         | 99                     | 98-100                 |
| Methyl red                                       | [MR]           | 168              | 75         | 25-99                  | 34-100                 |
| Voges-Proskauer <sup>3</sup>                     | [VP]           | 192              | 86         | 74-93                  | 59-90                  |
| Oxidase  |                | 224              | 100        | 100                    | 100                    |
| Gelatinase                                       |                | 224              | 100        | 52-90                  | 99-100                 |
| Amylase  | [Amy]          | 205              | 92         | NA                     | 88-97                  |
| Chitinase  | [Chi]          | 223              | 100        | NA                     | 84-100                 |
| Lipase (corn oil)                                | [Lip]          | 222              | 99         | 92                     | 90-100                 |
| Growth at 42°C                                   | [42C]          | 219              | 98         | 100 <sup>5</sup>       | 90-100                 |
| Sensitivity to O/129, 150 mg                     | [V150]         | 223              | 100        | 99                     | 100                    |
| Sensitivity to O/129, 10 mg                      | [V10]          | 223              | 100        | 99                     | 83                     |
| Sensitivity to PolymyxinB, 50 Units <sup>4</sup> | [PB]           | 35               | 16         | 22                     | 11                     |

<sup>1</sup> Summarized from Manual of Clinical Microbiology (55): The value varies depending on set of strains. Majority of the isolates were from clinical sources.

<sup>2</sup> Summarized from references (33, 122), and (190): The set included mostly environmental and some clinical isolates.

<sup>3</sup> Classical biovar is negative, El Tor is positive

<sup>4</sup> Classical biovar is positive, El Tor is negative

<sup>5</sup> The test was done at 40°C

The luminescence and acid production from mannitol tests produced the two most variable results in which isolates were evenly distributed across positive and negative responses. The rest of the biochemical tests showed significant preferences on one particular response over the other.

When overall profiles of phenotypic characteristics among the Chesapeake Bay isolates from this study were compared with those previously published (Table 3.5), most of the results were in agreement. In general, the environmental isolates showed more variable responses than clinical isolates. For example, environmental isolates gave 75-100% positive response to acid production from sucrose while 99% of the clinical isolates gave a positive response. The most consistent tests that gave the same results were arginine dihydrolase, growth in NaCl 1-3%, and oxidase.

Some tests that usually are positive for *V. cholerae*, for occasional strains yielded a negative response, namely growth in NaCl 0%, acid from sucrose, lysine and ornithine decarboxylase, gelatinase, amylase, lipase corn oil, chitinase, growth at 42°C, and sensitivity to O/129 vibriostatic agent. The tests that usually are negative for *V. cholerae* but positive strains have been reported (references in Table 3.5), were esculin hydrolysis, acid from arabinose, and growth in NaCl 8%. Notably, for low percent of the esculin hydrolysis positive strain reported, the result may be misinterpretation of the darkening of the medium, without actual hydrolysis of the esculin molecule, which required testing for loss of fluorescence (see Chapter 2).

Tests that gave variable results among different isolates were growth at 6% NaCl, acid from mannose, methyl red, Voges-Proskauer, and sensitivity to polymyxin B. The last two tests were used to differentiate between the classical and El Tor biotype. The

classical biotype is negative for Voges-Proskauer and sensitive to polymyxin B, while the El Tor is positive for Voges-Proskauer and resistant to polymyxin B.

Tests that differentiate the Chesapeake Bay isolates were acid from mannitol and luminescence, in which 51% of the Chesapeake Bay isolates were positive for mannitol while other reports cite more than 90% positive, and a higher percent of luminescence (52%) than previously reported (10%).

## B Genotypic characteristics

Presence of six toxin-related genes (*toxR*, *ctxA*, *zot*, *tcpA*, *stn* and *hlyA*), two outer-membrane porin genes (*ompU* and *ompW*) and luciferase gene (*luxA*) were determined by signal intensities of probes on dot-blot hybridization membranes (Table 3.6).

Table 3.6. Genotypic characteristics of the *V. cholerae* isolated from Chesapeake Bay

| Gene probe  | Percent of isolates for each level of signal intensity <sup>a</sup> |      |      |      |      |
|-------------|---|------|------|------|------|
|             | 0   | 1    | 2    | 3    | 4    |
| <i>ctxA</i> | 100.0   | 0.0  | 0.0  | 0.0  | 0.0  |
| <i>zot</i>  | 100.0   | 0.0  | 0.0  | 0.0  | 0.0  |
| <i>tcpA</i> | 100.0   | 0.0  | 0.0  | 0.0  | 0.0  |
| <i>toxR</i> | 1.3   | 7.1  | 4.5  | 2.7  | 84.4 |
| <i>hlyA</i> | 0.9   | 1.3  | 1.3  | 12.5 | 83.9 |
| <i>stn</i>  | 79.5  | 0.0  | 0.0  | 3.1  | 20.5 |
| <i>ompU</i> | 0.0   | 0.9  | 12.5 | 20.5 | 66.1 |
| <i>ompW</i> | 0.0   | 5.8  | 5.4  | 10.7 | 78.1 |
| <i>luxA</i> | 42.0  | 58.0 | 0.0  | 0.0  | 0.0  |

<sup>a</sup> Signal intensity was semi-quantitatively determined by darkness of the signal shown on x-ray film after hybridization with gene probe, to 5 levels: 0 = no signal (negative), 1 = low intensity, 2 = medium-low intensity, 3 = medium-high intensity, 4 = high intensity.

The major toxin genes associated with cholera pandemics, i.e., *ctxA* and *zot*, which are known to be part of the cholera phage CTXΦ, as well as the toxin-coregulated pilus gene *tcpA*, were not found in any of the 224 isolates. In contrast, the top regulator of cholera pathogenesis, *toxR* gene, was found in nearly every isolate (99%), with various levels of signal intensity, indicating different levels of gene homology with the probe *toxR* sequence.

When the hybridization signal intensity was considered, the strong signal intensity (level 4) indicating high homology to the probe *toxR* was observed in 84% of isolates. The rest of the isolates produced signals at intensity levels of medium-high (level 3), medium-low (level 2) and low (level 1) constituting 3, 5, and 7% of the total population, respectively. Surprisingly, three isolates did not show any signal to the *toxR* probe, indicating that these isolates did not have a gene homologous to the *toxR* probe used in this study. The three isolates (RC585-587) were isolated from the same water fraction at the S site on August 30, 1999 (S9830W) by direct plating. Clonal relationship of these three isolates was suggested by identical phenotypic and genotypic traits. These isolates also possess many atypical traits for *V. cholerae* i.e., gave negative results for sucrose, VP, amylase, and growth at 42°C. Although the first three traits suggest the isolates to be more similar to *Vibrio mimicus*, they were also luminous, a trait never reported before for *V. mimicus*. In addition, it should be noted that the three *toxR*-negative isolates possessed *ompU*, indicated by medium signal intensities for the *ompU* probe. Since *ompU* is known to be directly under the control of ToxR, how the *ompU* in these *toxR*-negative isolates is regulated is a very interesting issue that will benefit from further investigation.



Considering the consensus on the role of *toxR* as a global regulator that coordinates expression of many genes, including virulence and other genes, in response to environmental cues, and the universal distribution of *toxR* homologues in the genera *Vibrio* and *Photobacterium*, the finding of low- and no- *toxR* homology strains should be noted; and it suggested that *toxR* is not an essential gene for survival of *V. cholerae* in the Chesapeake Bay environment.

The hemolysin gene, *hlyA*, was found in 99% of isolates. The only two isolates that lacked the gene were from completely separate samples; RC341 was from Horn Point, on May 28, 1998, and RC544 was from the Susquehanna River on August 10, 1999. The only common feature of their source habitat was that both were isolated from water. It should be noted that isolate RC341 also showed a low *toxR* signal and was negative for VP, amylase, and lipase. However, the other *hlyA* negative isolate, RC544 had traits those common to the majority of *V. cholerae*.

The hemolysin gene, *hlyA* showed variability in its sequence similarity to the gene probe. Most of the isolates gave strong signals (intensity level 3-4, 97%). Among the six isolates with a low signal (intensity level 1-2), three were also *toxR* negative (RC585-587). Another three isolates (RC561-563) were VP negative. Interestingly, RC561-563 were from the same small plankton sample and were isolated by direct plating from H site on July 8, 1999 (H9708P), with clonal relationships indicated by their possessing identical traits.

In contrast to *hlyA*, the non-O1 toxin gene, *stn* was present at a lower frequency (21%) with less signal variation. All 46 positive isolates gave a strong signal with the probe. Notably, 91% of the isolates in this group were also luminous.

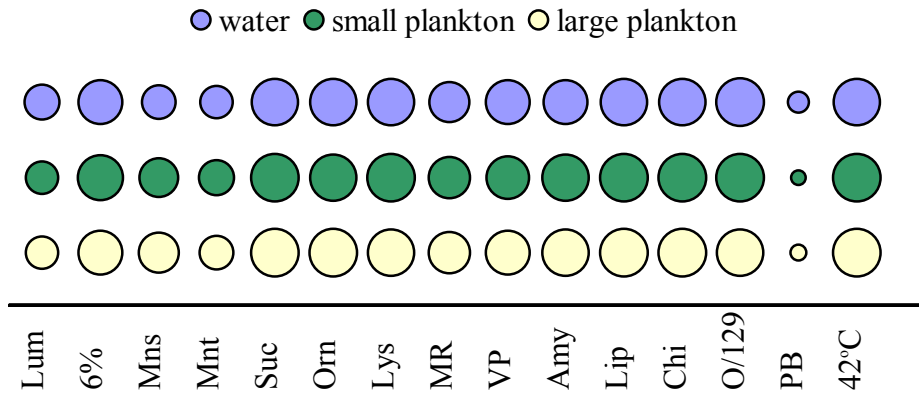
The two porin genes, *ompU* and *ompW*, were present in all isolates with a variable sequence homology with the gene probe. By grading the signal intensity of the two genes to 0-4 levels, a relatively higher level of gene fixation (i.e., less variability) was found with *ompW*. Although the prevalence of *ompW* high signals was greater than for *ompU*, the overall distribution of the alleles of the two genes across all intensity levels was not significantly different ( $P=0.08$ , marginal homogeneity test). Correlation in their distribution was also significant for the two genes ( $r=0.37$ ,  $P<0.01$ , Spearman rank correlation).

Similar to luminescence (Lum) among the phenotypic tests, the distribution of luciferase gene (*luxA*) was the most variable gene among those isolates included in this study. The gene was found in 58% (130/224) of the isolates. The hypothesis of predominance of *luxA* by a 6:4 ratio was accepted ( $P=0.59$ ). Variability within the *luxA* DNA sequence was very limited. All of the 130 positive isolates produced a strong hybridization signal, indicating a high level of homology in DNA sequence.

### **C Comparison of isolates from different sample fractions**

Average responses for phenotypic and genotypic traits of the isolates from different sample fractions were compared. Overall, all three fractions i.e., water, small plankton, and large plankton, yielded isolates with similar phenotypic and genotypic traits, indicating that they came from the same population (Fig. 3.5).

A: Phenotypic traits



B: Genotypic traits

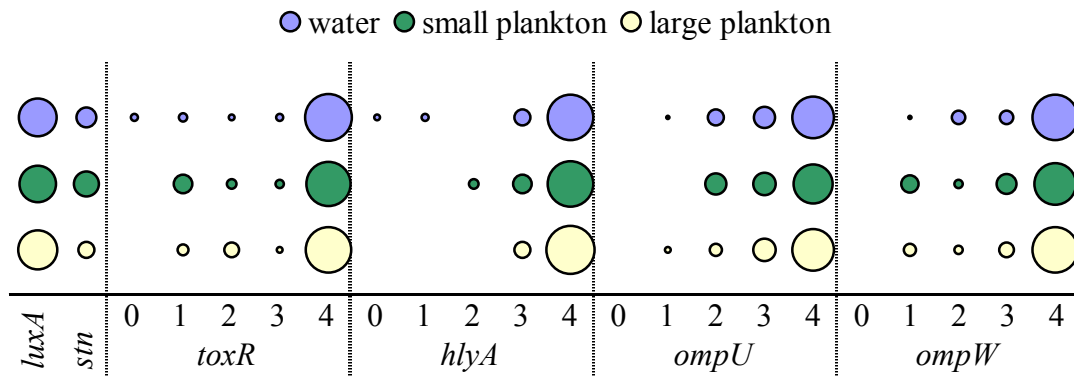


Figure 3.5. Bubble graph showing characteristics of *V. cholerae* from different samples. Size of the bubbles represent percent positive for the trait among the isolates in each sample. Panel A corresponds to phenotypic traits, and panel B corresponds to genotypic traits.

## **D Relationships among traits and environmental factors**

Correlation among pairs of each phenotypic and genotypic trait, as well as the environmental variables (pH, temperature and salinity), across isolates was estimated using the Spearman rank correlation. There were 5 pairs of traits that showed strong correlation  $r > 0.5$  (Table 3.8). The highest correlation was between Lum and *luxA* which is the phenotypic expression and its corresponding gene. *ompW*, *toxR*, and VP tests were highly correlated to each other. Marginal homogeneity test accepted the hypothesis of *toxR* and *ompW* hybridization signal showing the same kind of response i.e., the cell producing high signal to *toxR* also produced high signal to *ompW* probes and vice versa. The genes were highly correlated with VP test since almost all (28/29) isolates with low *toxR* were negative for VP. Sucrose fermentation and growth at 42°C were highly related. Out of 224 strains, five strains could not grow at 42°C, and three of those could not ferment sucrose.

Table 3.7. Correlation coefficient among traits<sup>a</sup>

|      | pH    | Temp  | Sal   | luxA        | toxR        | hlyA  | omU         | omW   | stn   | Lum   | Na6   | Mns   | Mnt  | Suc         | Orn   | Lys  | MR    | VP    | Amy  | Lip  | Chi  | V150 | V10  | PB   | 42C  |  |
|------|-------|-------|-------|-------------|-------------|-------|-------------|-------|-------|-------|-------|-------|------|-------------|-------|------|-------|-------|------|------|------|------|------|------|------|--|
| pH   | 1.00  |       |       |             |             |       |             |       |       |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| Temp | -0.31 | 1.00  |       |             |             |       |             |       |       |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| Sal  |       | 0.28  | 1.00  |             |             |       |             |       |       |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| luxA |       | 0.18  |       | 1.00        |             |       |             |       |       |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| toxR |       |       |       | 0.31        | 1.00        |       |             |       |       |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| hlyA |       |       |       | -0.33       | 0.29        | 1.00  |             |       |       |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| omU  |       |       |       | 0.43        | 0.30        | 0.16  | 1.00        |       |       |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| omW  | -0.18 |       |       | 0.33        | <b>0.68</b> | 0.26  | 0.37        | 1.00  |       |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| stn  |       | 0.36  | 0.17  | 0.34        | 0.14        | -0.18 | 0.16        | 0.20  | 1.00  |       |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| Lum  |       | 0.19  |       | <b>0.89</b> | 0.26        | 0.41  | 0.31        | 0.40  |       | 1.00  |       |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| Na6% | -0.16 |       |       | 0.29        | 0.38        | 0.22  | 0.23        | 0.38  |       | 0.28  | 1.00  |       |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| Mns  | 0.20  |       |       | 0.17        | -0.31       | 0.16  | -0.19       | 0.17  | 0.28  |       |       | 1.00  |      |             |       |      |       |       |      |      |      |      |      |      |      |  |
| Mnt  | -0.18 | -0.20 | -0.31 | -0.47       | -0.33       |       | -0.21       | -0.26 | -0.43 | -0.34 | -0.16 | 0.21  | 1.00 |             |       |      |       |       |      |      |      |      |      |      |      |  |
| Suc  |       |       |       | -0.19       | 0.32        | 0.31  | 0.15        | 0.23  |       |       |       |       |      | 1.00        |       |      |       |       |      |      |      |      |      |      |      |  |
| Orn  |       |       |       |             |             |       |             |       |       |       | 0.27  |       |      |             | 1.00  |      |       |       |      |      |      |      |      |      |      |  |
| Lys  |       |       |       |             |             |       |             |       |       |       |       |       |      |             |       | 1.00 |       |       |      |      |      |      |      |      |      |  |
| MR   |       | 0.21  | 0.16  | -0.17       | -0.22       |       | -0.19       |       | 0.23  | -0.13 | 0.21  | -0.23 |      |             |       |      | 1.00  |       |      |      |      |      |      |      |      |  |
| VP   |       |       | 0.38  | <b>0.83</b> | 0.29        | 0.34  | <b>0.70</b> | 0.18  | 0.32  | 0.39  | -0.27 | -0.35 | 0.29 |             |       |      | -0.24 | 1.00  |      |      |      |      |      |      |      |  |
| Amy  |       |       | 0.20  | 0.38        | 0.27        |       | 0.36        |       | 0.16  | 0.15  |       | -0.17 | 0.38 |             |       |      | -0.14 | 0.47  | 1.00 |      |      |      |      |      |      |  |
| Lip  |       |       |       | 0.23        |             |       | 0.20        |       |       |       | 0.27  |       |      |             |       |      |       | 0.23  | 0.31 | 1.00 |      |      |      |      |      |  |
| Chi  |       |       |       | 0.17        | 0.15        |       | 0.13        |       |       | 0.19  |       |       |      |             |       |      |       | 0.16  | 0.22 |      | 1.00 |      |      |      |      |  |
| V150 |       |       |       | 0.17        | 0.15        |       | 0.13        |       |       | 0.19  |       |       |      |             |       |      |       | 0.16  |      |      |      | 1.00 |      |      |      |  |
| V10  |       |       |       | 0.17        | 0.15        |       | 0.13        |       |       | 0.19  |       |       |      |             |       |      |       | 0.16  |      |      |      | 1.00 | 1.00 |      |      |  |
| PB   |       |       | -0.36 | -0.22       |             | -0.30 | -0.29       |       | -0.30 | -0.37 |       | 0.27  |      | -0.22       |       |      |       | -0.25 |      |      |      | 1.00 | 1.00 | 1.00 |      |  |
| 42C  |       |       |       | 0.31        | 0.29        |       | 0.22        |       |       |       |       |       |      | <b>0.77</b> | -0.22 |      |       | 0.37  | 0.39 |      | 0.44 |      |      | 1.00 | 1.00 |  |

<sup>a</sup> The values are calculated by Spearman rank correlation coefficient. Only significant correlations ( $P < 0.05$ ) are shown, blank space = non-significant. High correlation ( $> 0.5$ ) values are shown in bold inside a box.

Fig. 3.6 to 3.8 show plots of number of samples yielding interesting isolates, i.e., luminous, *stn* positive, and low *toxR*, plotted against environmental factors i.e., pH, salinity, temperature, and time of isolation.

For luminescence, there was no clear association of the trait with specific environmental factors (Fig. 3.6). In general, higher number of sample with luminescent isolates was found when higher number of sample was positive for *V. cholerae* isolation. In the year 2000, there was a higher ratio of luminous samples, however this could be chance, since the number of positive sample is small. Positive samples were obtained from pH range 7.3-9.1, with more samples positive at slightly alkaline pH 7.9-8.0. In salinity plot, the positive samples was isolated from salinity 0-21 ppt, however, bimodal distribution of positive samples was observed, at a peak of 5-7, and another peak at 9-13 ppt. More positive samples were observed at warm temperature 24-30°C and summer months of all the three calendar years.

There was no clear association of *stn* with the environmental factors tested (Fig. 3.7). Positive samples were found in a wide range of pH, from 7.3-9.1, with higher positive at slightly alkaline pH; salinity from 0-12 ppt, and, similar to luminescent samples, bimodal distribution of *stn* positive samples in the salinity plot was observed. More positive samples were observed at warm temperature 24-30°C and summer months of all the three calendar years.

The isolates containing a gene with a low sequence similarity to *toxR* are interesting, since it is a rare trait and suggests a uniquely different population of *V. cholerae* in Chesapeake Bay. Similar to luminescence and *stn*, the isolates with a low

reactive *toxR* (determined by gene probe signal intensity level 0-2) showed no clear association with pH, salinity, temperature, or time of isolation (Fig. 3.8).

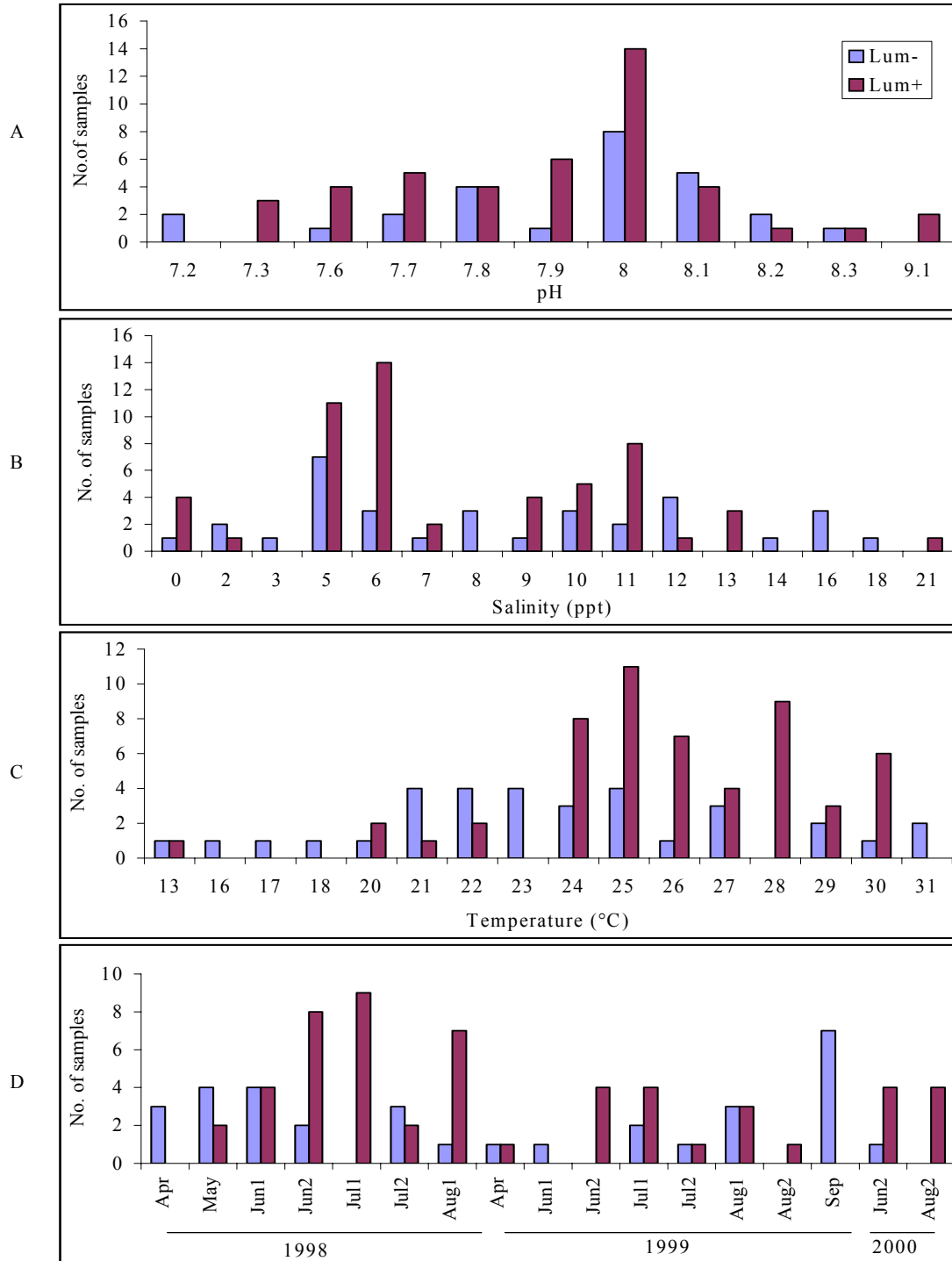


Figure 3.6. Column graphs plotted between number of samples yielding luminous (lum+) and non-luminous (lum-) isolates and environmental parameters, at different pH (A), salinity (B), temperature (C), and time of isolation (D).



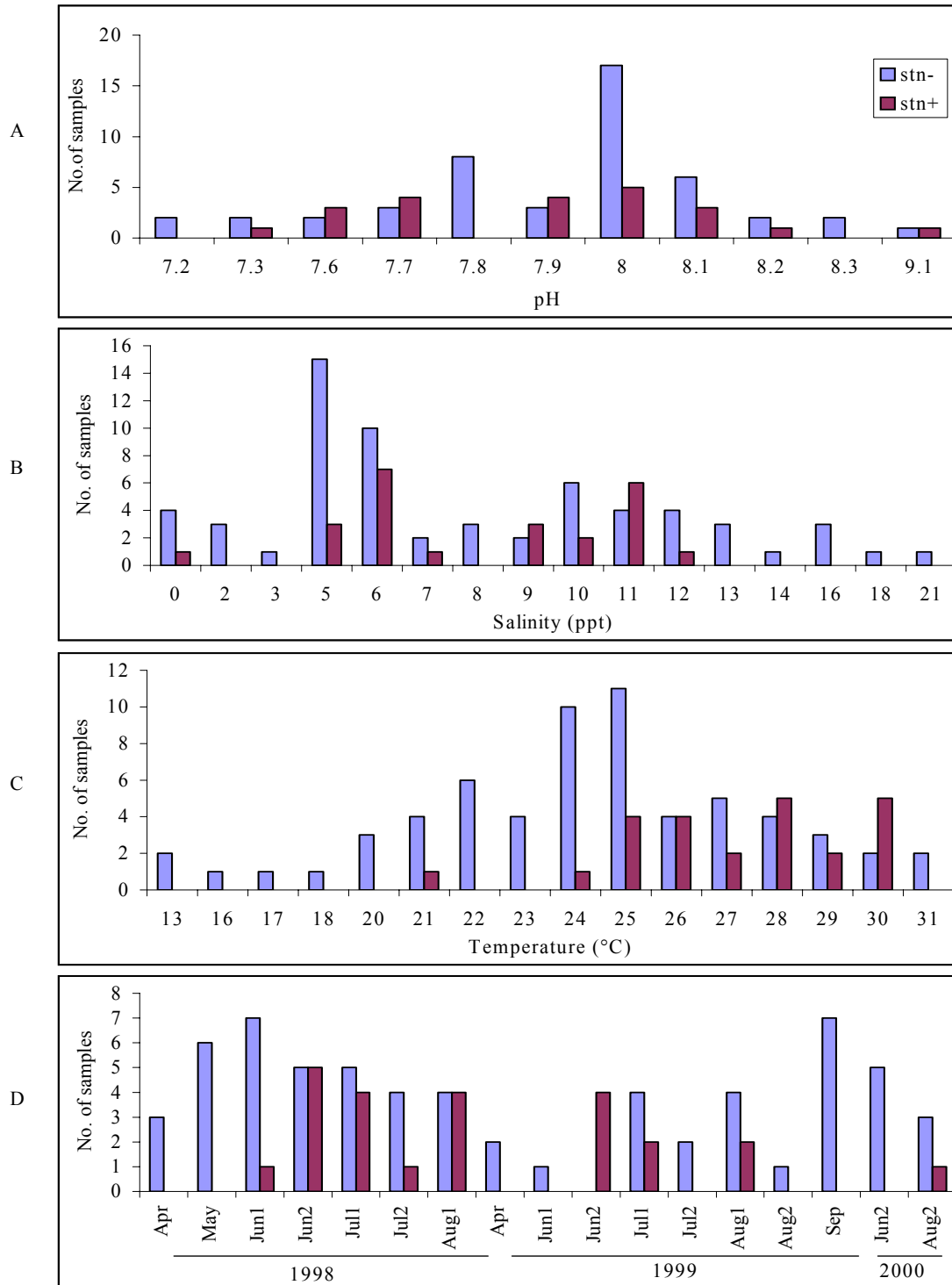


Figure 3.7. Column graphs plotted between number of samples yielding *stn* positive (*stn*+) and *stn* negative (*stn*-) isolates and environmental parameters, at different pH (A), salinity (B), temperature (C), and time of isolation (D).

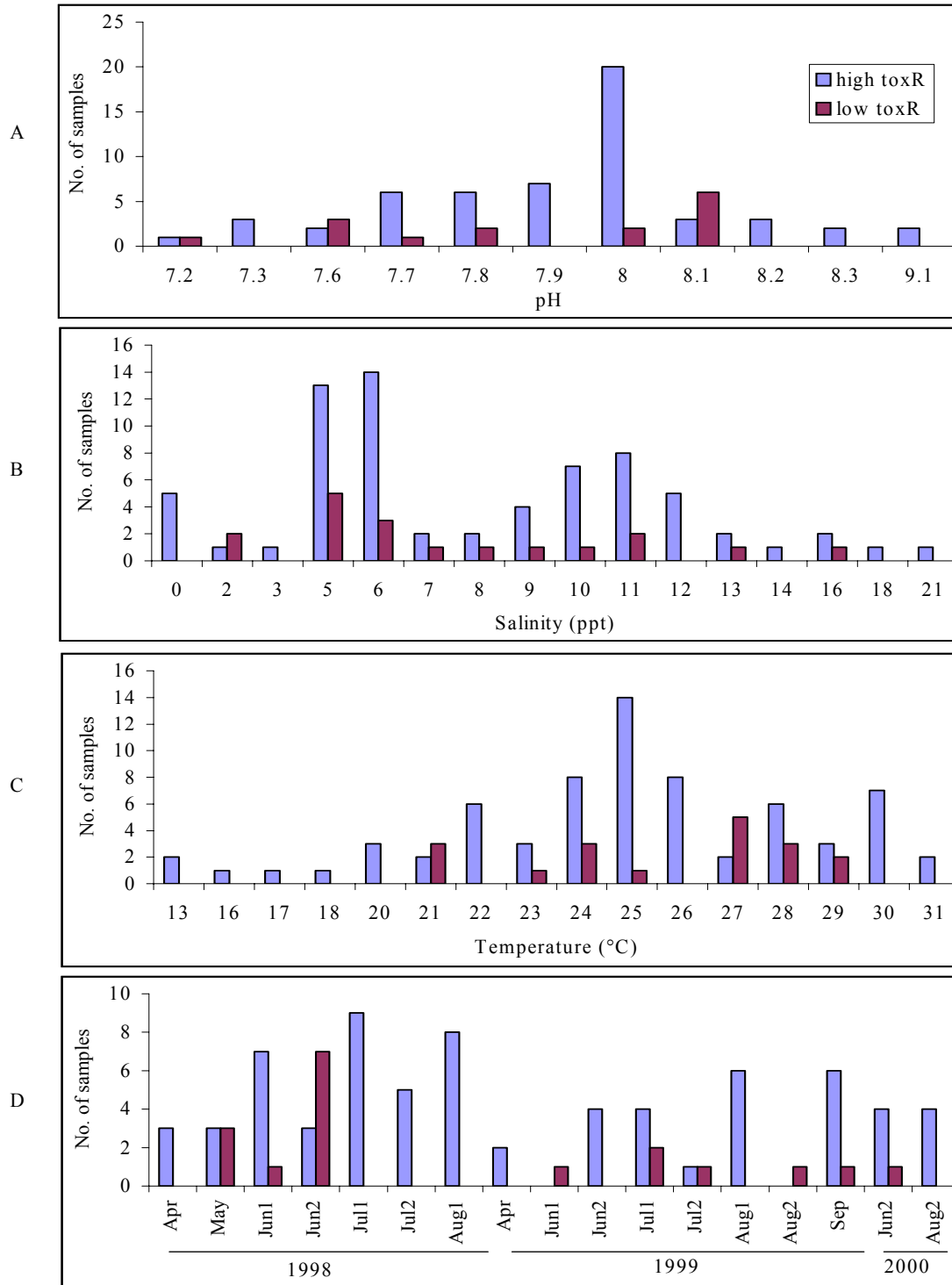


Figure 3.8. Column graphs plotted between number of samples yielding low *toxR* isolates (signal level 0-2) and high *toxR* (level 3-4) isolates and environmental parameters, at different pH (A), salinity (B), temperature (C), and time of isolation (D).

## 3.4 Discussion

### 3.4.1 Seasonality and annual variability

In this study, a clear seasonality of *V. cholerae* in Chesapeake Bay was observed, in that more isolates were obtained in the summer and significantly fewer in the winter. This finding is in agreement with that of previous studies (80, 192).

The multi-year span of this study revealed variation between years in abundance of *V. cholerae* in Chesapeake Bay. The number of isolates was more than five fold higher in 1998 than in 1999 when the enrichment method was used. The average salinity in 1998 was significantly lower than in 1999, while pH and temperature patterns were similar. Therefore, the observed annual variation was most likely related to the salinity differences between these two years. Louis *et al.* (111) reported that salinity differences between these two years were a result of the difference in amount of fresh water influx from the Susquehanna river, the major tributary located in the northernmost Chesapeake Bay. In 1998, there was higher than usual river flow, while in late 1998 and during most of 1999, there was unusually low flow, resulting in higher salinity in 1999. However, it should be noted that besides affecting salinity, there are other factors associated with high fresh water flow that could affect abundance of *V. cholerae*, such as nutrient runoff and water mixing, providing favorable conditions for multiplication of the bacterium.

Although *V. cholerae* was isolated only in the spring and summer (August-September), its presence was detected by PCR throughout the year. Exceptions were during the coldest months of December, 1998 in which the water at some sampling sites

was frozen (0-4°C) and the following month, January, 1999 (5-8°C). A similar finding was reported by Heidelberg *et al.* (80) in that detection of *V. cholerae* - *V. mimicus* populations in Chesapeake Bay by FODC was negative in the colder months. Therefore, failure to obtain isolates in the fall-winter months, where the DNA can be detected by PCR may be explained by a lower abundance of *V. cholerae*, i.e., lower than the detection limit of the isolation method or that the cells may exist in a viable but non-culturable (VBNC) state, hence failing to grow on the culture medium used.

During the coldest months of the winter, *V. cholerae* essentially disappear from the surface water, as indicated by PCR. It is important to find out whether the bacterium actually dies off or, through its association with planktons, over-winters in the sediment, as has been reported (79). Kaper *et al.*(96) found *V. cholerae* only in sediment during the winter months and zooplankton are known to undergo a form of dormancy in sediment of estuaries. Thus, the latter hypothesis appears to be more probable.

### **3.4.2 Comparison of enrichment and direct plating methods**

#### **A Uncoupling of the isolate**

It is well known that the enrichment method is very sensitive and able to detect very low numbers of target bacteria in a sample. Investigators have reported a superior sensitivity for the enrichment-based method, compared to the direct plating method (48, 71). In this study, the 14 samples that were positive only by enrichment and not by direct plating gave good agreement with this finding.

Surprisingly, eight samples were positive only by direct plating but not after enrichment, leading to uncoupling of the results (Fig. 3.4). Considering the lower

sensitivity of direct plating, it implies that the number of *V. cholerae* in these samples was large. Failure to detect a large number of target bacteria by enrichment was unexpected. Several investigators who used the MPN (Most probable number) method, a standard method for bacterial quantification based on growth in an enrichment medium, have reported failure in obtaining a positive result in some of the tubes with the largest amounts of inocula, especially when trying to enumerate *Vibrio* spp. (154). This phenomenon was explained as either a limitation of the MPN method or competitive overgrowth of other bacteria in the enrichment medium, or both. In contrast to the MPN, in which a relatively large inoculum size is used, it means more of other non-target bacteria are competitors. The inoculum size used in this study was fixed. However, the possibility of a high concentration of factors negatively affecting growth of the target bacterium, such as competition with other bacteria in these samples, should not be excluded.

Alternatively, the fact that most of the isolates from direct plating were from plankton provides a clue to the answer. It is entirely possible that the bacterium is strongly associated with plankton either internally or tightly bound on the surface of the zooplankton, in a biofilm. This zooplankton association and the biofilm formation of *V. cholerae* are well documented (76, 79, 86). *V. cholerae* either associated with zooplankton or in biofilms provide the bacterium protection from adverse environmental conditions (76, 79, 86, 87). A homogenizing step introduced into the direct plating method would allow a mechanical breaking down of this association, hence releasing the bacterium to grow individually on plates. Unfortunately, the plankton samples inoculated into enrichment cultures were not homogenized. Therefore, although the bacterium was

able to multiply, it was still associated with zooplankton and/or particulates or trapped inside a biofilm, hence undetectable by enrichment since only top layer of the static medium (without mixing) was used.

In sample S9830, the only water sample that yielded isolates by direct plating, the isolates were atypical, sucrose negative strains. It was not surprising that they were missed by enrichment since only sucrose positive colonies were selected.

## **B Limitations and advantages of direct plating**

Detection limit of the direct plating protocol used in this study varied with the number of total culturable bacteria in the samples. Due to the non-selective nature of the plating medium, to obtain discrete colonies, the samples sometimes had to be diluted up to 100-1000 fold. In that case, the detection limit of the method was as high as  $10^3$ - $10^4$  cells/ml.

There were some other difficulties associated with interpretation of the direct plating results. Colonies with medium to low signal intensities were often observed. Although most of these were not *V. cholerae*; it was not always the case. This may be due to uneven or poorly bound *V. cholerae* DNA on the filter paper. Colony size differences, such as very small colonies can be regarded as artifacts or vice versa, contributing to error in enumeration. Difficulty in discriminating positive colonies from artifacts does not provide confidence when this method is used as a quantification tool. Moreover, on several occasions, there was interference by swarming bacteria, rendering the results uninterpretable.

Clearly, additional optimization of the direct plating method is needed. To increase sensitivity of detection, a more selective media is needed. Although this may

result in underestimation of diversity and abundance, it would be more powerful than when a non-selective medium is employed. Higher percent agar in the plating medium could help reduce the problem of swarming bacteria.

Nonetheless, direct plating has advantages in that it provides information about the abundance of clones in the water, although only when the bacterium is present in large numbers. Furthermore, strains with an atypical phenotype, such as sucrose negative strains can be detected by this method.

### **C Limitations and advantages of enrichment**

In comparison, enrichment was more sensitive and yielded more diverse clones than direct plating. However, it is inferior to direct plating in that it does not reveal the true abundance of each specific clone. It is not possible to differentiate artificial clones that amplify in the enrichment from true abundance of the same clone in the environment. Furthermore, enrichment did not yield atypical strains since only sucrose-fermenting, *V. cholerae*-like colonies were selected. Additionally, *V. cholerae* tightly associated with plankton may be missed.

### **D An improved protocol for isolating *V. cholerae***

A combination of both methods could provide a better alternative. The protocol can be improved by enriching different dilutions of the homogenized sample, followed by plating on a selective medium, and detecting *V. cholerae* colonies by colony hybridization. This alternative method would be more sensitive, with diverse clones detected (a benefit of the enrichment). True abundance of the clones could then be estimated by dilutions of the inoculum, and atypical isolates detected (benefit of the

colony hybridization). Practicality and effectiveness of such a protocol in actual field application would be realized by further study.

### **3.4.3 Distribution of traits and their function**

#### **A Macromolecule degradation**

From the phenotypic and genotypic traits observed for the *V. cholerae* isolated from Chesapeake Bay, it is concluded that the species *V. cholerae* represents a diverse group of organisms with variability in both phenotype and genotype. However, most of the isolates (92-100%) were able to degrade macromolecules, i.e. were positive for gelatinase, amylase, lipase, and chitinase. Interestingly, chitin, the second most abundant polysaccharide in nature, usually exists as a complex with other polysaccharide macromolecules; it is a major component in the outer skeleton of crustaceans. These properties support the hypothesis that *V. cholerae* is ubiquitous in the aquatic environment and suggest that it has a significant role in nutrient cycles in the aquatic ecosystem.

#### **B Genes common to environmental *V. cholerae***

The presence of *toxR*, *hlyA*, *ompU*, and *ompW* genes in all (*ompU* and *ompW*) or most (*toxR* and *hlyA*) of the isolates suggests that these genes were acquired at an early state of the species evolution and that they play an important role in the persistence and geographic distribution of *V. cholerae* in the aquatic environment. Isolates that completely lacked a gene with sequence similarity to *toxR* are rare and usually have other atypical characteristics for the species. Isolates without *toxR*, as well as isolates with low



*toxR*, were confirmed as *V. cholerae* by other molecular methods, i.e., DNA-DNA hybridization and 16S-rDNA sequencing (Chapter 5).

It is noted that these four common genes have diverse sequences, as shown by different levels of hybridization signals. Three of the genes are known to be involved in sensing environmental signals (*toxR* and the two porin genes). Sequence variations in these genes may give *V. cholerae* advantages in adaptation and survival in a wider range of habitats. Interestingly, *hlyA*, considered to be a toxin gene, was also universally present, with some sequence variations similar to that of other environmental sensing genes. It also showed moderately high correlation with salinity. These characteristics suggest that the *hlyA* gene in *V. cholerae* may function in adaptation to the environment. Similar to *toxR*, rare isolates without the *hlyA* gene were found. The possible function of these genes in the association of *V. cholerae* in zooplankton would be fascinating to determine.

### **C Luminescence in *V. cholerae***

High incidence of luminescence in Chesapeake Bay isolates may be explained, in part, by a more sensitive detection method with superior sensitivity of the luminometer over the human eye. Grim *et al.* (73) showed the effect of culture medium, culture method, and length of time of cultivation on the light production by *V. cholerae*. Under optimized conditions, detection of the luminescence trait using the luminometer method was more sensitive than the method previously employed, i.e., observing colonies on agar plate after a period of dark adjustment. Approximately half of the strains originally characterized by West *et al.*, 1986 (190) were re-examined and the results showed a higher percentage of luminous strains, i.e., an increase from 10% to 17% (73).

Nonetheless, the 50% chance of finding isolates with luminescence was higher than any other reported. Occurrence of the luciferase gene, *luxA*, was even higher, as 60% of the isolates possessed the gene. Strains with *luxA*, but not luminous, may be explained by conditions provided in laboratory culture not being suitable for gene expression in some strains or that the gene may be defective, leading to dark mutants as commonly occurs in other luminous bacteria. This high incidence of luminescence occurred throughout the year, when *V. cholerae* was isolated, spreading in all three consecutive years when the sample was taken, from 1998 to 2000, and the high ratio of luminous strains was observed for all three sample types. Therefore, Chesapeake Bay may be mid-range in providing a specific environment favoring the luminous trait in *V. cholerae*, compared to the luminous bays of Puerto Rico.

High incidence of environmental isolates harboring the fully functional luminescence system infers a selective advantage in the environmental habitat. Otherwise, this highly energy consuming trait would not be conserved within the population. Czyz *et al.* (37) showed experimentally, using *V. harveyi* as the test organism, that lux(-) mutants become dominant over lux(+) strains under normal growth conditions but, in contrast, the lux(+) strain became dominant in cultures exposed to damage and stress by low level UV irradiation. It was proposed that expression of the luciferase gene helps facilitate DNA repair by providing an internal light source for the photo-mediated repair mechanism, photoreactivation (37, 38, 181). It can be speculated, therefore, that the free-living bacterium in water, which is more exposed to DNA damage conditions such as UV exposure, may use luminescence as a protective mechanism from DNA damage. When attached to plankton, the bacterium is even more protected from

UV damage; however, there may be different selective advantages for luminous isolates. The biological role of luminescence in bacteria is far from being fully known. However, it has been speculated that by making the substrate where the luminous bacteria attach luminous, the substrate, as well as the attached luminous bacteria would be more likely to be eaten by higher organisms, providing the bacterium with a nutrient rich environment in the gut, which would be a devious device for survival and multiplication.

#### **D Toxin genes in *V. cholerae***

The moderately high incidence of the *stn* gene (21%) and presence of *stn* positive isolates throughout the study from 1998 to 2000, suggests that this gene also plays a role in the natural environment. Although this gene has been linked to pathogenicity of non-O1 *V. cholerae*, enrichment of clones possessing the *stn* gene in the human host is hardly likely to contribute to the level of abundance and persistence observed in Chesapeake Bay where no outbreak of diarrheal case from non-O1 *V. cholerae* has ever been reported. As most of the *stn* positive strains are also luminous, it suggests that its role in exiting of the bacterium after a period of enrichment in the gut of its natural aquatic host attracted by plankton and/or particulates illuminated by the bacterium.

None of the environmental *V. cholerae* isolated in this study possessed genes coding for cholera toxin or for the cholera phage CTX $\Phi$  receptor and colonization factor, *tcpA*. None belong to the serogroups O1 and O139. This finding suggests that without enrichment in a human host, pathogenic strains of *V. cholerae* are rare in the environment or are very difficult to isolate. The reason may also be that O1 and O139 strains not common in the aquatic environment and may be out competed by environmental non O1 and non O139 strains. However, in part of the world where cholera is endemic, usually

in countries where sanitary conditions are a public health issue, discharge of the pathogenic clone into the drinking water reservoir is not rare. Furthermore, many of the cholera endemic areas have warm climates and mild winters, conditions hospitable to the bacterium, hence increased survival and persistence of the pathogenic clone in the environment. Nevertheless, isolation of O1 and O139 *V. cholerae* in those countries has been equally difficult. Only when direct probing with genes coding for O1 have these strains been detected in the environment with significantly high frequency.

### 3.5 Conclusions

The collection of 224 *V. cholerae* isolates in Chesapeake Bay was successfully established and permitted a study of population structure of *V. cholerae*. Both isolation methods used in this study were effective and complemented each other. A new isolation protocol, combining benefits of both methods is proposed; however the effectiveness of the new protocol will have to be evaluated. The relatively high incidence of *V. cholerae* and large number of isolates suggest that the species exists in micro-habitats both as free living and attached to particulates and plankton in the aquatic environment. The same population identified by similar phenotypic and genotypic traits moves freely from free-living to particulate-attached existence. As revealed by direct detection method by PCR, the species is truly ubiquitous in the natural aquatic environment, except in mid-winter in the temperate climate of Chesapeake Bay.

The seasonal dynamics of abundance of culturable isolates showed significantly contrast from a summer high (June to August) and a fall-winter decline and/or absence (September to March). Annual variation was also documented, most likely linked to fresh water and nutrient influx from the major Chesapeake Bay tributary, the Susquehanna River.

The environmental isolates of *V. cholerae* comprised diverse clones with a variety of phenotypic and genotypic traits. This diversity offers a large repertoire of resources to which the bacterium can adapt in a wide range of habitats and contributes to the success of the species. In fact, it thrives and persists in the natural environment of the Chesapeake Bay. The role of *V. cholerae* as a vigorous decomposer and agent of nutrient

cycling is suggested based on its universal ability to degrade a variety of macromolecules, including chitin, the most abundance macromolecule in the aquatic environment. The variability of base sequences of its porin genes also supports the versatility of *V. cholerae* in adapting to a wide range of habitats. On the other hand, the ubiquitous presence of hemolysin, even though, with sequence variability, and the surprisingly relatively high incidence of the *stn* gene in *V. cholerae* which are also frequently luminous, suggests that there is an important role for this gene in interactions of *V. cholerae* with aquatic eukaryotic organisms in the life cycle of *V. cholerae*, most probably zooplankton but not necessarily limited thereto. It was fascinating to find the 50:50 prevalence of luminescent *V. cholerae*, much higher than previously thought, a result, no doubt, due to a better testing protocol and more sensitive instrument. Nevertheless, luminescence should be recognized as one of the frequently occurring traits in environmental *V. cholerae*.

In conclusion, *V. cholerae* is a versatile aquatic inhabitant that can persist in either the free living state or attached to particulates under a variety of environmental conditions. It plays an important role in nutrient cycling and, perhaps, takes advantage of its aquatic host organisms such as plankton for growth and multiplication in the natural environment. Pathogenic clones were not detected in the Chesapeake Bay by the methods used in this study, possibly due to lack of enrichment through a human host, being out-competed by the more environmental-adapted clones, or the necessity to use direct gene probing of DNA extracted from water and plankton to assess presence of the genes associated with pathogenicity in humans.

## **Chapter 4 Determination of Clonality and Relatedness of *Vibrio cholerae* by Genomic Fingerprinting, using Long-Range Enterobacterial Repetitive Intergenic Consensus (ERIC) PCR**

### **4.1 Introduction**

#### **4.1.1 Importance of determining bacterial clonality**

Ability to identify clonal identity of bacterial isolates has important implications both in epidemiology and in general ecology surveys. Many outbreaks of disease have been associated with distinct bacterial clones. For example, hemorrhagic colitis and hemolytic uremic syndrome is caused by *Escherichia coli* O157:H7 (149), Brazilian purpuric fever by the specific clone of *Haemophilus influenzae* biogroup *aegyptius* (19), and toxic shock syndrome by *Staphylococcus aureus* ET4 (128). In addition to aiding identification of the causative agent and estimating the risk involved with specific pathogenic clones, knowledge of clonal identity can be used in epidemiology to track the source and spread of pathogenic bacteria. In general, in ecological surveys, such as was done in this study, identification of clones can be used to enhance efficiency and accuracy of the survey by making it possible to eliminate clonal redundancy in the strain collection. In the case of cholera, it is believed that all major outbreaks prior to the seventh pandemic were caused by *Vibrio cholerae* serogroup O1 classical biotype. Later, in the seventh pandemic, serogroup O1 El Tor and O139 emerged and became dominant. These shifts in epidemic clones emphasize importance of monitoring the bacterial

population as a whole to discern the presence of toxigenic clones as well as their close relatives.

#### **4.1.2 Available methods for determining bacterial clonality**

To determine clonal identity and the relatedness of the clones to the bacterial pathogen being sought in a large scale survey, such as in epidemiological studies, sensitive and robust methods are needed that are suitable for high throughput application. Several available methods for doing this are briefly reviewed below.

##### **A Phenotype-based method**

Early work on the clonal nature of bacterial pathogens was done largely by using specific phenotypic traits, such as serotyping based on bacterial surface antigens. Sakazaki type O1 and O139 of *V. cholerae* showed a strong correlation with cholera epidemic traits. However, since the test is based on only one characteristic i.e. expression of the O antigen, the analysis does not have enough resolution power for estimating the overall level of chromosomal diversity and relationships among isolates. The serotype trait also has been shown to be acquired by horizontal gene transfer from unrelated strains, rendering it unsuitable for this purpose. Moreover, the task of maintaining a collection of antisera, with more than 200 *V. cholerae* serogroups recognized to date (201), limits the use of this method to only a few reference laboratories.

To achieve enough resolution to differentiate among strains, a large number of phenotypic traits must be measured. Several biochemical test kits are commercially available, for example, Biolog microplate (Biolog, Inc., Hayward, CA), API identification system (bioMérieux, Inc., Hazelwood, MO) and Pheneplate system



(PhPlateAB, Stockholm, Sweden). Depending on the system selected, 23-95 tests can be done in a single multi-well plate. However, besides being costly and time and labor intensive, phenotypic methods have limited discriminatory power, especially in dealing with clonality of strains of the same species. It also has the inherent problem of reproducibility, dependence on culture conditions, and lack of stability during subculturing the strains being tested.

One method that is widely used in bacterial population genetics, multilocus enzyme electrophoresis (MLEE) (162), analyzes the electrophoresis profiles of several housekeeping enzymes and has proven to be more robust than other phenotype-base methods. Although it played a central role in establishing long-term bacterial epidemiology, MLEE is technically cumbersome and has not been adopted extensively for routine surveillance. Although the resolution power of MLEE is high, depending on the number of enzymes tested, time and cost associated with the method limits its use to small sets of strains and are prohibitive for the large scale surveys needed in ecological studies.

## **B Molecular methods**

Currently, many molecular approaches have been proposed and are increasingly dominant because they reflect relationships encoded in the DNA. They are less time consuming than extensive phenotypic analyses and provide better, more reproducible results. Ribotyping, first described by Grimont (74), detects polymorphism of rRNA gene loci by restriction digestion of the genomic DNA fragments separated by gel electrophoresis and hybridized with labeled ribosomal DNA probe. Pulsed-field gel electrophoresis (PFGE) uses a rare-cutting restriction enzyme to digest chromosomal

DNA and involves electrophoresing the large DNA fragments in a pulsed-field gel electrophoresis apparatus. Both methods are widely used in epidemiological studies of *V. cholerae* (22, 98, 104, 140). These methods, however, involve many steps and require specialized equipment which may limit their use to only the more resource-rich laboratories. Multilocus sequence typing (MLST) (115, 175) is an extension of MLEE, using DNA sequences of several selected housekeeping loci on the bacterial genome to identify and estimate genomic relationship among clones. This method is very powerful and highly reproducible, becoming more popular with advances in sequencing technology that make high throughput sequencing possible at a reasonable cost. However, pre-knowledge of the discriminatory enzymes that will provide useful information to differentiate clones for bacterial in each species is needed.

### **C Polymerase Chain Reaction (PCR)-based methods**

PCR-based genomic fingerprinting methods have many advantages over other molecular methods. The PCR method offers simplicity, high throughput ability with less demand for sophisticated instrumentation. With proper primers, the whole genome can be sampled randomly and evenly, generating robust fingerprint patterns without bias to any specific area. In MLEE, MLST, and ribotyping, only a selection of housekeeping genes is investigated, which might not reflect the whole genomic relationship. Several studies have shown that the discriminative capability of PCR-based fingerprint methods closely approaches or exceeds those of ribotyping and PFGE (118, 167, 197).

Several PCR-based methods, such as amplified fragment length polymorphism (AFLP), arbitrary primed PCR, and several repetitive sequences-based PCR (Rep-PCR) methods have been used to generate genomic fingerprints of *V. cholerae*. Jiang *et al.*

(2000) used a modified protocol for AFLP and successfully discriminated among closely related clones of *V. cholerae* O1 and O139 from different outbreak origins (93). AFLP is a high resolution and highly reproducible method (94). The procedure, however involves many steps and several days to complete. The steps include restriction enzyme digestion of the genomic DNA, addition of adapter sequences by a ligation reaction, DNA clean up by ethanol precipitation, labeling of primers, PCR amplification, polyacrylamide gel electrophoresis, and signal detection.

In contrast, arbitrary primed PCR (AP-PCR) or random-amplified fragment length polymorphism (RAPD) involves using a single arbitrary primer to amplify several discrete DNA products. This method is simple and has good discriminatory power, but poor reproducibility as shown by a test conducted by a network of several European laboratories (94). Taq polymerase and the thermocycling machines have been shown to play a role in influencing the RAPD patterns (114, 124). Therefore, it is not suitable for pooling data between laboratories or even within the same laboratory using different equipment and reagents.

Rep-PCR fingerprinting makes use of DNA primers complementary to naturally occurring, highly conserved, repetitive DNA sequences present in the bacterial genome. Several distinct families of the repetitive sequences have been found in most bacterial genomes, the most widely used are the 35-40 bp repetitive extragenic palindromic (REP) sequence and the 124-127 bp enterobacterial repetitive intergenic consensus (ERIC) sequence. These repetitive sequences have proven to be useful in providing unambiguous and highly discriminative DNA fingerprinting profiles for most bacterial genomes (184).

PCR fingerprinting, employing ERIC sequences (ERIC PCR), has been widely used and has proved to be highly discriminative, comparable to or exceeding that of MLEE, ribotyping, and PFGE (47, 167, 197). The ERIC PCR fingerprint method has been used to differentiate toxigenic and nontoxigenic strains of *V. cholerae* O1/O139 and non O1 strains (150) and to create genomic profiles for population genetic and clonal relationship analysis among clinical and environment isolates of *V. cholerae* (32, 209).

#### **4.1.3 A new method based on ERIC PCR fingerprinting**

Although ERIC PCR fingerprinting has been successfully used to resolve *V. cholerae* clonal lineages, the protocol can be improved to enhance its discriminative power and reproducibility. In the previous study of Rivera *et al.*(1995) and Colombo *et al.*(1997), ERIC PCR fingerprints yielded only 1-8 amplicons in the size range of 0.1-4 kb (32, 150). Although it can differentiate among O1 and non O1 *V. cholerae*, it failed to differentiate among the more closely related strains O1 and O139. In a later study, Zo *et al.* (2002) used a modified protocol with lower amplification stringency and higher percent agarose gel (3.6% Metaphore agarose), obtaining highly complex patterns with more than 40 amplicons from each isolate (209). However, only the amplicons in the size range 100-588 bp were able to be resolved and used in the analysis. Moreover, reproducibility of low-stringency, highly complex fingerprint patterns may be questionable.

Long-range ERIC PCR takes advantage of the commercially available, improved Taq polymerase quality and buffer system. The higher fidelity of the polymerase provides reliable and longer amplification fragments. The long-range PCR proved useful and indispensable in many areas of research including bacterial molecular typing (4, 20, 39,

121). The larger amplicon size means more genomic regions are sampled, hence a better reflection of the total genomic polymorphism of different strains. In this study, the objective was to improve the resolution of ERIC PCR fingerprinting, using the long-range PCR system. Other factors that may affect resolution of the fingerprint pattern, such as in the gel electrophoresis step, was also examined.

#### **4.1.4 Objectives of this study**

The main objective of this study was to develop a rapid, high throughput, high resolution long-range ERIC PCR protocol for determining clonal identity of *V. cholerae* environmental isolates. Suitable PCR and gel electrophoresis conditions were defined. Reproducibility and discriminative power of the method was assessed and suitable protocol for computer-assisted interpretation of fingerprint patterns is presented. Accuracy in genome relatedness estimation obtained from ERIC PCR fingerprinting was verified by comparing genomic DNA reassociation using the membrane DNA-DNA hybridization method (DDH).

## 4.2 Materials and Methods

### 4.2.1 Bacterial strains

A total of 380 bacterial isolates were subjected to long-range ERIC PCR fingerprinting. Among these, 278 were *V. cholerae* Chesapeake Bay isolates obtained earlier in the study. The rest were toxigenic and nontoxigenic *V. cholerae* from clinical and environmental isolates from several other geographical locations. *V. mimicus* and other *Vibrio spp.* and non vibrio bacteria were included as outgroup controls at different levels of genome relatedness (Table 4.1). Reference strain RC145 (*V. cholerae* O1 El Tor, ATCC39315, N16961) was included in all PCR reactions and gel electrophoresis to establish the reproducibility level of the method.

### 4.2.2 DNA preparation

Genomic DNA from the bacterial isolates was extracted using the DNeasy<sup>®</sup> Tissue Kit (Qiagen Inc., Valencia, CA) and eluted in 200 µl elution buffer AE (10 mM Tris-HCl pH 9.0, 0.5 mM EDTA). DNA concentrations were determined spectrophotometrically by absorbance at 260 nm ( $A_{260}$ ) (158) and/or by the PicoGreen<sup>®</sup> dsDNA quantitation kit (Molecular Probes, Inc., Eugene, OR) using CytoFluor<sup>®</sup> multi-well plate reader series 4000 (PerSeptive Biosystems, Inc., Framingham, MA) at excitation 480 nm and detection at 520 nm. The DNA were diluted to 20 ng/µl in AE buffer and stored at -20 °C until analyzed.

Table 4.1. (next page) Bacterial strains tested by long-range ERIC-PCR fingerprinting.

<sup>1</sup> Species: received as: VC = *Vibrio cholerae*; nonVC = other bacterial species.

<sup>2</sup> T/N: Toxigenicity of the isolates (presence of the CTX element); T = toxigenic; N = nontoxigenic.

<sup>3</sup> E/C: source of sample; E = environmental sample; C = clinical sample.

Table 4.1. Bacterial strains tested by long-range ERIC-PCR fingerprinting

| Species <sup>1</sup> | T/N <sup>2</sup>            | E/C <sup>3</sup> | Source of isolate   | Strain name   | Number               |   |
|----------------------|-----------------------------|------------------|---------------------|---|----------------------|---|
| VC                   |                             | N E              | CB, USA (1998-2000) | RC325-609 (with exception of RC442, 533-536, 540-541) | 278                  |   |
| VC                   |                             | N E              | CB, USA (1976-1979) | UM928,930,942-944,947,972,980,987-989,993,1001,1358   | 14                   |   |
| VC                   |                             | N E              | Peru                | P1, 5, 9, 14, 23, 30, 37, 39, 78                      | 9                    |   |
| VC                   |                             | N E              | Bangladesh          | ZB 120, 124, 125, 128, 129, 132-134, 136              | 9                    |   |
| VC                   |                             | N E              | Italy               | RC285-288   | 4                    |   |
| VC                   |                             | N E              | Fl, USA             | UM2870  | 1                    |   |
| VC                   |                             | N E              | Brazil              | RC66  | 1                    |   |
| VC                   |                             | N E              | Argentina           | UM4645-4646   | 2                    |   |
| VC                   |                             | N E              | Czechoslovakia      | UM1355, 1359  | 2                    |   |
| VC                   |                             | N E              | Denmark             | DK236, 239  | 2                    |   |
| VC                   |                             | N E              | Thailand            | UM1353, 1354  | 2                    |   |
| VC                   |                             | N E              | Phillippines        | UM1357  | 1                    |   |
| VC                   |                             | N E              | Ballast water       | RC785   | 1                    |   |
| VC                   | O1                          | El Tor           | N E                 | Chile   | RC23                 | 1 |
| VC                   | O1                          | El Tor, Ogawa    | N E                 | Brazil  | RC28                 | 1 |
| VC                   | O1                          |                  | N E                 | Mexico  | RC301-302            | 2 |
| VC                   | O1                          |                  | N E                 | India   | RC771, 774           | 2 |
| VC                   | O1                          |                  | N E                 | Fl, USA   | UM2711               | 1 |
| VC                   | O1                          |                  | N E                 | CB, USA (1977)  | UM972                | 1 |
| VC                   |                             |                  | N C                 | Mexico  | RC97, 98             | 2 |
| VC                   |                             |                  | N C                 | Florida, USA  | UM2269               | 1 |
| VC                   |                             | albensis         | N E                 | ATCC14547   | RC782                | 1 |
| VC                   | O1                          | Classical, Ogawa | T C                 | ATCC14035T  | RC2                  | 1 |
| VC                   | O1                          | Classical        | T C                 | 569B  | RC215                | 1 |
| VC                   | O1                          | Classical        | T C                 | ATCC11623   | RC3                  | 1 |
| VC                   | O1                          | Classical        | T C                 | DK64  | RC772                | 1 |
| VC                   | O1                          | Classical        | T C                 | O395  | RC776                | 1 |
| VC                   | O1                          | El Tor, Inaba    | T C                 | ATCC39315, N16961                                     | RC145                | 1 |
| VC                   | O1                          | El Tor, Inaba    | T C                 | Louisiana, USA  | RC290 (CDC2164-78)   | 1 |
| VC                   | O1                          | El Tor, Inaba    | T C                 | Peru  | RC90                 | 1 |
| VC                   | O1                          | El Tor, Ogawa    | T C                 | India   | RC105                | 1 |
| VC                   | O1                          | El Tor           | T C                 | Chile   | RC11                 | 1 |
| VC                   | O1                          | El Tor           | T C                 | India   | RC773, 775           | 2 |
| VC                   | O1                          | El Tor           | T C                 | Louisiana, USA  | UM2681, 2683, 2684   | 3 |
| VC                   | O139                        |                  | T C                 | AI 1877   | RC4                  | 1 |
| VC                   | O139                        |                  | T C                 | India   | RC33                 | 1 |
| VC                   | O139                        |                  | T C                 | Bangladesh  | RC30                 | 1 |
| VC                   | O139                        |                  | T C                 | Thailand  | RC12, 20             | 2 |
| VC                   | O37                         |                  | T C                 | DK13  | RC770                | 1 |
| nonVC                | <i>Vibrio mimicus</i>       |                  | N E                 | ATCC33653T  | RC5                  | 1 |
| nonVC                | <i>Vibrio mimicus</i>       |                  | N E                 | Japan   | RC6                  | 1 |
| nonVC                | <i>Vibrio mimicus</i>       |                  | N E                 | Bangladesh  | RC217-219            | 3 |
| nonVC                | <i>Vibrio mimicus</i>       |                  | N                   | Louisiana, USA  | RC54, 57, 59         | 3 |
| nonVC                | <i>Vibrio mimicus</i>       |                  | T                   | Louisiana, USA  | RC55, 56             | 2 |
| nonVC                | <i>Vibrio fluvialis</i>     |                  | N E                 | CB, USA (1998-1999)                                   | RC442, 534, 536, 541 | 4 |
| nonVC                | <i>Vibrio harveyi</i>       |                  | N E                 | deep sea vent   | RC93                 | 1 |
| nonVC                | <i>Vibrio spp.</i>          |                  | N E                 | deep sea vent   | RC95,96              | 2 |
| nonVC                | <i>Shewanella alga</i>      |                  | N E                 | deep sea vent   | RC94                 | 1 |
| nonVC                | <i>Shewanella spp.</i>      |                  | N E                 | CB, USA (1999)  | RC535                | 1 |
| nonVC                | <i>Aeromonas spp.</i>       |                  | N E                 | CB, USA (1999)  | RC540                | 1 |
| nonVC                | <i>Exiguobacterium spp.</i> |                  | N E                 | CB, USA (1999)  | RC533                | 1 |
| <b>Total</b>         |                             |                  |                     |   | <b>380</b>           |   |



### 4.2.3 Long-range ERIC PCR conditions

The amplification reactions were carried out in 50 µl volumes using the Takara *Ex Taq*<sup>TM</sup> DNA polymerase and buffer system (Takara Mirus Bio Corporation, Madison, WI). Final concentration of the PCR mixture was comprised of 1x *Ex Taq*<sup>TM</sup> buffer (with 2 mM MgCl<sub>2</sub>), 200 µM each of dNTPs, 800 nM each of ERIC primer, 1.25 units of *Ex Taq*<sup>TM</sup> DNA polymerase, and 150 ng of template DNA. PAGE (Polyacrylamide gel electrophoresis)-purified primers (Sigma-Genosys, Woodlands, TX) were used to ensure reproducibility and prevent variability among different batches of primers. The primer sequences were the universal ERIC primers as published by Versalovic *et al.* (184); ERIC-1, ATGTAAGCTCCTGGGGATTAC and ERIC-2, AAGTAAGTGAC TGGGGTGAGCG.

The PCR cycle was carried out using a thermocycler (MJ research PTC-200 Peltier Thermal cycler) and employing the calculated control option with initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 94°C for 45 second, annealing at 52°C for 1 min, and extension at 65°C for 10 min, and a final extension step at 65°C for 20 min.

### 4.2.4 Gel electrophoresis

After PCR reaction, the 50 µl reaction mixture was mixed with 10 µl of 6x modified loading buffer (15% Ficoll 400 in water, 0.01% bromophenol blue, modified from Sambrook *et al.* (158)), and 5 µl of the mixture was used for gel electrophoresis. Three microliters of molecular weight marker (HyperLadderI, Bioline USA Inc, Canton, MA) were loaded in the first, middle, and last lane of the gel for gel image normalization. Agarose gel, 1% wt/vol, was selected as the appropriate concentration for electrophoresis

separation of the long-range ERIC PCR product. Electrophoresis was performed at 120 V for 2:15 hr using Bio-Rad sub-cell<sup>®</sup> GT electrophoresis system (electrode distance 30 cm) with 15 cm x 20 cm tray. Two buffers commonly used for DNA electrophoresis, Tris-Borate with EDTA (TBE: 89 mM Tris-Borate, 1 mM EDTA) and Tris-Acetate with EDTA (TAE: 40 mM Tris-Acetate, 1 mM EDTA); and two different comb thicknesses, 0.75 mm and 1 mm were compared for their ability to provide sharp, high resolution fingerprint patterns. After electrophoresis, the gels were stained in ethidium bromide solution (5 µg/ml) for 3-5 min and destained in tap water for 20 min, with shaking. The fingerprint banding patterns were recorded using an imaging system, FluorImager<sup>™</sup>575 (Molecular Dynamics Inc., Sunnyvale, CA) at resolution 100 micron, 16 bits, at 650 PMT voltage output.

#### **4.2.5 Image analysis**

The fingerprint pattern in the gel was analyzed using a computer software package, GelComparII version 3.0 (Applied Maths BVBA, Belgium). After background subtraction and gel normalization, the fingerprint patterns were subjected to cluster analysis using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA). Two methods for similarity measuring, one based on binary data of occurrence of the band (band-based) calculated using the Dice coefficient and another method based on overall densitometric profile (curve-based) of the banding pattern calculated using Pearson's product moment correlation were compared.

#### **4.2.6 Verification of genome relatedness estimation by ERIC PCR fingerprinting**

To evaluate the resolution and accuracy of phylogenetic inference using the optimized long-range ERIC PCR fingerprinting, relatedness estimates among strains obtained from the ERIC fingerprints were compared to the estimates obtained by DNA-DNA hybridization, the standard method for estimation of relatedness among bacterial genomes (177, 189).

##### **A Estimation of genome relatedness by DNA-DNA hybridization**

Among the 380 strains subjected to ERIC PCR fingerprinting, 176 strains presenting unique fingerprint patterns were selected for DNA-DNA hybridization. While the fingerprinting method provided estimation of relatedness for all possible pairs of isolates in the experimental set, such an exhaustive pairwise comparison is not feasible with DDH in which pairwise DNA reassociation values of strains must be determined by 176 sets of labor-intensive experiments. An effective strategy for DDH is retaining the minimal number of probe strains while obtaining maximal diversity of relatedness between probes and target strains. In this study, this was achieved by selecting one strain at the proximity to the centroid of each major cluster (determined by ERIC PCR fingerprinting) to serve as hybridization probe.

*Selection of probes.* Two independent cluster analyses were performed: (1) principal component analysis (PCA) on binary coded data from band calling (band-based), and (2) Unweighted Pair Group Method with Arithmetic Mean (UPGMA) from a similarity matrix of Pearson correlation coefficients of intensity curves (curve-based). For the first, the scores from three major principal components of PCA, which explained about 20% of total variance in the correlation matrix, were plotted to visualize clusters.

The centroid areas for each cluster were arbitrarily determined by observing the relatively high density of strains to three areas (the red ellipses in Fig. 4.1). A selection of strains in the centroid proximity of each cluster was then compared with their positions on the dendrogram produced by the UPGMA analysis. For each of centroid-proximity strains from each cluster, the cluster membership and branching order were examined. When the strain showed consistent cluster membership from both of the cluster analyses, and when its branching order in the UPGMA cluster was not located on a peripheral branch, the strain was regarded as a centroid strain of the cluster. For further analysis, one centroid strain from each cluster was randomly selected as the probe strain. Exception to this process was the cluster containing epidemic strains, such as *V. cholerae* O1 El Tor and *V. cholerae* O1 Classical, in which RC145 (N16961) was preferentially selected to take advantage of the fully determined genomic sequence as a calibration genome (78). Finally, the three strains, RC145, RC466, and RC395 were selected as probe strains to represent three major clusters of *V. cholerae*.

*Strain selection and experimental design.* Duplicate sets of the 176 strains with unique ERIC PCR fingerprint patterns were selected for membrane DNA-DNA hybridization. To prevent systematic bias of the signal intensity generated from different DNA positions on the membrane, the genomic DNAs from each set were randomly assigned to a position among the two nylon membranes. Duplicates of DNA from the three probe strains were included in every membrane as positive control. Duplicates of the type strain of *V. cholerae* (RC2), *Vibrio mimicus* (RC5), *Vibrio fluvialis*, and *Aeropyrum pernix* were also included in every membrane as control strains providing

different levels of genome relatedness to the probe. The *A. pernix*, an archeal bacterium was included as the negative control.

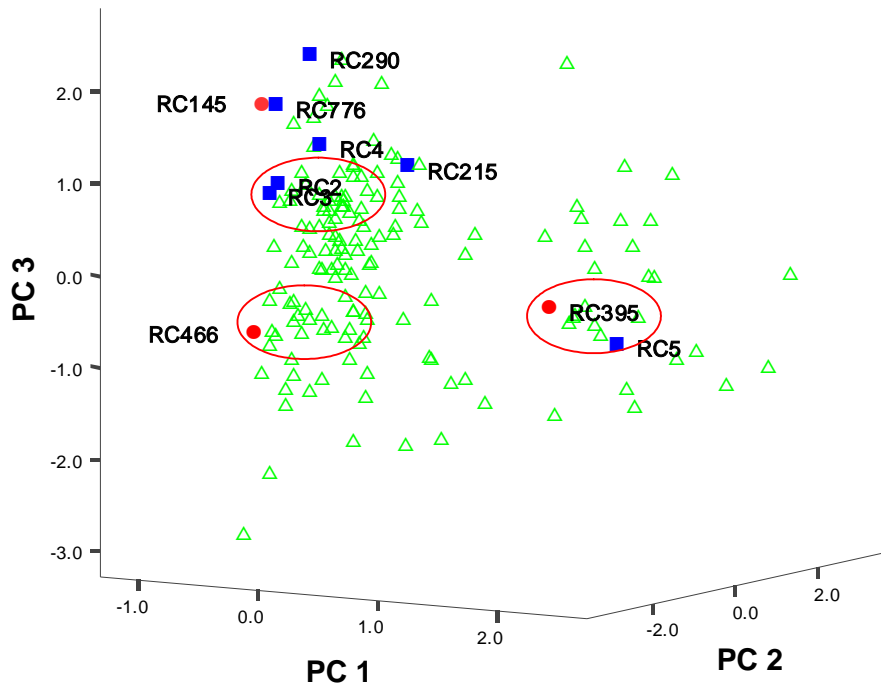


Figure 4.1. Three dimensional scatter graph of principle component analysis (PCA) from ERIC PCR fingerprints of representative *V. cholerae* and *V. mimicus* strains. The first three principal components (PC1, PC2 and PC3) are plotted. Eigen values were extracted from the correlation matrix of binary data (absence and presence of bands at 67 aligned band positions) of ERIC PCR fingerprints. The large red ellipses were identified as cluster centroid area. Red closed circles are the strains selected as DDH probe, and blue closed squares are well-characterized cholera pathogens except for RC5, the type strain of *V. mimicus*. All other strains are marked as green triangles.

*DNA-DNA hybridization.* The genomic DNA (500 ng) was dot blotted onto nylon membranes as described earlier (chapter 3). The probe genomic DNA were sheared to an approximate size of 400-600 bp by sonication and labeled by thermostable alkaline phosphatase enzyme using Geneimages™ AlkPhos Direct™ labeling kit (Amersham Biosciences Ltd, Buckinghamshire, England). Hybridization buffer and washing solution were prepared following the manufacturer's protocol. The membrane was prehybridized at 60°C for 30 min, and then hybridized (10 ng/ml probe) at 60°C overnight in a rotary hybridization tube. The membrane was then subjected to high stringency wash twice with primary wash buffer for 10 minutes at 70 °C, followed by low stringency wash twice with secondary wash buffer for 5 min at room temperature. Chemifluorescent signals were generated using ECF substrate (Amersham Biosciences). The fluorescent signals were recorded by an imaging system, Storm™840 (Molecular Dynamics Inc., Sunnyvale, CA) and the signal intensity was quantified by ImageQuant software version 5.1 (Molecular Dynamics, Inc.).

## **B Statistical comparison of the two methods for genome relatedness estimation**

To propose the use of ERIC PCR fingerprinting as a reliable method to measure genome relatedness, concordance of the estimate from this method was compared to that of the standard method (DDH) in terms of (1) 'agreement' of the two estimates by strong proportional relationship, and (2) 'accuracy' of the ERIC PCR estimate based on lack of bias in the functional relationship between the two methods. In addition, relative 'precision' of the two methods were assessed to comprehend the differences in the resolution power of the methods in discriminating distinctive strains. Schematic diagram showing the analysis procedures are shown in Fig. 4.2 and described below in detail.

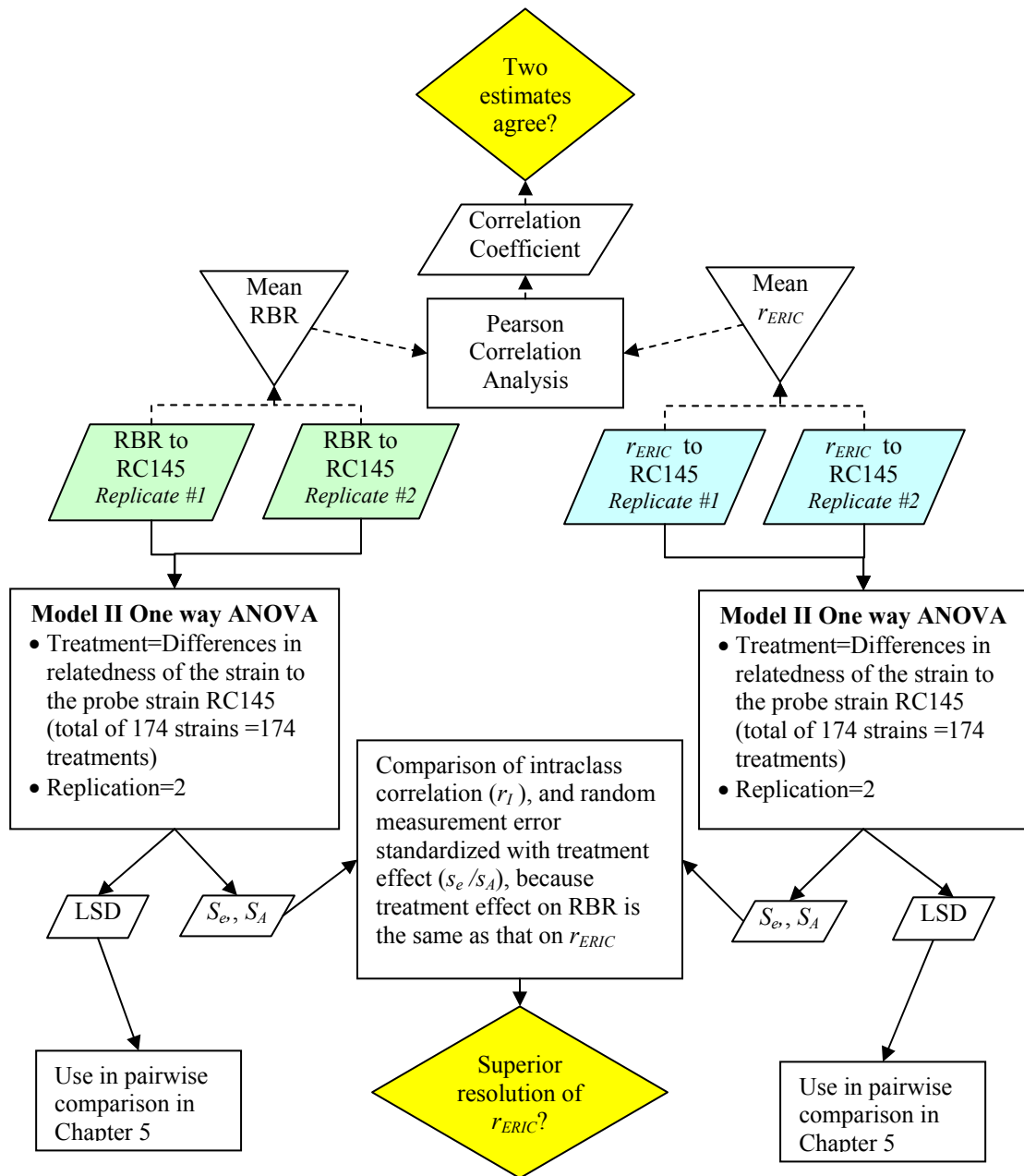


Figure 4.2. Schematic flow diagram of statistical analyses for comparison of DNA-DNA hybridization and ERIC PCR fingerprinting in estimating genome relatedness between a *V. cholerae* strain and a probe strain (RC145). Solid arrows denote flow of processes to compare resolution of the two methods; dashed arrows indicate the flow of processes to check the agreement of measurement between the two methods.

*Data Processing.* The results of the DNA-DNA hybridization are typically reported as relative binding ratios (RBR) using average of signals generated from the homologous DNA (the probe strain binding to itself) as 100 %. In this study, the average signals generated from *A. pernix*, which has the furthest relationship to *V. cholerae*, was used as base level (i.e., 0%) of hybridization signal. Typically, *A. pernix* DNA did not produce measurable level of signals after background subtraction. For the correlation analysis and ANCOVA analysis, the average RBR of the duplicate measurements were used for each strain. For Model II ANOVA, however, each RBR measurement was used to estimate the random measurement error.

For ERIC PCR fingerprinting, relatedness between two strains was measured as Pearson correlation coefficient ( $r_{ERIC}$ ) since it was determined to be a better method than the band-based calculation (see result, Fig. 4.4). Therefore, the theoretical maximum of the  $r_{ERIC}$  equals one while the minimum was set as zero. In actual estimation of  $r_{ERIC}$ , however, the definition of minimum estimation is ambiguous because the correlation coefficient was insignificant or even negative as the pair of target strains is less related. For the clarity of the analyses, those strains producing insignificant or negative  $r_{ERIC}$  value were trimmed from the data set because the genome deviation between the pair of strains was beyond the detection range of the method.

*Comparison of Resolution.* To compare the relative precision of the two methods, variances of RBR and  $r_{ERIC}$  determined for the RC145 probe in duplicate was decomposed and compared by Model II one-way ANOVA (172), regarding 174 strains of interest as a random factor of variance (Fig. 4.2). To meet the assumptions of homogeneity of variances and normality of error terms, RBR was transformed by square



root, and  $r_{ERIC}$  by arcsine transformation. Because RBR and  $r_{ERIC}$  are in different quantities (i.e., different units), direct comparison of random measurement error ( $s_e^2$ ) was not appropriate to give contrast in precisions of the two variables. To overcome this limitation, two standardized estimates were calculated and compared between RBR and  $r_{ERIC}$ . The first standardized estimate, the relative magnitude of measurement error and treatment effect ( $s_A^2$ ) for a variable was calculated as coefficient of intraclass correlation ( $r_I$ ; Sokal and Rohlf, 1995, pp 213), which standardize the treatment effect by the sum of treatment effect and measurement error ( $s_A^2 + s_e^2$ ), and  $(1 - r_I)$  provides estimation of relative magnitude of measurement error ( $s_e^2$ ) to the sum. The second standardized estimate, ratio of measurement error to treatment effect ( $s_e/s_A$ ) was used, and it allowed direct comparison between RBR and  $r_{ERIC}$  measurements because the treatment effect is identical in both methods when identical set of strains was used in the samples.

*Functional Relationship between RBR and  $r_{ERIC}$ .* Using mean values of duplicate RBR and  $r_{ERIC}$  against each of the three probe strain (RC145, RC395 and RC466), relationship of  $r_{ERIC}$  to RBR was analyzed by regression as a part of analysis of covariance (ANCOVA; Sokal and Rohlf, 1995(172)). This analysis was performed to meet the two goals: (1) checking on the parametric accuracy of  $r_{ERIC}$  in estimating the genome relatedness estimated by DDH; and (2) possibility of generalized (i.e., not specific to each probe) relationship between the two estimates of genome relatedness.

For the latter, the effect of differences in the probe strain was used as treatment of the model, where  $r_{ERIC}$  was used as covariate to fit the variation of the dependent variable RBR. The model also incorporated interaction between the covariate and the treatment effects. Therefore, the equation of model was  $Y_{ij} = \alpha_i + \beta_i X_j + \alpha \beta_i X_j + \varepsilon_{ij}$ , where Y is square

root of RBR, the covariate  $X_j$  is arcsine of square root of  $r_{ERIC}$  for strain  $j$ ,  $\varepsilon$  is error term, and  $\alpha_i$ ,  $\beta_i$  and  $\alpha\beta_i$  is the intercept, slope and the coefficient for interaction between the covariate and the treatment for probe  $i$ , respectively.

To evaluate the accuracy (i.e., absence of bias) of  $r_{ERIC}$  in estimation of genome relatedness, the functional relationship between  $r_{ERIC}$  and RBR was produced from the linear regression using the ANCOVA model described above. Because RBR is an unbiased estimate of genome relatedness, the accuracy of  $r_{ERIC}$  as an unbiased alternative estimate can be proved if accurate RBR value can be predicted from a given  $r_{ERIC}$  value using the functional relationship. The prediction of RBR=1 when  $r_{ERIC}=1$ , which occurs in the estimation of relatedness between identical target and probe strains, is an ideal case for this purpose because this prediction does not involve any measurement error. Therefore, this study checked if the functional relationship produced from ANCOVA model was compatible with the prediction by examining whether 95% confidence interval (CI) of regression lines and slopes include the prediction value.

## 4.3 Results

### 4.3.1 Optimization of the ERIC PCR fingerprinting protocol

#### A Long range-ERIC PCR conditions

*Taq enzyme and PCR buffer.* An improved, higher fidelity Taq polymerase enzyme and PCR buffer system (Takara *Ex Taq<sup>TM</sup>*) was used to provide higher reproducibility of the PCR reaction and increase the range of the amplicon size. In this experiment, the amplicon size as large as 10 kb was reliably amplified (Fig. 4.3), an improvement from the previous report of maximum amplicon size of 3-4 kb with normal Taq conditions.

*PCR conditions.* To optimize the PCR conditions, the extension temperature at 65 and 70°C, for 5 and 10 min, reaction volumes of 50 and 20 µl, and template DNA from 1 to 200 ng per reaction were tested on three randomly selected *V. cholerae* genomic DNA with three different DNA extracts. The results from all conditions yielded similar banding patterns with slightly different background and band intensity in some positions. The conditions giving the most reproducible pattern from all the DNA templates tested were selected and used in this study (see Materials and Methods).

#### B Electrophoresis

One way to increase the resolution power of the fingerprint, especially from complex banding patterns such as the one obtained from PCR-based fingerprinting, is to optimize the electrophoresis conditions. To do this, several factors were considered, such

as concentration of the agarose gel, choice of loading buffer, running buffer and comb thickness.

*Agarose gel concentration.* Since ERIC PCR randomly amplifies genomic DNA, the exact size and number of amplified products in each size range was not known. In order to select a suitable electrophoresis condition for the band separation, several agarose gel concentrations were tested. Among the gel concentrations tested (0.7-3%), 1% agarose gel gave the best banding pattern providing the most bands with good band separation and was selected. It effectively separated bands in the size range 400-10,000 bp from amplicons obtained in this long-range PCR protocol.

*Loading buffer.* To reduce the smiling effect of the bands, a non-glycerol based loading buffer was recommended. The dye concentration was also reduced to minimize hindering the banding pattern by the dye color. Using commercially available loading buffer, as well as the buffer prepared according to the standard formula yielded a background irregularity of the banding pattern in the dye front area. Although this effect was considered minor in other applications, it could reduce overall sensitivity of the fingerprinting profile when digitized signals of every image pixel were compared. The modified loading buffer reported here (Materials and Methods) provide sufficient visual aid in sample loading with minimal effect on the banding pattern. However, since the dye will be almost invisible after the electrophoresis starts, the run was stopped at a preset time instead of visually inspected the running distance of the dye front.

*Comb thickness.* Effect of the comb thickness on fingerprinting patterns is shown in Fig. 4.3 A-B. The thinner comb yielded sharper bands and better resolution, especially

when two bands were very close to each other (see black arrows pointing right).

Therefore, the thinner comb was concluded to be preferable to the thicker comb.

*Electrophoresis buffer.* Choices of electrophoresis buffer also had an effect on the banding pattern. Two commonly used buffers, 1x TAE and 0.5X TBE, were tested. The results shown in Fig. 4.3 B-C clearly showed the superiority of TBE over TAE in resolving the fingerprint banding pattern obtained from long-range ERIC-PCR. Many bands that were not visible in TAE were revealed when TBE buffer was used (see white arrows pointing left). Overall banding patterns obtained using TBE buffer appeared to be sharper, with less smearing and providing higher resolution. Although the TBE buffer provided slower running rates, longer running times can be used to obtain the same running distance. Therefore, TBE was the buffer of choice in the optimized electrophoresis protocol for this study.

*Staining of the gel.* Two DNA staining dyes, SYBR Gold and ethidium bromide were compared. The SYBR Gold is known to be more sensitive in staining DNA and causes less background fluorescence from the gel matrices than ethidium bromide. However, to discern distinct banding patterns from the background of other smearing, non specific products such as the genomic fingerprint obtained in this study, less sensitive dye was deemed preferable. When SYBR Gold was used, the overall fluorescent intensity was so high that distinct bands could not be obtained, especially large DNA fragments and the area where several DNA bands appeared to be close to each other. Better banding resolution was obtained when the less sensitive ethidium bromide was used.

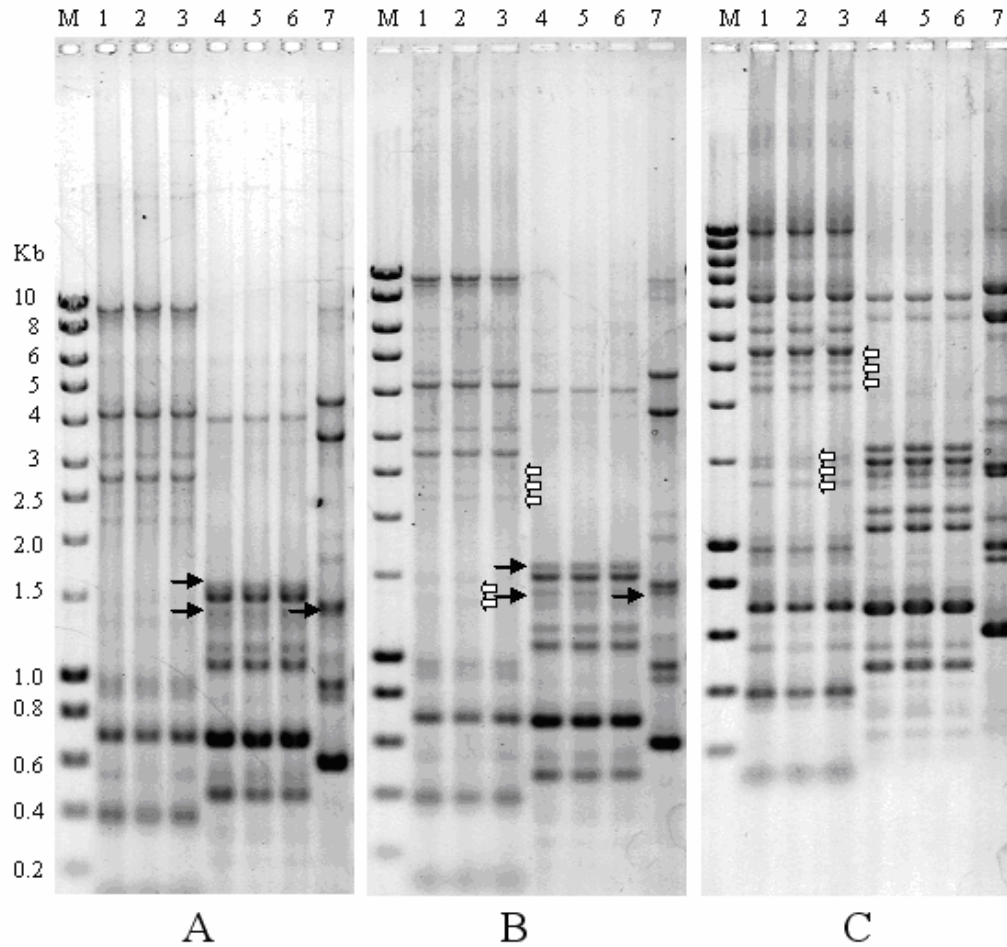


Figure 4.3. Effects of comb thickness and running buffer on the banding patterns in ERIC PCR fingerprints. Panel A = 1 mm-comb, 1x TAE buffer; B = 0.75 mm-comb, 1x TAE buffer; C = 0.75 mm-comb, 0.5x TBE buffer. Legend: M = molecular size marker (HyperLadderI); lane 1-3, *V. mimicus* (RC 217); lane 4-6, *V. cholerae* (RC 561); lane 7, *V. cholerae* O1 (RC 145). Black arrows pointing to the right in panel A and B show that the thinner comb resulted in better resolution and sharpness of the bands. White arrows pointing to the left in panel B and C show that more bands are visible using TBE as a running buffer compared to TAE running buffer.

### 4.3.2 Interpretation of fingerprint

After the PCR and electrophoresis conditions were selected, genomic fingerprints of the different bacterial isolates were obtained. The fingerprint images were imported to image analysis software GelComparII. After gel normalization, the fingerprint patterns were subjected to cluster analysis. Two options for calculating the pattern similarity for the cluster analysis, band-based and curve-based, were compared.

*Band-based.* Similarity was calculated from binary data of common and different bands in each band position. In order to calculate similarity, the bands and their positions needed to be assigned in a process termed “band calling”. Difficulties in the band calling process were recognized. First, accuracy in assigning presence of the bands, especially for faint bands from different fingerprints, had to be carefully inspected. This was overcome by setting the minimum criteria, such as the intensity curve height or area. However, the criteria were arbitrary, depending on the decision of the investigator. Second, difficulty in assigning the band position (band class) did not provide confidence in the result. Assigning the band position of thick bands, which is common in PCR-based fingerprinting, was arbitrary and not reproducible. The same thick band from different fingerprints had a tendency to be wrongly assigned to a different band class. On the other hand, different thin bands that appear to be close to each other sometimes were wrongly assigned to the same band class. Although computer assisted, the majority of the results needed exhaustive manual inspection and were subjected to individual decision. The reliability and reproducibility of the band calling process were reduced when the fingerprint banding pattern became more complex and when more patterns were included in the analysis.

*Curve based.* Similarity value was calculated based upon densitometric curve of the fingerprint pattern. This method provided more consistent results with less manual intervention. Since the method takes into account the overall pattern of the fingerprint, it is less sensitive to small variations due to faint bands or small shifts in the overall patterns.

*Selection of the calculation method.* In selecting suitable options for the fingerprint similarity calculation, a set of 21 fingerprint patterns obtained from the same isolate (RC145) was used as a sample set. Since all fingerprint patterns were from the same isolate, a higher similarity value indicates better calculation protocol. The similarity calculated based on band-based (using Dice coefficient) and curve-based (using Pearson product-moment correlation) were compared using percent similarity value obtained from a UPGMA clustering tree. We found that the curve-based calculation gave the higher similarity value than the band-based method (as example of a pair of fingerprint shown in Fig.4.4). For the band-based calculation and after close inspection, position and member of each band class had to be reassigned several times to achieve the same high similarity level. Therefore, for robustness and consistency, the similarity value calculated by curve-based using Pearson correlation was selected.

*Comparison range.* The DNA size range for comparison was set at 400-10,000 bp. From the same data set of RC145 fingerprints, we found that the DNA band size smaller than 400 bp was more diffused and less consistent. This was probably the result of the property of the 1% agarose gel electrophoresis used. The 10 kb upper limit of the comparison range was set, considering fidelity of the long-range Taq enzyme system.



Although larger than 10 kb amplicon size may be obtained, consistency of the large amplicon may not be reliable in all PCR reactions.

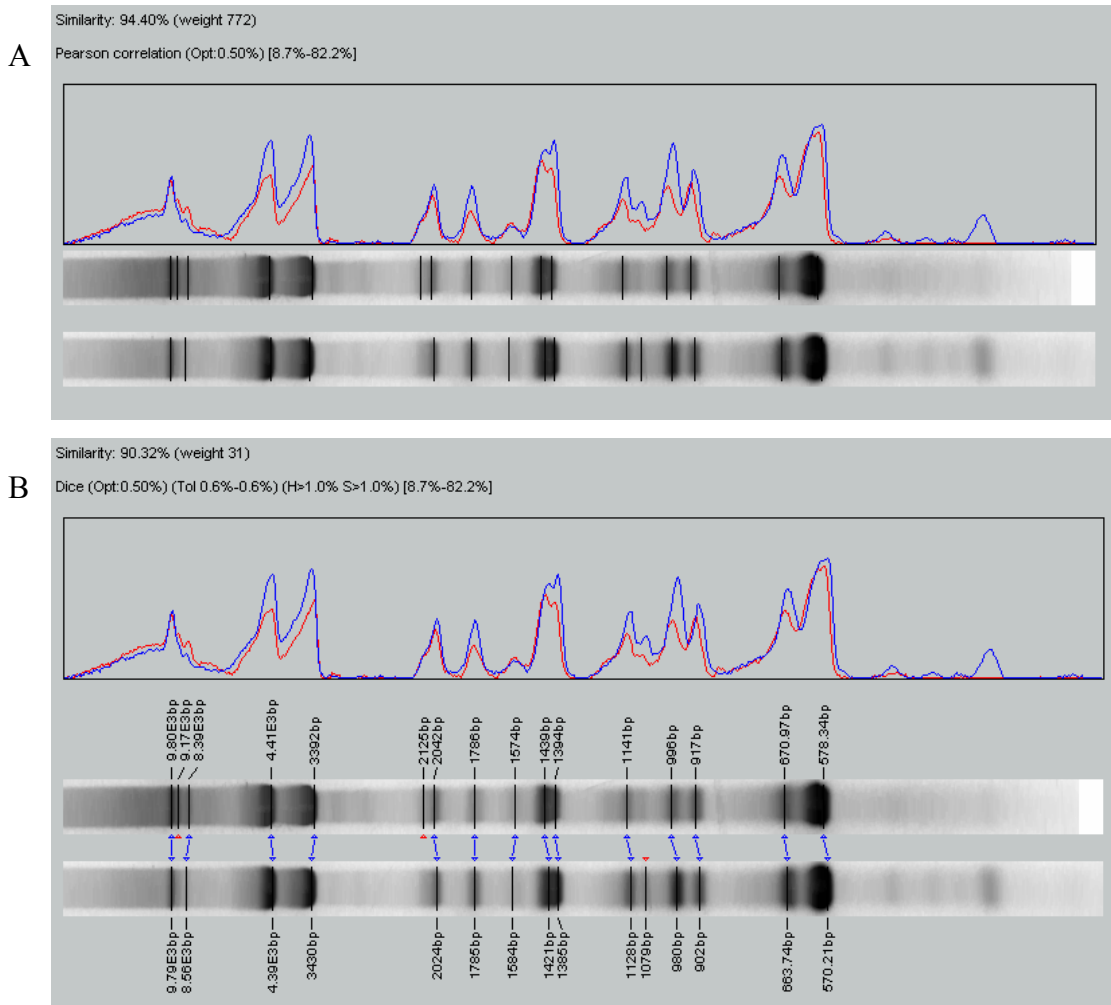


Figure 4.4. Example of curve-based and band-based calculation for obtaining similarity value between pair of ERIC-PCR fingerprint. Panel A is a curve-based calculation, by Pearson's correlation coefficients on the densitometric curves generated from ERIC-PCR fingerprints; panel B is a band-based calculation using Dice coefficient of band positions that was assigned by the computer software and manually corrected. The same pair of ERIC fingerprint from identical strain was used for the comparison.

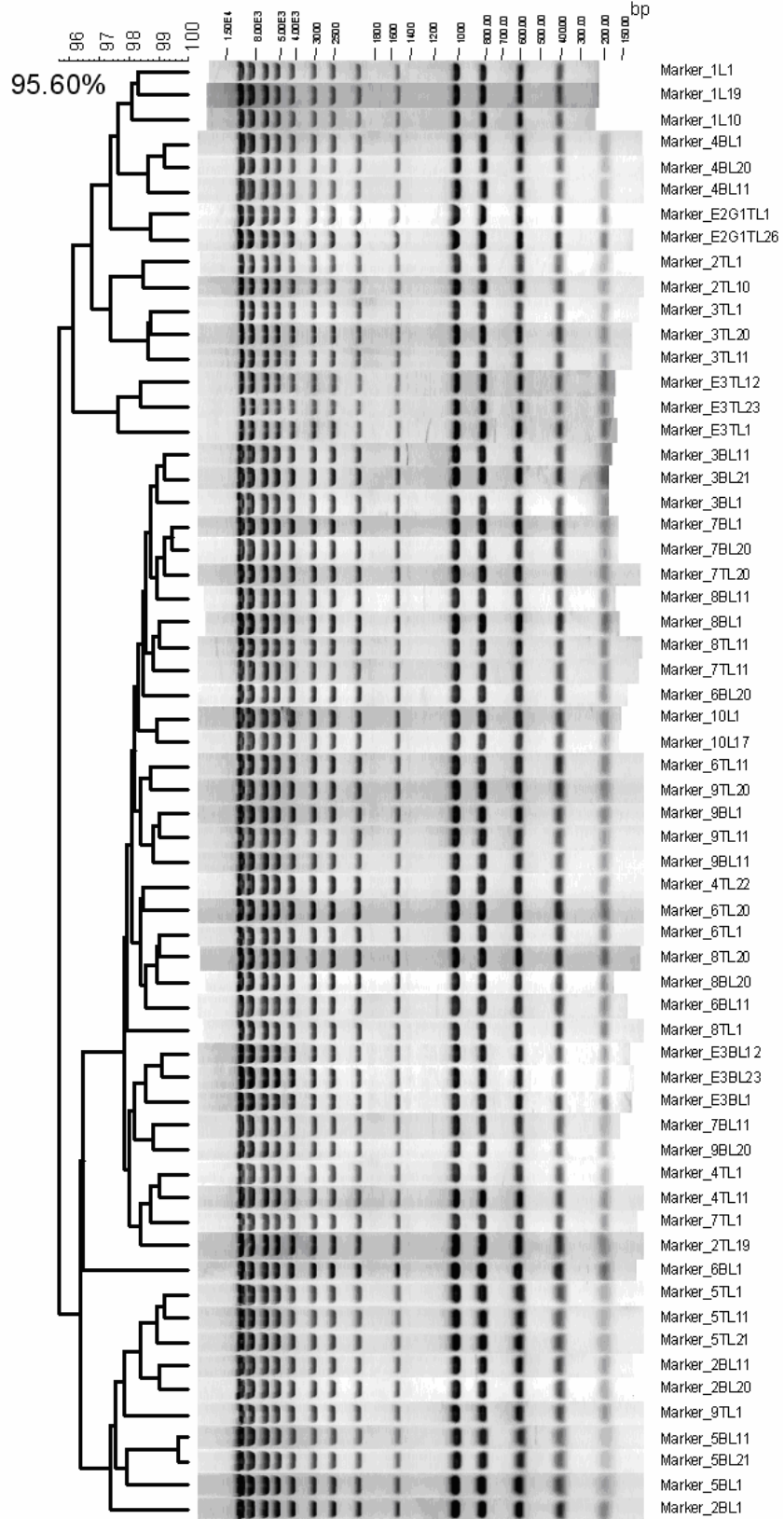
### **4.3.3 Reproducibility of the long-range ERIC PCR fingerprinting**

Overall reproducibility of the ERIC PCR fingerprinting was estimated by inspecting the variability (error) due to different steps in the protocol.

*Variation from electrophoresis and image analysis.* The variability due to electrophoresis and image analysis was evaluated from the DNA size markers that were run on different electrophoresis gels. Each gel image was normalized and the similarity values were calculated by curve-based Pearson's correlation. The results shown in Fig. 4.5 indicate that all 61 banding patterns from the DNA size marker clustered at >95% similarity on the UPGMA tree. The 5% variation can be explained by differences in the electrophoresis running distance, different background of each gel, and error in the normalization of the gel images.

Figure 4.5. (next page) Variability in fingerprint similarity caused by gel electrophoresis. The dendrogram of DNA size marker (HyperLadderI) from different electrophoresis gels was based on similarity value calculated in the range of 400-10,000 bp (8.7-82.2% running distance on the gel, with 0.5% optimization) by Pearson's product moment correlation (Pearson correlation) and clustered using the unweighted pair group method with arithmetic mean (UPGMA). The minimum similarity coefficient of the cluster is presented in percent at the top left of the dendrogram.

Pearson correlation (Opt:0.50%) [8.7%-82.2%]



*Variation from PCR reaction.* In addition to variation from electrophoresis and image normalization, genomic DNA fingerprints of a reference DNA (RC145) amplified from different PCR reactions were used to estimate variability caused by the PCR reaction. The results in Fig. 4.6 showed that all 21 fingerprints obtained from RC145 had the same banding pattern, but slightly different in band intensity, and all were clustered at >90% similarity on the UPGMA tree. Therefore, the PCR amplification step added 5% variability to the analysis. The sources of variation due to PCR reactions can be explained, in part, by pipetting error, different batches of enzyme and buffer, and different PCR machines (two different PCR boxes with different ramping rate were used throughout the study).

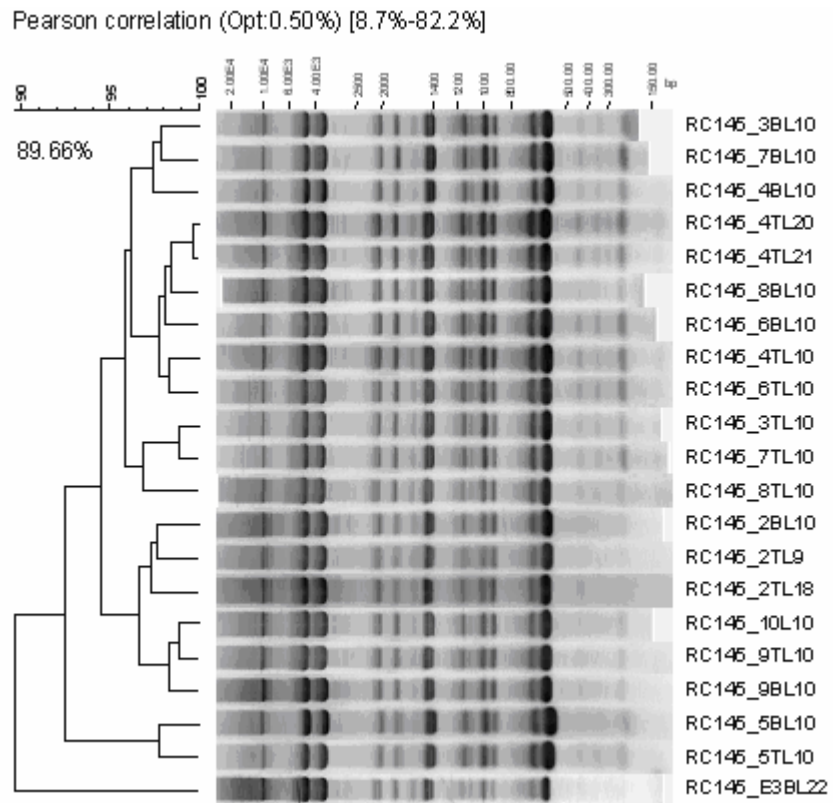


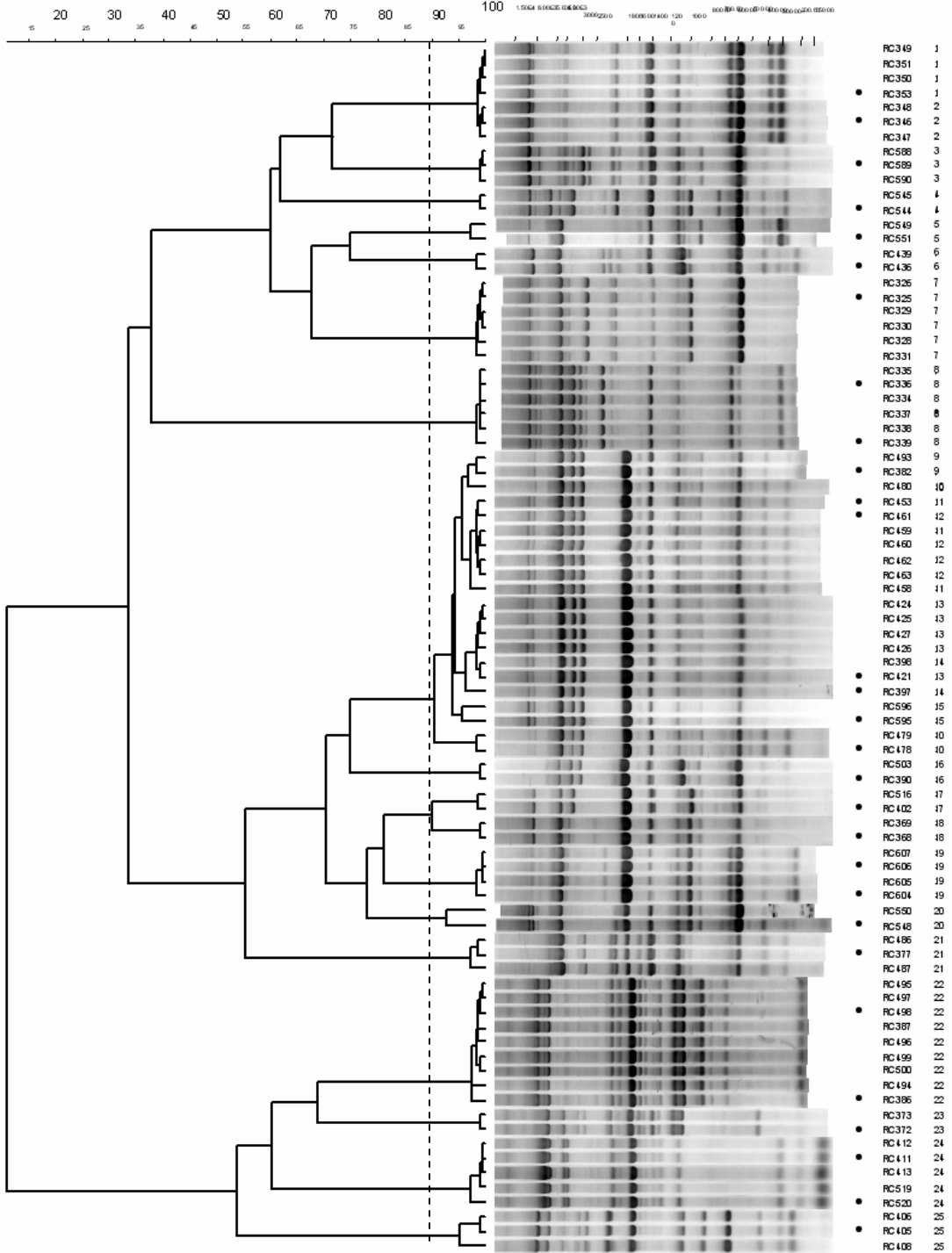
Figure 4.6. Variability in fingerprint similarity caused by different PCR reactions. The dendrogram of RC145 fingerprint pattern obtained from different PCR reactions was based on similarity calculated in the range of 400-10,000 bp by Pearson correlation and clustered using UPGMA method. The minimum similarity coefficient of the cluster is presented in percent at the top left of the dendrogram.

#### 4.3.4 Identification of artificial clones

Isolation of *V. cholerae* from environmental samples usually involves enrichment steps. Enrichment provided a suitable environment for the bacterium to multiply, hence increasing the chance of being detected. However, it created redundancy of specific clones that multiplied in enrichment media, not reflecting a real abundance of the clone in the sample. With the use of ERIC PCR fingerprinting, such redundancy can be detected. In this study, we defined cells presumably having multiplied from the same cell in the enrichment step as an “artificial clone”. The artificial clones were identified when three criteria were met. First, the isolates were from the same enrichment flask; second, they had same phenotype and genotype; and third, they provided the same fingerprint pattern. The isolates were considered to have the same fingerprint pattern when there was no band difference by visual inspection and showed >90% similarity from UPGMA tree using the Pearson correlation coefficient. In total, 25 groups of artificial clones were detected from the Chesapeake Bay strain collection (Fig. 4.7). One representative from each group of artificial clones was selected and the rest (54 isolates) were removed from the set designated for further study.

Figure 4.7. (next page) Cluster analysis of ERIC PCR fingerprints for identification of artificial clones. Same number at the end of each fingerprint indicates that the isolates were from the same enrichment flask. The isolates selected as the clone representative are marked by a dot in front of the strain name. The 90% similarity level is indicated by a vertical dashed line.





#### 4.3.5 Resolution power of the long-range ERIC PCR fingerprinting

Ability of the ERIC PCR fingerprinting protocol to differentiate different bacterial strains, a measure of the protocol resolution power, was tested. Fingerprint patterns obtained for bacteria with a different level of genome similarity were used, i.e. different genus, species, and *V. cholerae* isolates from different environmental sources, as well as clinical isolates with known clonal complexes, such as O1 classical, O1 El Tor, and O139. An example of the strains tested is shown in Fig. 4.8. The ERIC PCR fingerprint was able to differentiate *V. cholerae* from other species, i.e. *Vibrio harveyi*, *Shewanella sp.*, *Vibrio mimicus*, *Vibrio fluvialis*, and *Aeromonas sp.*, as well as reveal diversity of the environmental isolates of *V. cholerae*.

Among the toxigenic strains of O1 classical, O1 El Tor, and O139; a close-up picture of the ERIC PCR fingerprints (Fig. 4.9) showed similar banding patterns and strains of different serogroups were clustering together at >90% similarity in the UPGMA tree, i.e., within the error range of the ERIC-PCR method. However, after careful inspection of the banding patterns, two band positions at 1.4 kb and 1 kb were found to discriminate the three closely related clones (black arrows, Fig. 4.9). Both bands were present in all the O1 El Tor, while only the 1.4 kb, but not the 1 kb band, was present in O139. The 1.4 kb was absent from the O1 classical strains.

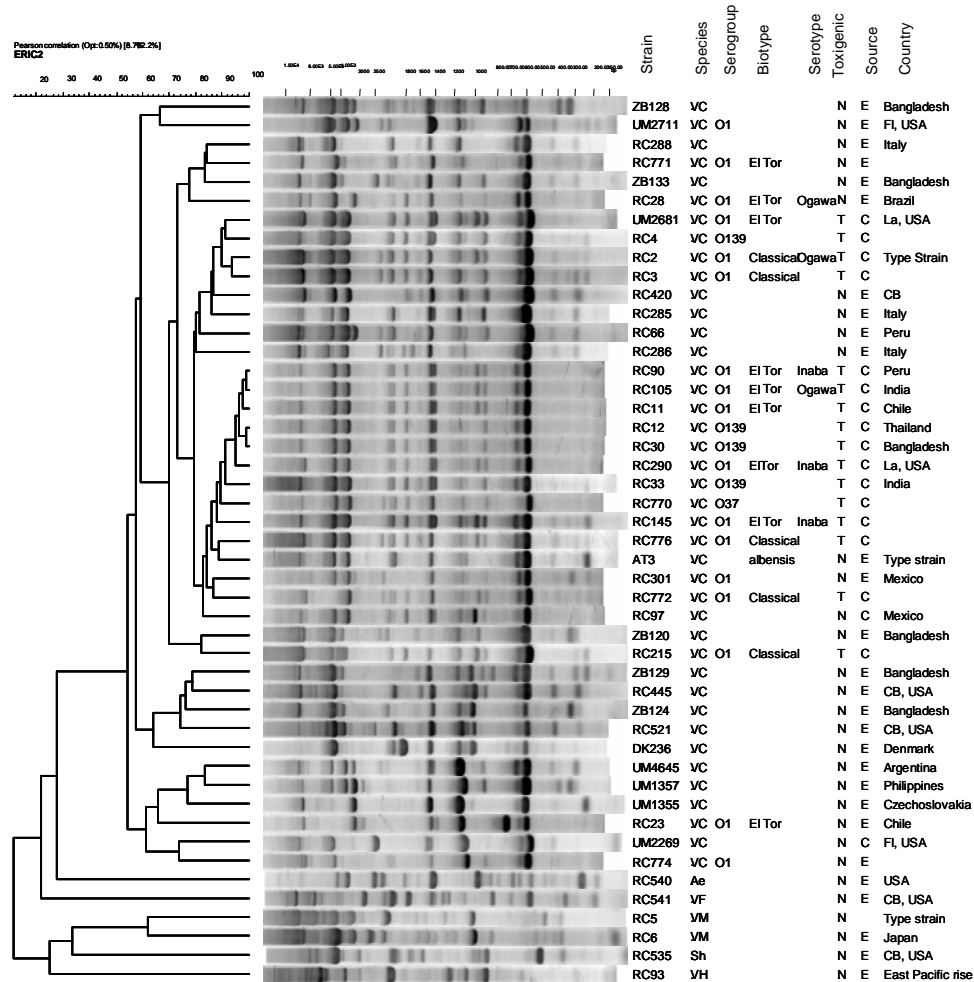


Figure 4.8. Example of fingerprints and UPGMA dendrograms obtained for different bacterial strains. Species abbreviation: VC = *Vibrio cholerae*; Ae = *Aeromonas spp.*; VF = *Vibrio fluvialis*, VM = *Vibrio mimicus*, VH = *Vibrio harveyi*, and Sh = *Shewanella spp.* Toxigenic: N = non-toxigenic; T = toxigenic (presence of CTX gene). Source: E = environmental; C = clinical.

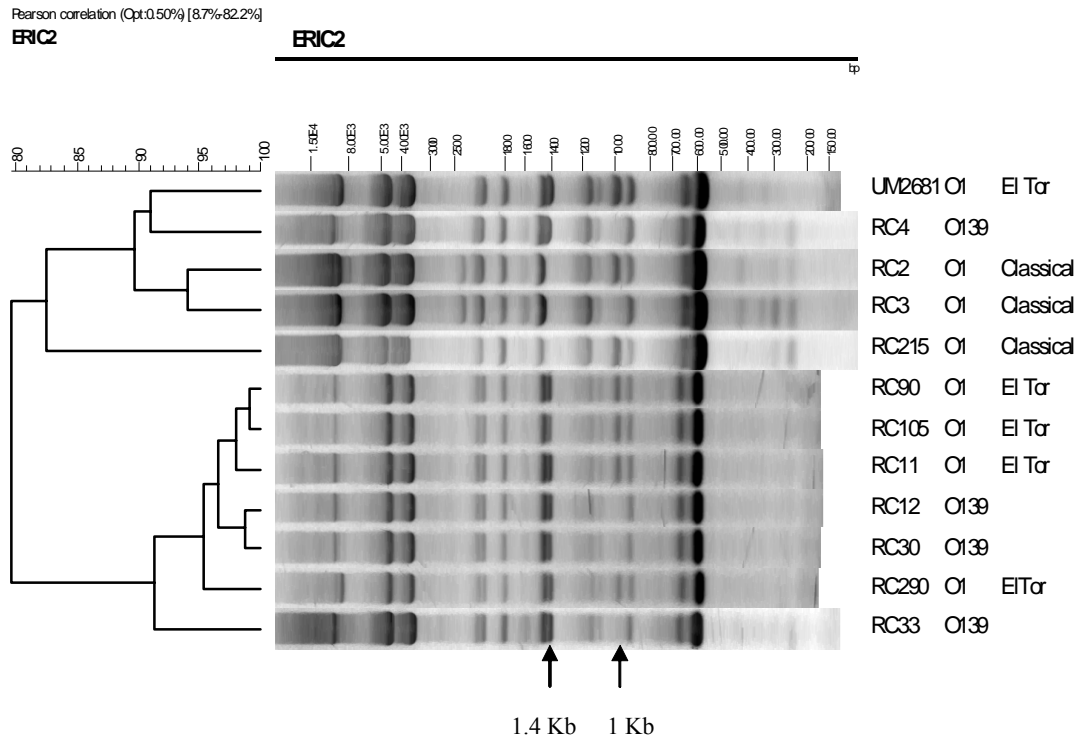


Figure 4.9. Genomic fingerprints of toxigenic *V. cholerae*. Arrows show position of the discriminative bands; 1.4 kb and 1 kb. *V. cholerae* O1 El Tor has both bands, while *V. cholerae* O139 has only the 1.4 kb band and not the 1 kb band, and *V. cholerae* O1 Classical lacks the 1.4 kb band whether or not the 1 kb band is present.

Fig. 4.9 also shows the limitation of the clustering protocol and importance of the electrophoresis step in the analysis of very closely related strains. The factors that separate these toxigenic clones into two major clusters were most likely caused by differences in band intensity (either a result of loading volume or differences in PCR efficiency) and running distance, rather than in the banding pattern itself. In the gel normalization step, when the gels with different running distances were compared against the same reference position, the shorter gel was stretched and the band appeared broader. This process affected the densitometric profiles of the fingerprints and reduced the similarity value among strains that were visually identical or very similar (e.g., among El Tor strains and among O139 strains). These effects were realized later and the causes were found to be differences in gel size as well as gelling temperature in the gel preparation step.

Therefore the ERIC PCR fingerprinting protocol reported in this study has limitations in similarity calculations of closely related strains. The error in the same banding pattern clustering at different similarity levels can be minimized by using the same PCR reaction and strict control of electrophoresis conditions. For the high similarity level (> 90%) fingerprints, the banding pattern was determined by visual inspection or the band-calling process to identify identical fingerprints.

In summary, with the ERIC PCR fingerprinting protocol presented here, it is possible to differentiate strains of *V. cholerae* O1 classical, O1 El Tor, and O139 by inspecting the banding pattern visually, especially by the differentiating band sizes of 1.4 and 1.0 kb.

### 4.3.6 Comparison of genome relatedness estimation with DNA-DNA hybridization

#### A Agreement between RBR and $r_{ERIC}$

To meet the assumptions of normality and homoscedasticity in ANOVA, it was necessary to transform RBR and  $r_{ERIC}$  into the square root of RBR ( $\sqrt{RBR}$ ) and arcsine of the square root of  $r_{ERIC}$  ( $\sin^{-1}\sqrt{r_{ERIC}}$ ), respectively. The transformed values from the two techniques were highly correlated ( $r = 0.7$ ,  $P < 0.01$ , Pearson correlation coefficient, Fig. 4.10). This result revealed that the two measurement variables are in good agreement with strong proportional relationship.

#### B Comparison of resolution

As target strains were samples of the global population of *V. cholerae*, the decomposition of variances were interpreted in the frame of Model II ANOVA (172). For a standardized comparison of variances from the variables of two different units, the treatment effects of  $\sqrt{RBR}$  and  $\sin^{-1}\sqrt{r_{ERIC}}$  are regarded to be identical because the samples strains of the two ANOVA models are identical and so is genome divergency of the sample strain set. The high correlation between  $\sqrt{RBR}$  and  $\sin^{-1}\sqrt{r_{ERIC}}$  supports the postulate by revealing the presence of a strong common treatment factor(s).

In Table 4.2, standard deviation of the source of variances (as  $s_e$  and  $s_A$ ) is shown along with two standardized variance estimates ( $s_e/s_A$ ) and intraclass correlation ( $r_I$ ). The intraclass correlation indicates that total variance in  $\sin^{-1}\sqrt{r_{ERIC}}$  was due, in most part, to treatment effect rather than by measurement error (99:1 ratio of treatment effect to measuring error, respectively). On the other hand, contribution of the measurement error was higher in  $\sqrt{RBR}$  (7:3 ratio of treatment effect to measuring error, respectively),

indicating higher levels of measurement error associated with DNA-DNA hybridization than that of ERIC-PCR fingerprinting. The larger measurement error was directly contrasted with  $s_e/s_A$  ratio, which was 5.4 fold higher in RBR. Therefore, it was shown that  $r_{ERIC}$  can provide >5 times better resolution in differentiating bacterial genomes. LSD value, which can be readily used for pairwise comparison of strains to tell the significant difference in their genomes, also showed 3 times better resolution of  $r_{ERIC}$  method.

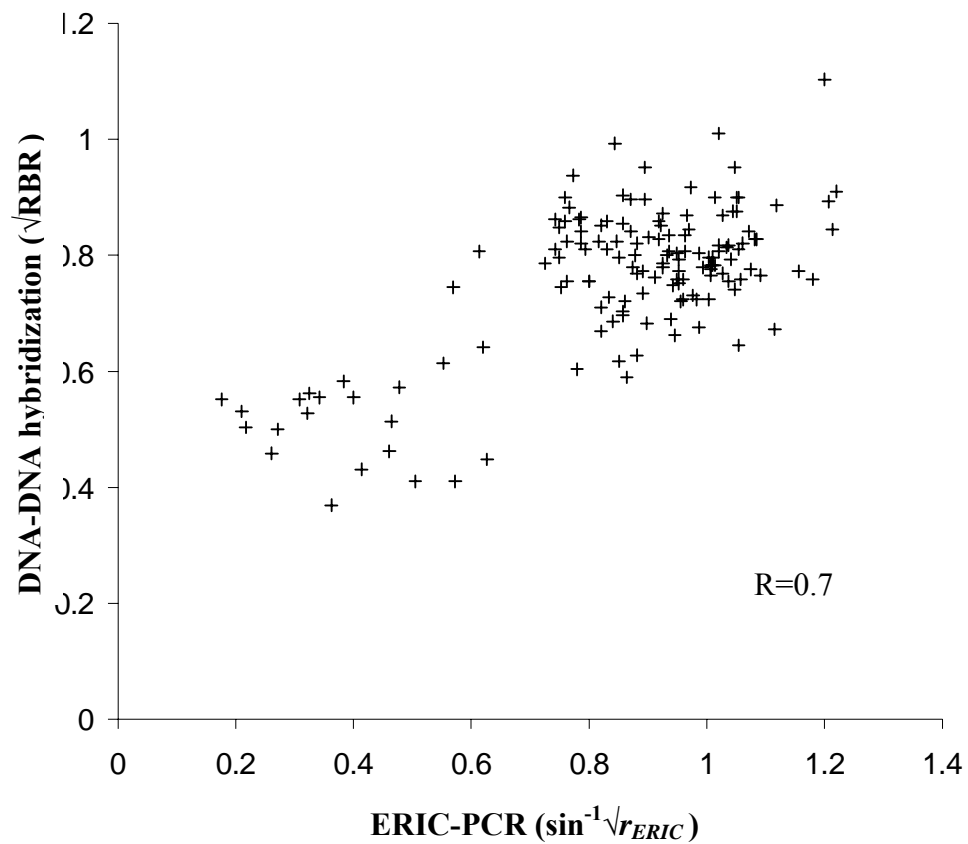


Figure 4.10. Scatter plot showing distribution of the transformed means of RBR and  $r_{ERIC}$ . Correlation between the two methods was calculated using the Pearson correlation coefficient.

Table 4.2. Estimates of variance components from ANOVA on transformed RBR and  $r_{ERIC}^a$

| Variable                   | $s_e$ | $s_A$ | $s_e/s_A$ | $r_I$ | $LSD_{\alpha=0.05}$ |
|----------------------------|-------|-------|-----------|-------|---------------------|
| $\sqrt{RBR}$               | 0.077 | 0.119 | 65%       | 70%   | 0.15 (0.0225)       |
| $\sin^{-1}\sqrt{r_{ERIC}}$ | 0.027 | 0.225 | 12%       | 99%   | 0.05 (0.0025)       |

<sup>a</sup>Description of columns

$s_e$ : Standard deviation of error; estimated as the square root of the mean square of error (MSE)

$s_A$ : Standard deviation of treatment effects by differences of target strains; estimated from the square root of the value (mean square among groups - MSE)/ $n_0$ , where  $n_0$  is effective number of samples in a group

$r_I$ : Coefficient of intraclass correlation [ $s_A / (s_A^2 + s_e^2)$ ]

$LSD_{\alpha=0.05}$ : Least significant difference at type I comparisonwise error value 0.05, number in parenthesis indicates reverse transformed back to the original scale



### C Functional Relationship between RBR and $r_{ERIC}$

Because both  $\sqrt{\text{RBR}}$  and  $\sin^{-1}\sqrt{r_{ERIC}}$  have random measurement error, Model II regression is considered as an appropriate regression between them. However, Model I regression was also feasible because of the following two reasons: First, the measurement error of  $\sqrt{\text{RBR}}$  is considerably greater (ca. 5 fold) than that of  $\sin^{-1}\sqrt{r_{ERIC}}$ , (107). Second, the measurement error of  $\sin^{-1}\sqrt{r_{ERIC}}$  is occurring randomly around the mean of the measurements (Berkson's case; Sokal and Rohlf, 1995 (172)). Because the functional relationship between RBR and  $r_{ERIC}$  may differ according to probe strain, ANCOVA with model I regression on  $\sqrt{\text{RBR}}$  was performed with  $\sin^{-1}\sqrt{r_{ERIC}}$  as covariate and three probe strains (RC145, RC369, and RC466) as treatments.

Partitioning of variance by ANCOVA indicated that  $\sqrt{\text{RBR}}$  has significant linear relationship with  $\sin^{-1}\sqrt{r_{ERIC}}$  ( $F_{1,875} = 365, P < 0.01$ ) and that the effect of difference in probe strains is also significant ( $F_{2,875} = 11.7, P < 0.01$ ). Because the interaction between  $\sin^{-1}\sqrt{r_{ERIC}}$  and probe strain turned out to be significant ( $F_{2,875} = 4.52, P < 0.05$ ), the functional relationship was analyzed separately for each probe strain.

Results of the regression analysis are shown in Table 4.3 and Fig. 4.11. All intercepts of the regressions had a positive sign and was significantly different from zero ( $P < 0.01$ ). Therefore, the identical range of genome deviation produces  $r_{ERIC}$  range of zero to one while RBR ranges 23% to 100% (determined by reverse transformation of intercepts in Table 4.3). By the *t*-test, the slope and intercept of RC145 probe was significantly different from those of RC395 probe ( $P < 0.01$ , contrast in ANCOVA) while those of RC466 were indistinguishable from both RC145 and RC395. Therefore, it is concluded that the relationship between RBR and  $r_{ERIC}$  is probe-specific.

To assess the presence of bias in the functional prediction of RBR with  $r_{ERIC}$ , the 95% CI of the each regression slope was examined for the inclusion of the expected slope from the line connecting intercept and the point  $(\sin^{-1}\sqrt{r_{ERIC}}, \sqrt{RBR}) = (\text{arcsine of } 1, 1)$ . This expected slope was calculated for each probe (Table 4.3) and plotted in Fig. 4.11. From those results, it was found that 95% CI of the estimated slopes always included the expected slope for all three probes. Therefore, it was concluded that  $r_{ERIC}$  is an unbiased estimate of genome relatedness by accurately predicting the standard unbiased estimate (RBR).

Table 4.3. Estimates of regression analysis of the three probes by ANCOVA<sup>a</sup>

| Probe | Intercept            |                     |                                    | Slope                |                     | Expected Slope <sup>d</sup> |                     |
|-------|----------------------|---------------------|------------------------------------|----------------------|---------------------|-----------------------------|---------------------|
|       | Estimate ( $\pm$ SE) | 95% CI <sup>b</sup> | 95% CI <sub>RBR</sub> <sup>c</sup> | Estimate ( $\pm$ SE) | 95% CI <sup>b</sup> | Estimate                    | 95% CI <sup>b</sup> |
| RC145 | 0.48 $\pm$ 0.02      | 0.44 - 0.52         | 0.19 - 0.27                        | 0.37 $\pm$ 0.03      | 0.32 - 0.42         | 0.33                        | 0.30 - 0.36         |
| RC395 | 0.61 $\pm$ 0.01      | 0.58 - 0.63         | 0.34 - 0.40                        | 0.25 $\pm$ 0.03      | 0.19 - 0.31         | 0.25                        | 0.23 - 0.27         |
| RC466 | 0.57 $\pm$ 0.02      | 0.53 - 0.61         | 0.28 - 0.38                        | 0.31 $\pm$ 0.03      | 0.26 - 0.36         | 0.27                        | 0.25 - 0.30         |

<sup>a</sup> For all estimates of intercepts and slopes, the null hypothesis  $H_0$ :estimate=0 was tested at  $P < 0.01$ .

<sup>b</sup> 95% CI = 95% confidence interval of  $\sqrt{RBR}$ .

<sup>c</sup> 95% CI<sub>RBR</sub> = 95% confidence interval of reverse transformed  $\sqrt{RBR}$  ( squared to convert the unit of intercept into that of the original unit, RBR).

<sup>d</sup> Expected slope was calculated from the intercept on  $\sqrt{RBR}$  axis to the point of (arcsine1,1), the expected value when the two methods are in complete agreement.

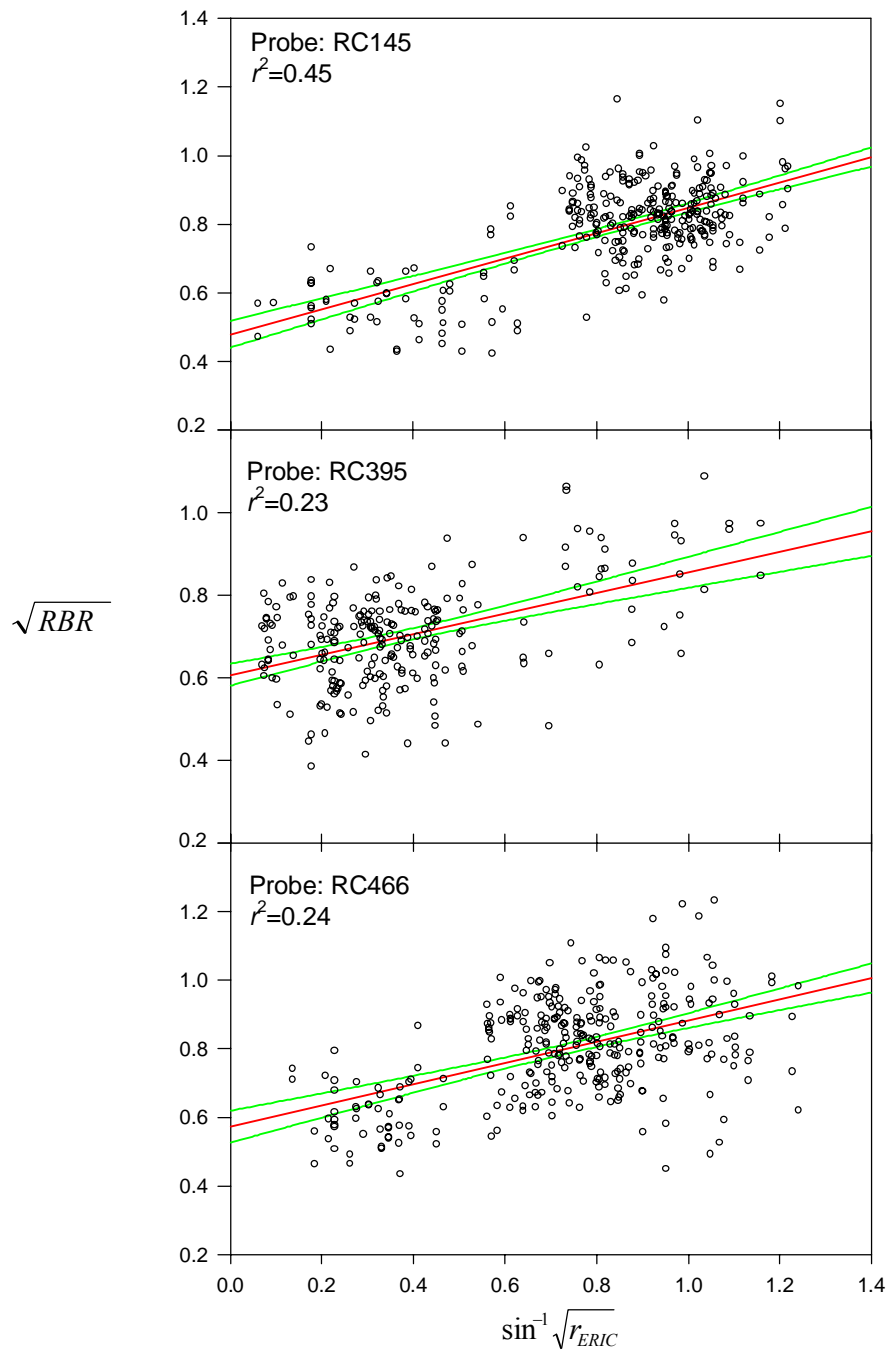


Figure 4.11. Regression analysis of transformed RBR and  $r_{ERIC}$ . Probe strains and  $r^2$  are shown on the top left of each plot. Red line shows regression line of the observed value. Green line shows upper and lower limit of the 95% confidence interval of both intercept and slope.

## 4.4 Discussion

### 4.4.1 Selection of fingerprinting protocol

To determine the identity of bacterial clones, ERIC PCR was selected as a method of choice to for obtaining genomic fingerprints of each isolate in the collection. The highly conserved 124-127 bp enterobacterial repetitive intergenic consensus (ERIC) sequence were first reported in enterobacteria *Escherichia coli* and *Salmonella typhimurium*, and later in several other bacteria, including *V. cholerae* (85, 184). Although ERIC sequences are highly conserved at the nucleotide sequence level, their chromosomal locations differ between species. Zo *et al.* (2002) simulated ERIC primers binding sites along the two chromosomes of *V. cholerae* O1 El Tor N16961 complete sequence and found that the primer binding sites are evenly dispersed throughout the entire region of the two chromosomes (209). This characteristic makes the ERIC sequence a suitable target for genomic fingerprinting because it permits sampling the entire genome without bias to any specific region.

In the long-range ERIC PCR protocol reported here, the fingerprint range was increased from the usual 3-4 kb to at least 10 kb, with highly reproducible patterns. The number of bands that were amplified from different *V. cholerae* chromosomes also increased from 1-8 (150, 167) to 6-17 bands in each strain. The increase in both size range and number of amplified fragments raised the complexity of the overall fingerprint and yielded an improved resolution power from the ERIC PCR fingerprinting protocol.

To improve reproducibility of the fingerprint, several factors influencing the fingerprint pattern need to be taken into account. Although different DNA concentrations, were found to yield similar banding patterns, only a slight variation in the

fingerprint pattern can have an adverse effect when highly discriminative and reproducible patterns are sought. Therefore, DNA quality and concentration must be the same level throughout the study. The same is true for other PCR conditions, reaction volume, extension temperature, and time in the range tested, all provided similar fingerprint patterns, however, they were not always identical. Therefore, once the desired condition was determined, it had to be kept consistent throughout the study. In less demanding cases, where rapid identification is the intent, whole cells obtained from pure culture have been used successfully in PCR genomic fingerprinting (112, 208).

#### **4.4.2 Improvement in fingerprint banding patterns by adjustment of the electrophoresis setting.**

Separation of DNA in horizontal agarose gel electrophoresis is a simple technique with which most laboratories are familiar. The method usually provides a satisfactory result within the broad range of conditions used. However, the effect of the simple parameters that are often overlooked, if controlled carefully, can improve band sharpness and overall resolution of the complex banding patterns, such as those obtained by PCR-based genomic fingerprinting. The effect of some parameters, such as choice of electrophoresis buffer, agarose concentration, comb thickness and loading buffer, has been discussed in the various technical applications in the FMC Bioproducts catalog and in a laboratory manual for molecular cloning (158). Thus, effects of some of the parameters on fingerprint patterns obtained by long-range ERIC PCR have been demonstrated. The most profound improvement of the fingerprint patterns was obtained when a thin comb (0.75 mm) was used with 0.5x TBE running buffer.

Other parameters that can affect overall resolution of the fingerprint pattern were also defined. The use of a modified loading buffer with Ficoll-400 as sinking agent and reduced dye concentration is a simple way to improve fingerprint resolution. The glycerol-based loading buffer has a lower molecular weight and allows DNA to stream up the sides of the well before electrophoresis is started, resulting in a U-shaped band (FMC catalog). We also found that for a complex banding pattern, such as those obtained in PCR fingerprinting, the less sensitive ethidium bromide was more suitable for revealing distinctive bands than highly sensitive DNA stains, such as SYBR Gold.

Differences in running distance of the fingerprint DNA were found to reduce the reliability of the method. This was unexpected because all electrophoresis gels are run with preset timing at the same constant voltage (4 V/cm of electrode distance). It was later realized that size of the gel matrix, as well as size of the tray used to hold the gel in the electrophoresis, also affect running rate of the DNA, probably by affecting overall resistance in the electrophoresis environment. Since these factors were not mentioned in the reference protocols, they are overlooked and use of each gel/tray size varied, depending on availability, resulting in differences in running distance of the fingerprints. The differences in running distance of each fingerprint lane in the same gel, in which the lane closer to the edge of the gel runs slower than the lane in the middle of the gel (gel smiling) has also been recognized. Although the gel smiling effect was well known, the cause was not. Researchers usually avoided the use of several lanes on both sides of the gel. However, this solution is not suitable and results in waste of time and resources, especially when high throughput result must be done. We found that pouring the gel suspension shortly after melting, while the gel is still very hot (but not boiling) can

reduce the smiling effect to a barely discernable level, whereas pouring cooler gels just before the solidifying temperature was reached produced very big smiling gels. The most probable explanation for this discovery is that in the hot-setting gel, the heat in the agarose is more evenly dispersed before polymerization takes place, resulting in a more homogeneous gel. With cooler gels, the temperature difference of the agarose in the middle and at the edge of the tray was greater because of travel distance of the gel from the pouring point in the middle through the cold tray until reaching the edge, causing the gel to polymerize at different times and resulted in non-homogeneous gel, causing smiling.

Unfortunately, this discovery was made at the last stage of the study after all the fingerprints had been analyzed. Therefore, visual inspection was required to resolve closely related strains. On the other hand, it showed that the resolution power of the protocol could be improved and, at the current stage of the technology, the protocol was robust and despite sub-optimum conditions, closely related strains could be differentiated.

#### **4.4.3 Interpretation of fingerprint patterns using the band-based algorithm**

When a large number of complex fingerprint patterns must be compared, computer analysis is essential. A commercial software package dedicated to fingerprint analysis, GelComparII (Applied Maths, Belgium) was used in this study. Several fingerprints from different gels can be compared by standardization with a reference system created from the molecular size marker. It is essential to use the same size marker in the gels being compared. Several marker lanes are also essential; with at least three positions at both ends and one in the middle of the gel to allow the program to correct for possible gel smiling using the normalization algorithm.

Fingerprints can be analyzed on “band-based” or “curve-based” analysis. As was observed in this study, band-based analysis required tedious and laborious band assignment and can be subjective. The conclusion is that it is not suitable for large data sets or for complex fingerprinting patterns, such as those generated from PCR fingerprinting. In contrast, curve-based calculation is more robust and it takes into account the full complexity of the genomic fingerprint that is reflected in the densitometric curve, providing higher similarity values without the subjective band-calling step. It is easy to standardize and fully automate. These properties make it attractive in high throughput applications such as epidemiology and ecology. In agreeing with our conclusion, the curve-based analysis is the method of choice for genomic fingerprinting analyses that have been reported by other investigators (25, 77, 144).

#### **4.4.4 Reproducibility and resolution of the long-range ERIC PCR fingerprinting protocol**

With these defined PCR and electrophoresis conditions, genomic fingerprints obtained from the method used here was highly reproducible providing >90% similarity when the same strain was used. The same reproducibility was obtained in other studies using rep-PCR genomic fingerprints, confirming that PCR-based genomic fingerprinting is a highly reproducible method (25, 144, 186). The protocol also has high resolution in revealing the diversity of *V. cholerae* environmental isolates, as well as in differentiating among closely related epidemic clones, O1 classical, O1 El Tor, and O139. The O1 El Tor environmental isolates were found to be clonally diverse and different from clinical isolates. All of the El Tor clinical isolates provided similar fingerprints which were very similar to the O139 fingerprint, with only the 1 kb band missing in the O139 fingerprint.



This similarity in fingerprint indicates a close relationship among the clinical El Tor isolates and the O139 isolates, which is in agreement with the results of other investigators (15, 16, 98, 187). Thus, the long-range ERIC PCR fingerprinting protocol described here provided a rapid, simple method for determining *V. cholerae* clonality with highly sensitive and reliable results.

#### **4.4.5 Estimation of genomic relatedness**

Several studies suggested that genomic fingerprinting methods, such as the repetitive-sequence-based (rep)-PCR and AFLP provide accurate determination of taxonomic and phylogenetic relationships between and among bacteria (31, 88, 91). In a large scale study using *Xanthomonas* strains as a model, Rademaker *et al.* (2000) reported good correlation between DNA-DNA homology and PCR-based genomic fingerprinting using three primers for rep-PCR (ERIC, REP, and BOX primers) and AFLP genomic fingerprinting (143).

The results shown in this study provide evidence supporting that estimation of genomic relatedness of *V. cholerae* by the new genomic fingerprinting method using long-range ERIC PCR is in complete agreement, with high correlation ( $r = 0.7$ ), to those obtained from DNA-DNA hybridization. Estimation of genome relatedness obtained from ERIC-PCR fingerprints is un-bias and independent of probe strain (because for all three probes, the expected slopes fell within the 95% CI of the estimated slope, showing complete agreement to the expected un-bias value). Differences in the functional relationship of the two methods among different probes were expected since organization (i.e., positions of ERIC sequences) and degree of genome divergence of the test strains to each probe strain are different. Although functional relationship of the two methods is

probe specific, all probes provided good agreement between the two methods. The ERIC PCR fingerprinting was also more sensitive, with 5-fold higher precision (less measurement error) i.e., able to differentiate among closely related strains for which DNA hybridization was not able to differentiate. However, ERIC PCR has limitation in that it is suitable for differentiating only closely related strains, i.e. within the same or closely related species, since the correlation value cannot be assessed when more distantly related strains are examined. This limitation was also recognized as problem in DNA-DNA hybridization, since reassociation of DNA with less than 30% homology is not stable. Therefore, that is the lower limit of reliable DNA homology determination.

The genomic fingerprinting using long-range ERIC PCR, as described in this study, is highly reproducible and has high resolution capability for assessing clonality and relatedness among closely related bacteria. The technique requires only the simple equipment that most laboratories already have, i.e. thermal cycler, horizontal gel electrophoresis, a gel documentation system, and a computer. For gel documentation, if a sensitive fluorescence scanner is unavailable, digitized gel images can be obtained by using a digital camera or scanning of the gel picture. The process is rapid and less cumbersome than many other molecular methods. It can be done within two days with less hands-on operation, including approximately eight hours for the PCR step, which can be conveniently done overnight, and a few hours for electrophoresis.

## 4.5 Conclusion

In this study, we report an improved protocol for genomic fingerprinting using long-range ERIC PCR. The optimized protocol takes advantage of the improved long-range Taq polymerase enzyme to produce a highly efficient PCR reaction and defines the gel electrophoresis conditions for high resolution of the fingerprint pattern. For the fingerprint similarity calculation, a curve-based analysis by Pearson's product moment correlation was found to be accurate (i.e., unbiased) and precise (5 times higher precision than RBR measured by DNA-DNA hybridization). Because it involves little manual labor, the estimate is more suitable for high throughput application than band-based analysis. The protocol yielded highly reproducible fingerprints and has the resolution power to discriminate closely related *V. cholerae* O1 El Tor and O139 serogroups. Genomic fingerprinting, using long-range ERIC PCR as described in this study, is quick, simple, less demanding, and provides robust results that reflect genetic relatedness among strains, all accomplished with high throughput. The method can be used to identify clonal identity, relatedness of strains, diversity, and phylogenetic structure, especially for large collections of bacterial isolates, such as is needed in epidemiological and ecological investigations.

## **Chapter 5 Analysis of the population structure of *V. cholerae* in Chesapeake Bay**

### **5.1 Introduction**

*Vibrio cholerae* is recognized as a human pathogen, as well as a natural inhabitant of the aquatic environment. Therefore, its population structure can provide important insight into the origin of the species and public health risk of the environmental strains. The genetic structure of bacterial populations may be explained by using one of three models, clonal, non-clonal, or weakly clonal, depending on the degree of recombination and horizontal gene exchange. A clonal model exhibits a high degree of linkage disequilibrium, an index measured by the level of allele distribution at each genetic locus, indicating the limits of the frequency of recombination and whether the population is predominantly descendent directly from a single clone. In the case of the non-clonal (panmictic) model, the population exhibits linkage equilibrium, indicating that recombination of alleles occurs freely and frequently. However, the two extremes of strictly clonal and non-clonal population structures are rarely found in species of bacteria. Most bacterial populations occupy a middle ground, whereby recombination is highly significant in evolution but not sufficiently frequent to prevent emergence of clonal lineages. More detail discussion about bacterial population models is provided by Spratt and Maiden (176). In a large and diverse population, such as *V. cholerae*, a complex model showing a high degree of recombination in some branches and clonal lineages in others appears to be the best fit (14, 53).

A number of studies have been done to analyze the population structure of *V. cholerae*. However the focus has been on the epidemic clones (54, 69). An extensive study of the genetic diversity and population structure of *V. cholerae* as an entire species was conducted by Beltran *et al.* (14). Using multilocus enzyme electrophoresis (MLEE) analysis of seventeen enzyme loci and a collection of 397 *V. cholerae* isolates, including 143 serogroup reference strains, it was observed that the genetic diversity of *V. cholerae* was high, with the mean diversity of each locus being 0.436, which was smaller than the diversity value of 0.627 for *Salmonella enterica* (164), but larger than the comparable value of 0.343 reported for *Escherichia coli* (163). The genetic structure of the *V. cholerae* population was found to be composed of one major branch with several subclusters and exhibited a degree of linkage equilibrium; and a minor but significant branch composed of a few isolates with high degree of linkage disequilibrium (14).

In this chapter, ERIC-PCR genomic fingerprinting was used to determine relatedness among strains of the census collection of *V. cholerae* from Chesapeake Bay, taking into account the rich background data, including phenotypic and genotypic characteristics, and the environment from which the strains were isolated. The population structure of the species was assessed from the relatedness measurements by cluster analysis, with the support of DNA-DNA hybridization and 16S rRNA sequence analyses. The results of these analyses provided new insight into the structure and evolution of *V. cholerae* by revealing the presence of a multi-layered solid substructure of the species as a whole and the significant divergence of the *V. cholerae* genome at many levels, but with the 16S rRNA sequence relatively fixed.

## 5.2 Materials and Methods

### 5.2.1 Bacterial strains

A subset of bacterial isolates collected during a census survey of *V. cholerae* in Chesapeake Bay was subjected to long-range ERIC PCR fingerprinting (Chapter 4). The non-redundant collection, i.e., collection without artificial clone, of 224 *V. cholerae* isolates from Chesapeake Bay was used to analyze the population structure associated with phenotype, genotype, and habitat and to provide a scaffold for the total population structure. Additional isolates from diverse geographical, as well as clinical, sources were included to provide greater understanding of the structure obtained. The additional isolates comprised 56 environmental isolates and 24 clinical isolates. The environmental isolates were from previous Chesapeake Bay studies carried out in 1976-1979 (14 isolates), Peru (10 isolates), Bangladesh (9 isolates), Italy (4 isolates), Florida, Argentina, Czechoslovakia, Denmark, Thailand, Mexico, India (2 isolates each), and one isolate each from the Philippines, Chile, Brazil, ballast water in a ship arriving to United States, and Germany (previously *V. albensis* type strain ATCC 14547). The clinical isolates included three non-O1 isolates, two from Mexico and one from Florida, *V. cholerae* O37 (1 isolate), and the *V. cholerae* O1 and O139 epidemic strains (20 isolates) from six different countries. A close relative, *V. mimicus*, (10 isolates), was also included in the analysis. A strain of *Vibrio fluvialis* isolated from Chesapeake Bay was included as an outlier. More details, as well as strain names and numbers are listed in Table 4.1 (Chapter 4).

### 5.2.2 Determination of operational taxonomic units (OTUs)

Long-range ERIC PCR described in Chapter 4 was used to produce genomic fingerprints of the isolates. Relatedness among isolates was estimated by the Pearson correlation coefficient ( $r_e$ ) of signal intensity profile (curve-based calculation) of the genomic fingerprint.

To analyze the structure of *V. cholerae* population, using isolates from the census survey, an explicit definition of clonal entity was required because the entire collection of isolates contained clonal redundancies. That is many identical isolates originating from one cell in the sample comprises clonal redundancy. For example, enrichment culture can artificially produce many identical isolates, with the result that a single cell from a given sample will be over-represented, since many replicates of the cell will be present in the analysis of the population structures in nature.

Depending on the method used for isolation, the level of clonal redundancy will vary. In this study, two isolation methods were used: enrichment and direct plating-colony hybridization. In the case of direct plating, each isolate was from a single colony forming unit (CFU) on direct plate culture. Therefore, clonal redundancy by an artificial process was not considered. Each individual isolate from the hybridization plate was considered to correspond to an individual cell in the sample.

In the case of enrichment, individual enrichment flasks were the source of significant artificial clonal redundancy. Therefore, clonal identity among isolates from the same enrichment flask was determined by comparing isolates by phenotype, genotype, and genomic fingerprint profile. Any discrepancy in phenotype or genotype was interpreted as evidence of a different clonal entity. By following this procedure, 54

artificial clones were identified from 25 of the enrichment flasks. Clonal redundancy was eliminated from the census collection by retaining only one randomly-picked representative from each redundant clone to include in the subsequent analyses.

### 5.2.3 Cluster analysis

Cluster analysis of the isolates was performed based on phylogenetic interpretation of relatedness measurements among OTUs. Relatedness among isolates was estimated by the Pearson correlation coefficients ( $r_e$ ) of signal intensity profiles of ERIC-PCR fingerprints. The distance between a pair of isolates ( $d_e$ ) was determined as  $d = 1 - r$ , so that a distance matrix of all possible pairs could be constructed. The individual distance values ranged from 0 to 1, where 0 represented identical isolates and 1 represented completely unrelated isolates.

The distance matrix was used to construct phylogenetic trees for cluster analysis. Since true topology of the clusters is unknown, three different phylogenetic tree algorithms i.e., neighbor-joining (NJ) (157), unweighted pair group method with arithmetic mean (UPGMA) (171), and the Fitch-Margoliash (Fitch) (68) methods were employed and relationships calculated using the computer program PHYLIP version 3.6 (66) to provide confidence of the clustering by consensus among the different methods. To determine the support level of each branch, cophenetic correlation and consensus of a branch occurrence across the three clustering methods were calculated. A branch was determined to be significantly supported when the cophenetic correlation was  $>0.7$  and the branch was consensus to all three trees. The significance of support for a branch by consensus occurrence in all of three trees was estimated as the probability being true topology as  $>0.7$ , according to simulation by Kim *et al.*(103). Taking this number, the



occurrence of two layers of consensus branches (by nesting of a consensus branch under a consensus branch) provided the probability of the two layered clusters being a correct topology  $> (1 - (1 - 0.7)^2) = 0.91$ . With three layers of consensus branches, the probability is larger than 0.97.

## **5.3 Results and discussion**

### **5.3.1 Structure of the *V. cholerae* population in Chesapeake Bay**

#### **A Cluster analysis**

The non-redundant collection of 224 *V. cholerae* isolates from Chesapeake Bay was used to construct a population structure for *V. cholerae* in the Chesapeake Bay. Five additional strains were added to the cluster analysis, one each of *V. cholerae* serogroup O1 classical (RC2), O1 El Tor (RC145), and O139 (RC33), these were included as representative of epidemic clones and the a type strain of *V. mimicus* (RC5) was added as the representative of a closely related species. A *V. fluvialis* isolate from the Chesapeake Bay (RC541) was used as a representative outlier.

To summarize the three trees, the rectangular NJ tree with consensus support is shown in Fig. 5.1. Most of the branches were well supported by cophenetic correlation, with a correlation coefficient  $> 0.7$ . Exceptions were found mostly on those branches where distance among the OTUs was very small, so that the distance measurement was well within the measurement error range. Support of clustering by occurrence of consensus branches from different trees was far more conservative. According to the location of branches relative to the terminal nodes, the clusters of OTUs defined by the

consensus branches could be classified into three groupings; clonal cluster, primary cluster, and intermediate cluster.

A clonal cluster was a set of OTUs under a consensus branch with distance to terminal nodes  $<0.1$ , the upper limit of distance estimation of identical isolates by genomic fingerprinting (reproducibility of long-range ERIC PCR fingerprinting, Chapter 4). When fingerprints were visually examined, significant differences in fingerprint patterns were not detected in the clonal clusters. Clonal clusters consisted of OTUs whose source was different from each other by one or more of the following criteria: site; time; or fraction. This observation indicates that the same clone was often isolated from samples from different sites, collected at different times, and from different types of samples (water or plankton).

A primary cluster was defined as a set of OTUs under the consensus branches directly originating from the presumed root branch (i.e., the top-most branch connecting to the outgroup OTU, RC 541). A significant primary branch is shown in Fig. 5.1, namely branch B which forms Cluster B. Presence of the primary Cluster B implies that there are two distinctive lineages of *V. cholerae* in Chesapeake Bay. One lineage is represented by Cluster B and the other comprised the rest of the population (group A; the term group is used here instead of cluster because the OTUs comprising this group were not a single significant cluster), including the type strain and cholera pandemic strains of the species. Presentation of the tree as an unrooted radial tree (Fig. 5.2) demonstrates separation of at least two of the separate lineages from the root branch.

The intermediate cluster comprised those clusters formed by a significant, i.e., consensus, inner branch that did not belong to either a clonal or a primary cluster. Most

of the branches in the case of the inner branch orders were not consensus by the three clustering methods. This limitation in consensus of clustering is believed to be caused by resolution in the estimation of distances by  $r_e$ . When relatedness is small, the effect of spurious correlation is more profound because the true correlation is absent or close to zero, while randomly occurring, spurious correlation has a greater influence. Therefore, the location of the outgroup OTU (RC541) is believed to contain more error from spurious correlation than the other OTUs. The lack of strong support for the clustering of OTUs in group A by the consensus tree method is a consequence of the difference in location of the outgroup OTU in each of the three trees. This phenomenon was also a factor in most of the inner branches of the tree portion of group A.

A strongly supported inner branch was A1, which defined Cluster A1. Cluster A1 was the largest, having 110 OTUs out of a total of 229 OTUs. The A1 cluster also contained at least two separate lineages: A1 $\alpha$ , comprising RC390, RC502, and RC434, and the rest of the 107 OTUs. In addition to A1, group A contained five smaller intermediate clusters (cluster A2 – 6, Fig. 5. 1). Although the number of OTUs in these clusters was small (6-8 isolates), they consisted of at least three layers of consensus branches, which provided highly significant support. Among them, Cluster A3 was notable because it contained the 6<sup>th</sup> pandemic type strain of *V. cholerae* (RC2). Therefore, six other members of this cluster can be regarded as close relatives of RC2, namely *V. cholerae* O1 classical biotype and the type strain of *V. cholerae*.

From the NJ tree in Fig 5.1, two unstable branches of relatively long length were found. One was reference strain RC5, the type strain of *V. mimicus*. The other was the clonal cluster of RC584, RC585 and RC586. In the UPGMA and Fitch trees, the RC5

branch commonly occurred between branch B and the outgroup branch. Therefore, the odd location of RC5 in the NJ tree can be attributed to the unique principle of the NJ method, which can cause unbalanced topology to meet the principle of minimum evolution. Considering that the other two trees used the principle of least square to find the optimum topology and that RC5 is a type strain of *V. mimicus*, the topology of early branching from the outgroup branch (as shown in UPGMA and Fitch tree) is believed to be the correct location for RC5. However, location of the clonal cluster of RC586 varied across all three trees. In the Fitch tree, it branched close to the clonal cluster of RC345, but its branch length was also the longest of all branches as in the case of the NJ tree, making its topology unbalanced in an otherwise well balanced tree. In the UPGMA tree, an ultrametric tree with least square optimum, the cluster of RC586 is the first branch of outgroup RC541. With balanced branch length, the most plausible position of the RC586 cluster was the early branching position from the root, as the UPGMA tree shows (Fig. 5.2). This conclusion was also supported by the phenotypic and genotypic characteristics and 16S rRNA base sequences, discussed below. For the rest of the discussion, the clonal cluster RC586 is referred to as primary Cluster C.

## **B Characteristics of clusters**

To assess the relationship between the phenetic characteristics of the isolates and their genetic lineages, the distribution of 22 variable traits (16 phenotypic and 6 genotypic traits) was contrasted for each cluster or combination of clusters (Table 5.1 and Table 5.2).

Figure 5.1. (next page) Dendogram showing the overall population structure of the *V. cholerae* Chesapeake Bay isolates using the Neighbor-Joining (NJ) clustering method. The branched NJ tree is shown with *V. fluvialis* (RC541) as outgroup. Level of consensus branch (1-3) among the three clustering methods (NJ, UPGMA, and Fitch) is shown above the branch, and cophenetic correlation is shown below the branch. Cluster A1 and B were reduced; see Fig. 5.3 and Fig.5.4 for details. Explanation of colors and gradient is provided in Fig. 5.3b.

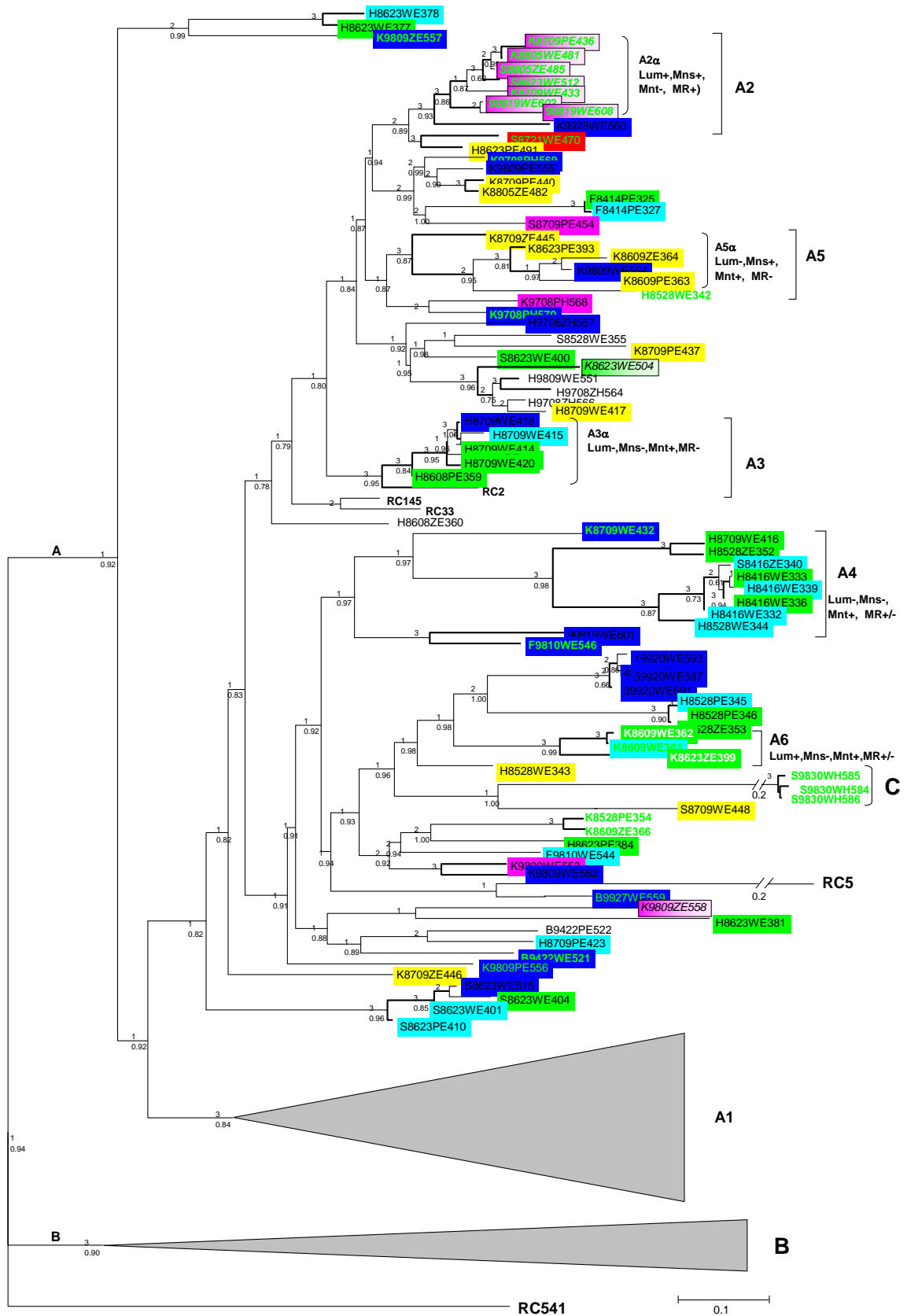


Figure 5.1.

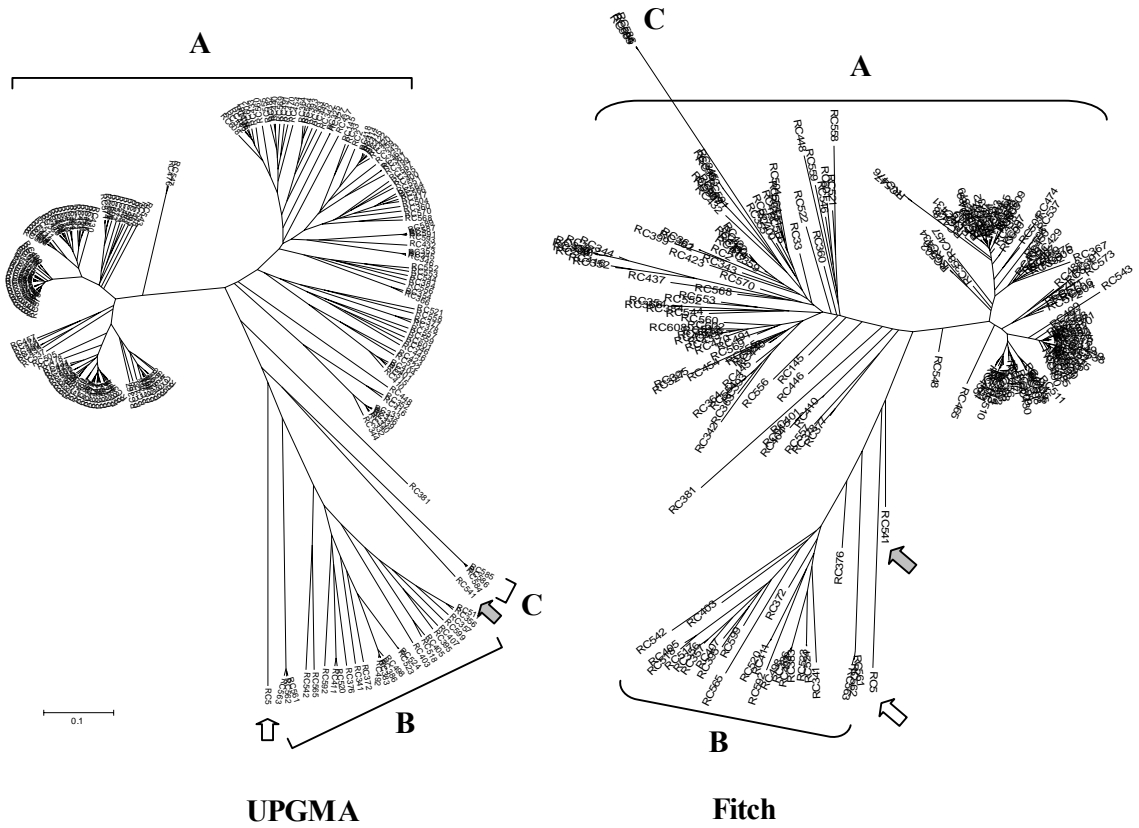


Figure 5.2. Unrooted radial tree of *V. cholerae* Chesapeake Bay isolates obtained using the unweighted pair group method with arithmetic mean (UPGMA) and Fitch-Margoliash (Fitch) clustering. Gray arrow shows the outgroup (RC541) and white arrow *V. mimicus* (RC5).

Table 5.1. Percent positive reactions in the phenotypic tests of strains in each cluster

| Test           | Cluster |      |      |      |      |      |      |             |             | VC <sup>a</sup> | VM <sup>b</sup> |      |
|----------------|---------|------|------|------|------|------|------|-------------|-------------|-----------------|-----------------|------|
|                | A1      | A2   | A3   | A4   | A5   | A6   | Ax   | A           | B           |                 |                 | C    |
| Lum            | 86%     | 88%  | 0%   | 0%   | 17%  | 100% | 14%  | <b>58%</b>  | <b>0%</b>   | <b>100%</b>     | 10%             | 0%   |
| NaCl6%         | 97%     | 88%  | 86%  | 75%  | 100% | 100% | 93%  | <b>94%</b>  | <b>50%</b>  | <b>100%</b>     | 60%             | 54%  |
| Mns            | 70%     | 88%  | 14%  | 0%   | 83%  | 0%   | 41%  | <b>57%</b>  | <b>100%</b> | <b>100%</b>     | 88%             | 100% |
| Mnt            | 30%     | 0%   | 86%  | 100% | 83%  | 100% | 61%  | <b>45%</b>  | <b>100%</b> | <b>100%</b>     | 98%             | 100% |
| Suc            | 100%    | 100% | 100% | 100% | 100% | 100% | 100% | <b>100%</b> | <b>100%</b> | <b>0%</b>       | 100%            | 0%   |
| Orn            | 100%    | 100% | 100% | 100% | 100% | 100% | 96%  | <b>99%</b>  | <b>100%</b> | <b>100%</b>     | 99%             | 98%  |
| Lys            | 98%     | 100% | 100% | 100% | 100% | 100% | 100% | <b>99%</b>  | <b>100%</b> | <b>100%</b>     | 99%             | 98%  |
| MR             | 78%     | 100% | 43%  | 50%  | 33%  | 33%  | 68%  | <b>72%</b>  | <b>100%</b> | <b>100%</b>     | 24%             | 15%  |
| VP             | 99%     | 100% | 86%  | 88%  | 100% | 100% | 96%  | <b>97%</b>  | <b>4%</b>   | <b>0%</b>       | 65%             | 0%   |
| Amy            | 99%     | 88%  | 100% | 100% | 100% | 100% | 89%  | <b>96%</b>  | <b>69%</b>  | <b>0%</b>       | 88%             | 0%   |
| Lip            | 100%    | 100% | 100% | 100% | 100% | 100% | 100% | <b>100%</b> | <b>92%</b>  | <b>100%</b>     | 95%             | 10%  |
| Chi            | 100%    | 100% | 100% | 100% | 100% | 100% | 100% | <b>100%</b> | <b>96%</b>  | <b>100%</b>     | 100%            | 100% |
| V150           | 100%    | 100% | 100% | 100% | 100% | 100% | 98%  | <b>99%</b>  | <b>96%</b>  | <b>100%</b>     | 99%             | 98%  |
| V10            | 100%    | 100% | 100% | 100% | 100% | 100% | 98%  | <b>99%</b>  | <b>96%</b>  | <b>100%</b>     | NA              | NA   |
| PB             | 1%      | 38%  | 57%  | 75%  | 0%   | 67%  | 14%  | <b>12%</b>  | <b>46%</b>  | <b>0%</b>       | 22%             | 87%  |
| 42C            | 99%     | 100% | 100% | 100% | 100% | 100% | 100% | <b>99%</b>  | <b>96%</b>  | <b>0%</b>       | 97%             | 87%  |
| No.of isolates | 110     | 8    | 7    | 8    | 6    | 3    | 56   | <b>198</b>  | <b>26</b>   | <b>3</b>        |                 |      |

<sup>a</sup> VC = *Vibrio cholerae*, <sup>b</sup> VM = *Vibrio mimicus*, average reactions obtained from Davis, *et al.* (1981) and West *et al.* (1986). Values from primary clusters A, B, and C are bolded.



Table 5.2. Genotypic traits of strains in each cluster

| Gene probe      | signal level <sup>a</sup> | Cluster |      |      |      |      |      |     |            |             |             |
|-----------------|---------------------------|---------|------|------|------|------|------|-----|------------|-------------|-------------|
|                 |                           | A1      | A2   | A3   | A4   | A5   | A6   | Ax  | <b>A</b>   | <b>B</b>    | <b>C</b>    |
| <i>luxA</i>     | 0                         | 5%      | 13%  | 100% | 100% | 83%  | 0%   | 80% | <b>36%</b> | <b>100%</b> | <b>0%</b>   |
|                 | 4                         | 95%     | 88%  | 0%   | 0%   | 17%  | 100% | 20% | <b>64%</b> | <b>0%</b>   | <b>100%</b> |
| <i>stn</i>      | 0                         | 66%     | 13%  | 100% | 100% | 100% | 100% | 96% | <b>77%</b> | <b>100%</b> | <b>100%</b> |
|                 | 4                         | 34%     | 88%  | 0%   | 0%   | 0%   | 0%   | 4%  | <b>23%</b> | <b>0%</b>   | <b>0%</b>   |
| <i>ompU</i>     | 0                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 0%  | <b>0%</b>  | <b>0%</b>   | <b>0%</b>   |
|                 | 1                         | 0%      | 0%   | 14%  | 0%   | 0%   | 0%   | 0%  | <b>1%</b>  | <b>4%</b>   | <b>0%</b>   |
|                 | 2                         | 1%      | 13%  | 71%  | 13%  | 0%   | 0%   | 16% | <b>9%</b>  | <b>38%</b>  | <b>33%</b>  |
|                 | 3                         | 13%     | 13%  | 0%   | 88%  | 0%   | 0%   | 25% | <b>18%</b> | <b>31%</b>  | <b>67%</b>  |
|                 | 4                         | 86%     | 75%  | 14%  | 0%   | 100% | 100% | 57% | <b>72%</b> | <b>27%</b>  | <b>0%</b>   |
| <i>ompW</i>     | 0                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 0%  | <b>0%</b>  | <b>0%</b>   | <b>0%</b>   |
|                 | 1                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 0%  | <b>0%</b>  | <b>50%</b>  | <b>0%</b>   |
|                 | 2                         | 1%      | 0%   | 0%   | 0%   | 0%   | 0%   | 2%  | <b>1%</b>  | <b>27%</b>  | <b>100%</b> |
|                 | 3                         | 4%      | 38%  | 14%  | 0%   | 33%  | 0%   | 16% | <b>10%</b> | <b>19%</b>  | <b>0%</b>   |
|                 | 4                         | 95%     | 63%  | 86%  | 100% | 67%  | 100% | 80% | <b>89%</b> | <b>4%</b>   | <b>0%</b>   |
| <i>toxR</i>     | 0                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 0%  | <b>0%</b>  | <b>0%</b>   | <b>100%</b> |
|                 | 1                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 0%  | <b>0%</b>  | <b>62%</b>  | <b>0%</b>   |
|                 | 2                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 0%  | <b>0%</b>  | <b>38%</b>  | <b>0%</b>   |
|                 | 3                         | 5%      | 0%   | 0%   | 0%   | 0%   | 0%   | 2%  | <b>3%</b>  | <b>0%</b>   | <b>0%</b>   |
|                 | 4                         | 95%     | 100% | 100% | 100% | 100% | 100% | 96% | <b>96%</b> | <b>0%</b>   | <b>0%</b>   |
| <i>hlyA</i>     | 0                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 2%  | <b>1%</b>  | <b>4%</b>   | <b>0%</b>   |
|                 | 1                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 0%  | <b>0%</b>  | <b>0%</b>   | <b>100%</b> |
|                 | 2                         | 0%      | 0%   | 0%   | 0%   | 0%   | 0%   | 0%  | <b>0%</b>  | <b>12%</b>  | <b>0%</b>   |
|                 | 3                         | 13%     | 0%   | 0%   | 0%   | 17%  | 0%   | 13% | <b>11%</b> | <b>23%</b>  | <b>0%</b>   |
|                 | 4                         | 87%     | 100% | 100% | 100% | 83%  | 100% | 84% | <b>88%</b> | <b>62%</b>  | <b>0%</b>   |
| No. of isolates |                           | 110     | 8    | 7    | 8    | 6    | 3    | 56  | <b>198</b> | <b>26</b>   | <b>3</b>    |

<sup>a</sup> Signal level, 0 = negative, 1-4 = levels of signal intensity (or homology) to the gene probes, 1 is the lowest and 4 is the strongest signal intensity. Values from primary clusters A, B, and C are bolded.

**Group A** consisted of at least six heterogeneous clusters; however, 5 out of 16 phenotypic traits were completely fixed among the 195 isolates, namely acid from sucrose, lipase production, chitinase production, and susceptibility to two concentrations of the vibriocidal compound O/129. Six other characteristics, growth in 6% NaCl and at 42°C, ornithine and lysine decarboxylase, amylase production and VP test were nearly fixed, i.e., >90% of the isolates gave positive reaction in the tests, reflecting that group A may be a well defined primary cluster with common features among the several intermediate subclusters. Notably, growth in 6% NaCl and amylase production were strongly conserved in group A, with highly variable response in Cluster B. None of the 19 isolates from 16 samples which were negative for growth in 6% NaCl and for amylase was negative for both tests. This implies that the variability observed for the two relatively conserved traits in group A was more or less random and that the small variability which was observed was not due to some systematic relatedness among the isolates. Therefore, the presence of conserved traits in group A that were not conserved in Cluster B strongly supports the hypothesis that group A is indeed a well defined phylogenetic lineage of *V. cholerae*. Phenotypic traits that were not fixed in this group were Lum, Mns, Mnt, MR and PB.

**Cluster A1** is the largest cluster, containing 110 isolates. Yet strong fixation of PB negative (not sensitive to polymyxin B) and moderate fixation of Lum positive were observed. In the case of luminescence, a total of 15 non-luminescent isolates was found. However, 10 of the 15 carried the luminescence structural gene, *luxA*, in their genome. Thus, the luminescent genetic trait was nearly fixed in this cluster, with 95% positive response. The five *luxA* negative isolates in cluster A1 were not supported as members of

the cluster, that is, only by the consensus branch A1 (Fig. 5.3), implying a possibility of inadequate clustering due to spurious correlations of fingerprint intensity curves. It is concluded that strains in Cluster A1 are, in general, luminescent and resistant to polymyxin B.

The three most differentiating of the phenotypic traits were the Mns, Mnt, and MR tests. When the distribution of isolates with a diverse combination of responses to the tests was examined, a strong correlation for combinations with clustering by genome fingerprints was observed, yielding 7 subclusters, A1 $\alpha$  to A1 $\eta$  (Fig. 5.3). The three phenotypic responses were surprisingly strongly fixed in each subcluster, with a few exceptions. Subcluster  $\alpha$ ,  $\zeta$  and  $\eta$  was supported by consensus while the other four subclusters were supported solely by strong correlation of the four phenotypic characteristics. An interesting feature of  $\epsilon$  subcluster was the absence of luminescence, in spite of the presence of its gene, suggesting a unique difference in regulation of luminescence in the epsilon subcluster.

Another important characteristic of the A1 cluster was that 37 of 110 (34%) isolates were positive for heat-stable enterotoxin (*stn*<sup>+</sup>). However, the gene was not evenly distributed among strains of the subcluster. Notably, all of the 28 isolates in the  $\eta$  subcluster were positive for *stn*, while none of the  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\epsilon$  subcluster were positive for the gene.

The habitat of isolates in the A1 cluster was not uniform. All members of the cluster were isolated in June-August from 1998-2000 (Table 5.3). The isolates were from all 3 size fractions, 9 different sampling sites, a wide range of pH (7.3-9.1), temperature (19.5-31°C), and salinity (0-9.5 ppt).

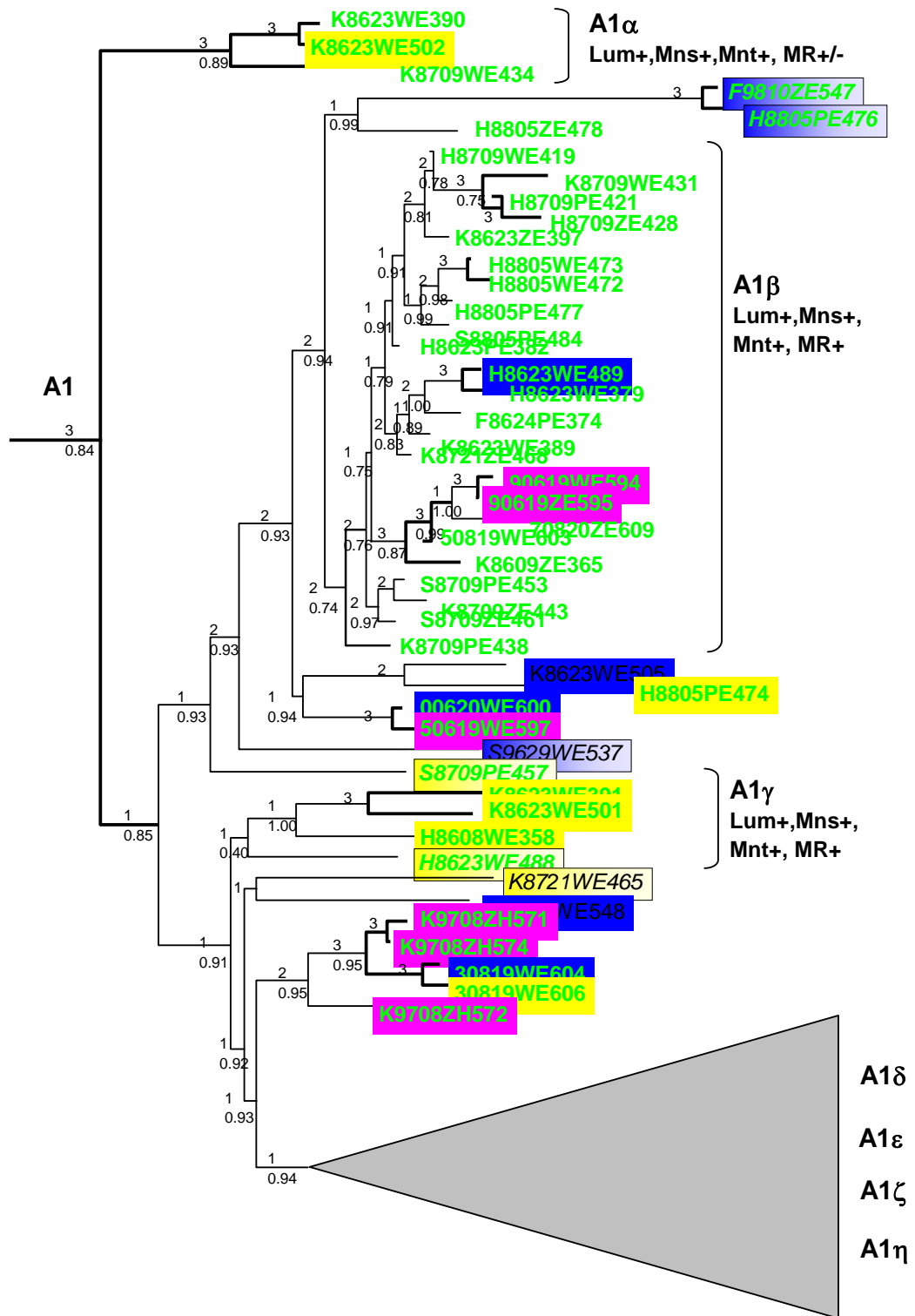


Figure 5.3a. NJ-Dendrogram of Cluster A1, subcluster  $\alpha$ ,  $\beta$ , and  $\gamma$ . See explanation of colors and gradient in Fig. 5.3b.

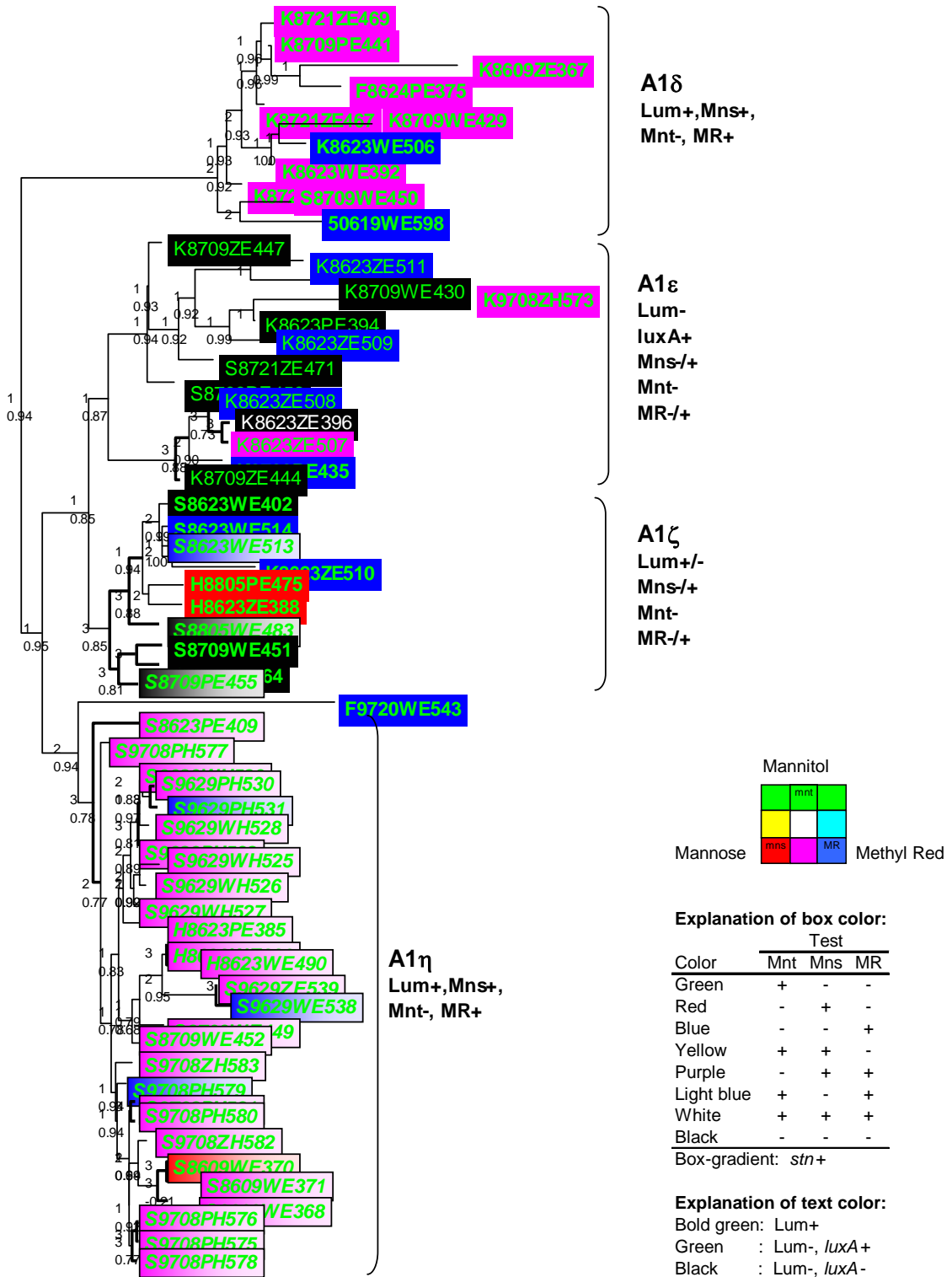


Figure 5.3b. NJ-Dendrogram of Cluster A1, subcluster  $\delta$ ,  $\epsilon$ ,  $\zeta$ , and  $\eta$ . Explanation of colors and gradient is provided at the bottom right corner.

**Cluster A2 and A6:** Except for A1, luminescence was not a common feature among the group A strains. A total of 22 isolates were luminescent (Lum+ and *luxA*+), besides those in the A1 cluster. Seven of the non-A1 luminescent strains formed cluster A2, together with RC560, an exception in the cluster supported by consensus. Three other non-A1 luminescent strains clustered as A6 (Fig. 5. 1). The rest of the non-A1 luminescent isolates were poorly supported as members of a significant cluster and were scattered throughout group A, making their position relative to luminescent cluster A1, A2 and A6.

In Cluster A2, the seven luminescent strains (A2 $\alpha$ ) had the same phenotypic profile, Lum+, Mns+, Mnt- and MR+, in spite of the fact that they were from five different samples collected on different cruises and at different shore sites and from all of the three sample types. Two toxicity related genes, *hlyA* and *stn*, were fixed with high homology with the probes.

**Cluster A3** comprised the A3 $\alpha$  subcluster and RC2, the epidemic type strain of *V. cholerae*. The traits that differentiated RC2 from the A3 $\alpha$  subcluster were high homology with the *ompU* probe (level 4 vs. level 2) and Mns+ and VP-. A3 $\alpha$  strains were isolated only from site H, from water and plankton collected on two separate sampling occasions.

**Cluster A4** consisted of seven strains from site H and one strain from site S. Five were isolated from water, and two were from zooplankton. The unique aspect of the strains was that they were isolated during April and May, 1998, except for RC416. All were isolated from samples with a pH range 7.96 – 7.97 and salinity 5-6 ppt. Traits that

differentiated A4 from other group A clusters were high frequency of isolates sensitive to polymyxin B.

**Cluster A5** consisted of subcluster A5 $\alpha$  and a single luminescent isolate, RC 342. All 5 members of the A5 $\alpha$  subcluster were from a single station (K). With the exception of one strain, RC554, all were isolated from four different plankton samples collected during June-July 1998. RC554 was an exception in that it was isolated from water a year later in August, 1999. Members of the A5 $\alpha$  subcluster were isolated from a wide variation of salinity (5-12 ppt) and water temperature (19.5-25°C).

**Group AX** is a collection of group A isolates that did not cluster with any of the other six clusters. Phylogenetic relationships among the 57 strains were diverse.

**Cluster B:** As shown in Table 5.1, Cluster B isolates had phenotypic profiles that are clearly different from group A. Traits that varied in group A at the subcluster level (Lum, Mns, Mnt and MR) were completely fixed or nearly fixed, while variance in growth in 6% NaCl, Amy, and PB was higher. Among the fixed phenotypic traits, Lum and VP were differentiated group A from Cluster B. The genetic profile of Cluster B is given in Table 5.2.

Traits variable in Cluster B were 6%NaCl, Amy, and PB, and homology to *ompU*, *ompW*, *toxR* and *hlyA*. Figure 5.4 shows seven variables plotted along the NJ tree.

Except for the clonal cluster B2, the response of strains to the three phenotypic tests was different even inside clonal clusters. In most cases, fixation of phenotypic and genotypic traits was not observed at any levels of clustering. Regarding the relatively long branch lengths and high variability of individual traits in Cluster B (Fig. 5.4), the divergence of niches of Cluster B strains was considered to be greater than those of group A member.

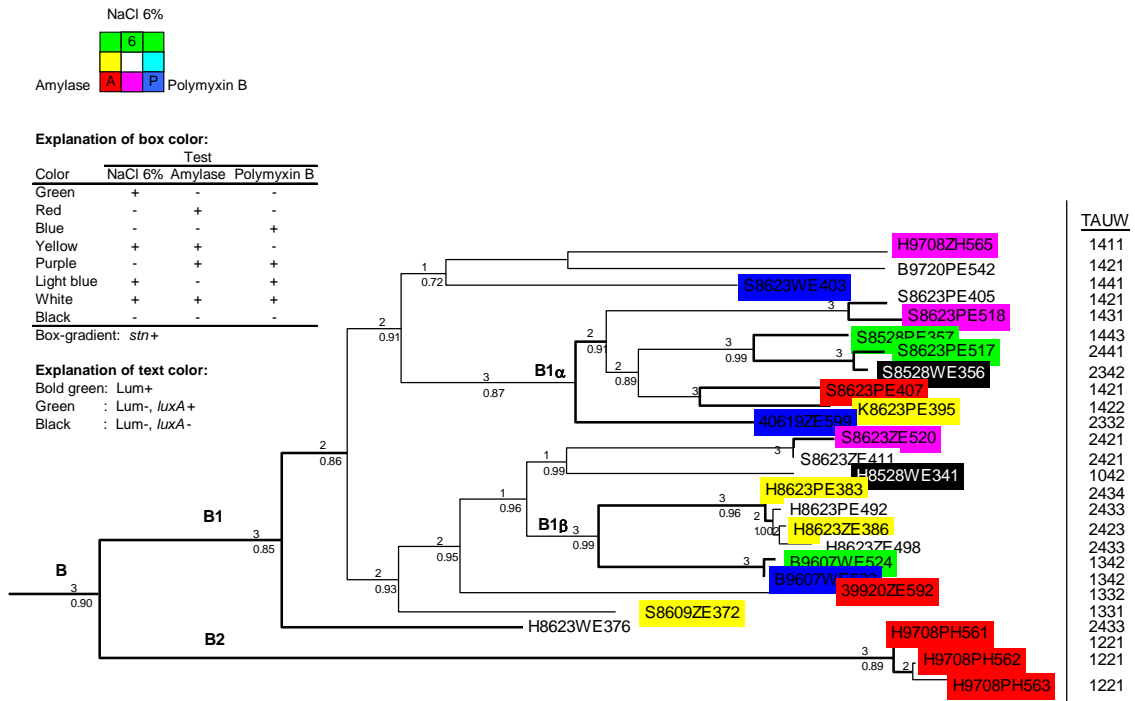


Figure 5.4. NJ-Dendrogram of Cluster B. Explanation of colors and gradient is provided at the top right corner. Noted that no luminescent (green letter) or *stn* positive (gradient) strains in this cluster. The four-digit number on the right lists level of homology for *toxR*, *hlyA*, *ompU* and *ompW* gene probes, respectively.



**Cluster C** comprised the clonal cluster of sucrose negative strains isolated from a single water sample by hybridization. These isolates resemble *V. mimicus*, except that for luminescence, not known to occur in *V. mimicus* (Table 5.1). While gene probe hybridization for *ompU*, *ompW* and *hlyA* genes showed low homology to those genes for the epidemic strain from which the probes were derived, the *toxR* gene homology was completely absent. From its unique characteristic and exceptionally long branch length in the NJ tree, the proper branch location of this cluster was the early branching location as shown in the UPGMA tree (Fig. 5.2).

### **C     Distribution of clusters**

The environment from which each significant cluster was isolated is described in Table 5.3.

In comparison among primary clusters, group A isolates were distributed evenly between free-living (water) and attached to particulates (sum of the two plankton fractions), while Cluster B isolates preferred the particulate fractions. The two primary clusters were also different in their distribution across different sites of sampling. Two notable differences were that Cluster B member was not isolated from F site and low incidence of Cluster B isolates from site K was observed, the site that yielded the most isolates in Group A (Table 5.3). Seasonal dynamics of the incidence rate for Group A was characterized as showing a major peak in late June and early July, and a minor peak in early August. Cluster B strains, although composed of fewer isolates (26 isolates) were obtained throughout the three calendar years, peaking in late June or early July, depending on year of isolation.

In summary, the seasonal dynamics associated with incidence were, in general, similar across all clusters, with the highest peak in late June and early July. An exception was Cluster A4, which was isolated in the spring.

Table 5.3. Frequency of isolates in each cluster by habitat and chronology (time)

|              |                      | Cluster    |           |          |          |          |          |           |            |           |          | Total      |    |
|--------------|----------------------|------------|-----------|----------|----------|----------|----------|-----------|------------|-----------|----------|------------|----|
|              |                      | A1         | A2        | A3       | A4       | A5       | A6       | AX        | A          | B         | C        |            |    |
| Fraction     | Water                | 49         | 7         | 4        | 6        | 2        | 2        | 25        | <b>95</b>  | <b>6</b>  | <b>3</b> | <b>104</b> |    |
|              | small plankton       | 30         | 2         | 2        |          | 2        |          | 17        | <b>53</b>  | <b>12</b> |          | <b>65</b>  |    |
|              | large plankton       | 31         | 1         |          | 2        | 2        | 1        | 10        | <b>47</b>  | <b>8</b>  |          | <b>55</b>  |    |
| Site         | Pelagic <sup>a</sup> | 9          | 2         |          |          |          |          | 5         | <b>16</b>  | <b>2</b>  |          | <b>18</b>  |    |
|              | B                    |            |           |          |          |          |          | 3         | <b>3</b>   | <b>3</b>  |          | <b>6</b>   |    |
|              | F                    | 4          |           |          |          |          |          | 4         | <b>8</b>   |           |          | <b>8</b>   |    |
|              | H                    | 20         | 1         | 6        | 7        | 1        |          | 15        | <b>50</b>  | <b>10</b> |          | <b>60</b>  |    |
|              | K                    | 37         | 4         |          |          | 5        | 3        | 17        | <b>66</b>  | <b>1</b>  |          | <b>67</b>  |    |
|              | S                    | 40         | 3         |          | 1        |          |          | 8         | <b>52</b>  | <b>10</b> | <b>3</b> | <b>65</b>  |    |
| Year         | Round <sup>b</sup>   |            |           |          |          |          |          |           |            |           |          |            |    |
| 1998         | 4                    |            |           |          | 5        |          |          | 2         | 7          |           |          | 7          |    |
|              | 5                    |            |           |          | 2        | 1        |          | 6         | 9          | 3         |          | 12         |    |
|              | 6                    | 6          |           | 1        |          | 2        | 2        | 2         | 13         | 1         |          | 14         |    |
|              | 7                    | 30         | 2         |          |          | 1        | 1        | 10        | 44         | 13        |          | 57         |    |
|              | 8                    | 23         | 2         | 5        | 1        | 1        |          | 8         | 40         |           |          | 40         |    |
|              | 9                    | 6          | 1         |          |          |          |          |           | 7          |           |          | 7          |    |
|              | 10                   | 9          | 2         |          |          |          |          | 1         | 12         |           |          | 12         |    |
|              | 1999                 | 4          |           |          |          |          |          |           | 2          | 2         |          |            | 2  |
|              |                      | 6          |           |          |          |          |          |           |            |           | 2        |            | 2  |
|              |                      | 7          | 11        |          |          |          |          |           |            | 11        |          |            | 11 |
| 8            |                      | 13         |           |          |          |          |          | 6         | 19         | 4         |          | 23         |    |
| 9            |                      | 1          |           |          |          |          |          |           | 1          | 1         |          | 2          |    |
| 10           |                      | 2          |           |          |          | 1        |          | 9         | 12         |           |          | 12         |    |
| 11           |                      |            |           |          |          |          |          |           |            |           | 3        | 3          |    |
| 12           |                      |            | 1         |          |          |          |          | 5         | 6          | 1         |          | 7          |    |
| 2000         | 7                    | 5          |           |          |          |          |          |           | 5          | 1         |          | 6          |    |
|              | 11                   | 4          | 2         |          |          |          |          | 1         | 7          |           |          | 7          |    |
| <b>Total</b> |                      | <b>110</b> | <b>10</b> | <b>6</b> | <b>8</b> | <b>6</b> | <b>3</b> | <b>52</b> | <b>195</b> | <b>26</b> | <b>3</b> | <b>224</b> |    |

<sup>a</sup> Pelagic: summary of all sampling sites from cruises

<sup>b</sup> Round: sampling events corresponding to once every month and twice during summer months (15 rounds/year). Round 1-5 corresponds to January-May; 6 = early June; 7 = late June; 8 = early July; 9 = late July; 10 = early August; 11 = late August; 12-15 = September-December.

### **5.3.2 Structure of *V. cholerae* populations outside Chesapeake Bay**

#### **A Global population structure**

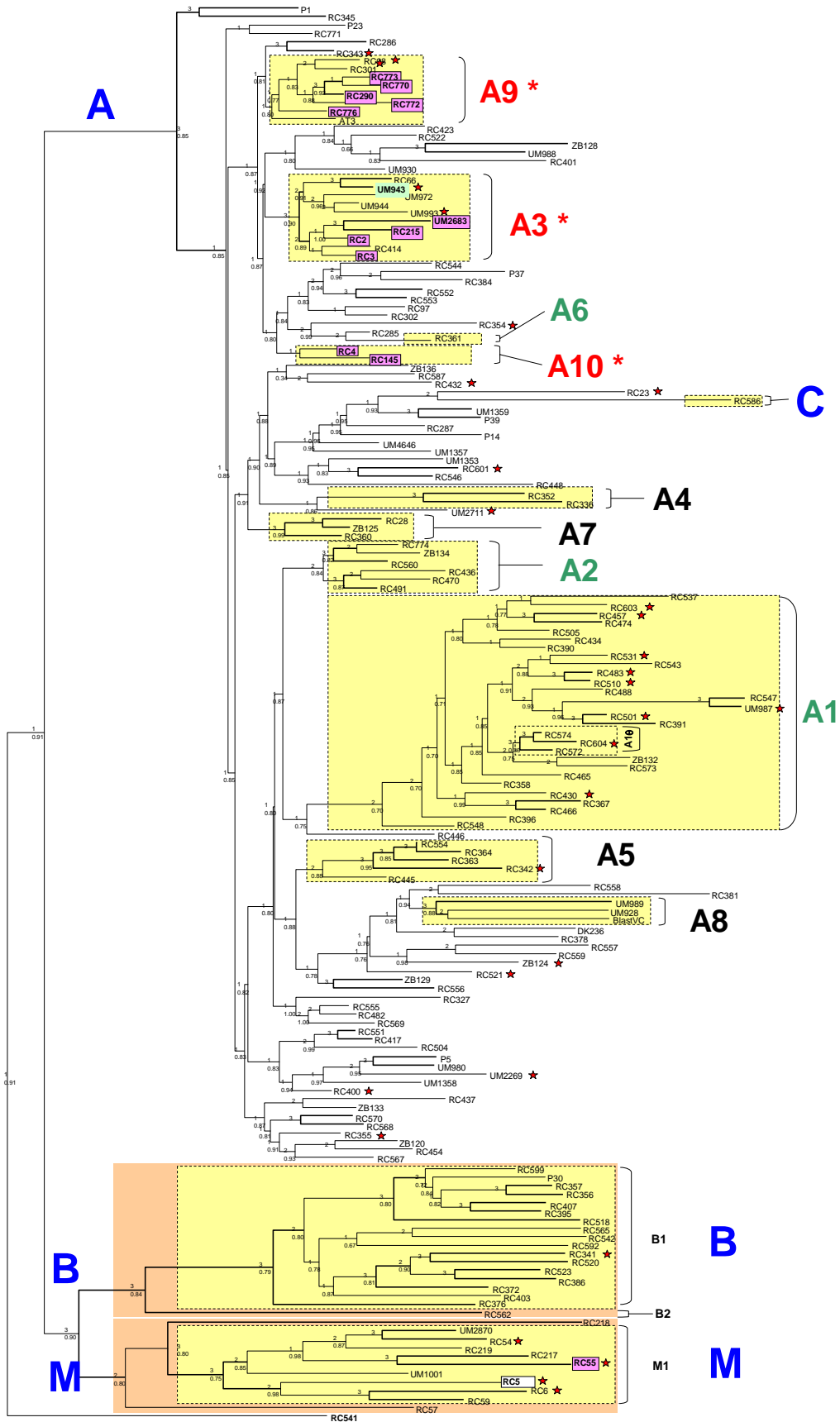
A total of 94 Chesapeake Bay isolates were included in the analysis.

Geographical extension of population structure was achieved by the additional analysis of 39 environmental isolates from 13 countries (Peru, Bangladesh, Italy, Argentina, Czechoslovakia, Denmark, Thailand, Mexico, India, Philippines, Chile, Brazil, , and Germany), one isolate from ballast water in a ship arriving in the United States and two isolates from Florida. Fourteen isolates collected from Chesapeake Bay in 1976-1979 were included to broaden diversity in time. Twenty-four *V. cholerae* clinical isolates from diverse geographies and serogroups were included to identify the location of pathogenic strains in the natural populations. Ten isolates of *V. mimicus* (environmental and clinical isolates) were included to define the relative taxonomic standing of Cluster B, which had characteristics resembling those of *V. mimicus*.

#### **B Cluster Analysis**

In the cluster analysis of the extended collection, the following aspects were considered: (1) concordance of clusters with clusters identified with Chesapeake Bay, (2) determination of new clusters not found in Chesapeake Bay, (3) clustering of *V. mimicus* and its location in relative to *V. cholerae*, (4) clustering of isolates by geographical location of origin, and (5) presence of epidemic lineage. The result of the cluster analysis is shown in Figs. 5.5 and 5.6.

Figure 5.5. (next page) Population structure of the extended set of *V. cholerae* strain collection by Neighbor-Joining clustering. Strain names are highlighted as: toxigenic (CTX+) = pink; the single *V. cholerae* sucrose-negative strain (UM943) = green; and *V. mimicus* type strain (RC5) = white. Clusters associated with toxigenic *V. cholerae* (A3, A9, A10) are red with red astericks, clusters consist of mostly luminescent strains (A1, A2, A6) are green, primary cluster (A, B, C, M) are blue. Red stars mark strains susceptible to CTX $\Phi$  infection (see Chapter 6).



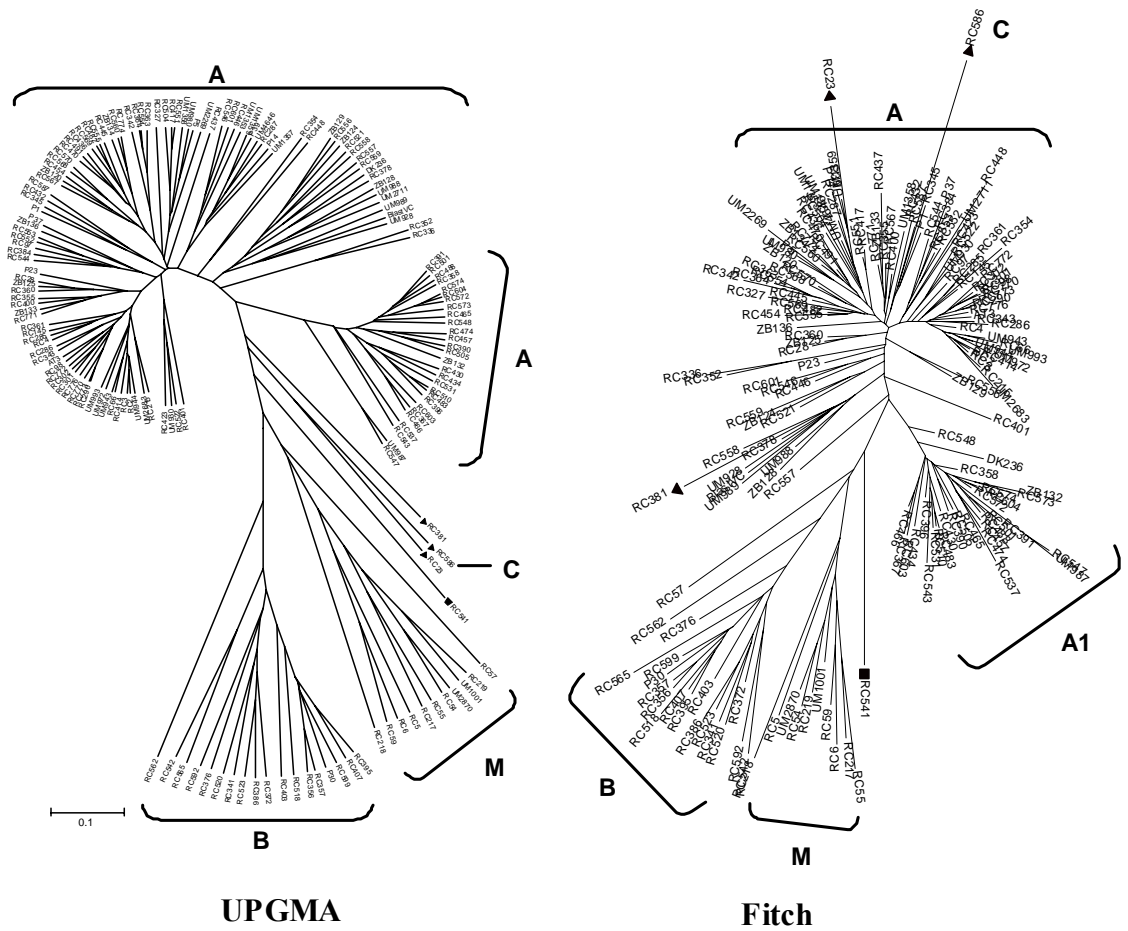


Figure 5.6. Unrooted radial tree of the *V. cholerae* strain collection based on the unweighted pair group method with arithmetic mean (UPGMA) and Fitch-Margoliash (Fitch) clustering. Outgroup (RC541) is indicated by the solid square, and strains with an exceptionally long branch are indicated by a solid triangle.

**Primary Cluster A, B, C and M.** At the level of primary clusters, the structure of clusters A and B was consistent with the outcome of a previous cluster analysis done using the Chesapeake Bay collection of strain. Notable is the support for group A in the previous analysis, well supported by consensus as a significant Cluster A.

**Additional primary cluster, Cluster M,** was defined as a well supported cluster, branching close to Cluster B. When divergences of the members in Cluster B and Cluster M from the majority of *V. cholerae* (Cluster A) were compared, no significant difference was found. Therefore, the two clusters, Cluster B and Cluster M, are concluded to have the same level of relatedness, or distance, to Cluster A.

Most of the strains in Cluster M were identified as *V. mimicus*. Exceptions were two non-O1 environmental isolates from Florida and Chesapeake Bay, UM1001 and UM2870, sucrose negative *V. cholerae*. They were isolated and identified in 1978 and 1981, before *V. mimicus* was proposed as a separate species by Davis *et al.* (46). Using *V. cholerae* specific primers (see Chapter 3) targeted to the intergenic spacer region, the two strains gave positive signal. However, two of ten *V. mimicus* isolates (RC54 and RC 218) were also positive by PCR, indicating that the intergenic sequence was shared by the two species. Additional information from DNA-DNA hybridization with three *V. cholerae* representative probes gave the same response as *V. mimicus*. Therefore, from the phenotypic and genomic composition, it is concluded that UM1001 and UM2870 are *V. mimicus* and that cluster analysis by ERIC-PCR fingerprinting was effective in identifying the phylogenetic position of the strains as *V. mimicus*.

**Intermediate clusters and sub-clusters in Cluster A.** Clusters A1 to A6, intermediate clusters in the cluster analysis, were represented by 1-25 strains from



Chesapeake Bay. Except for Cluster A6, which was represented by a single strain, five intermediate clusters, A1-A5, were detected (Fig. 5.5).

In addition to confirming the A1-A6 clusters, two additional intermediate clusters, A7 and A8, were identified in Cluster A and their characteristics listed below.

Cluster A1, containing luminescent strains, comprised 110 isolates from Chesapeake Bay, was supported by two of the three clustering methods (Fig. 5.6). RC547-UM987 cluster is considered a close relative of the A1 cluster.

Among the strains in Cluster A1, ZB132 was the only strain whose source was outside Chesapeake Bay. ZB132 was isolated from pond water in Bakergonj, Bangladesh. One strain from Chesapeake Bay, isolated in 1977 (UM947), had an ERIC fingerprint identical to RC510, isolated in 1998. Thus, members of Cluster A1 were abundant and persisted in Chesapeake Bay for more than 20 years, and were also present in freshwater of tropical locations.

An additional subcluster (A1 $\theta$ ) was identified close to ZB132. In the cluster analysis, RC571, RC574, RC604 and RC606 formed a consistent cluster (Fig. 5.1), but were not classified as a subcluster because two of the four strains gave a different phenotype in utilization of mannitol and mannose. In Fig. 5.5, RC572, located outside of the four strain cluster in Fig. 5.1, consistently clustered with two representatives of the four strain cluster. RC571, RC572 and RC574 were isolated from the same zooplankton homogenates by hybridization and may be clonal.

In Cluster A2, RC774, *V. cholerae* O1 from an environmental sample of unknown location and ZB134, an isolate from pond water in Bakergonj, Bangladesh, were included in the Chesapeake Bay strain cluster.

Cluster A3, regarded as a cluster of potentially toxigenic strains in the Chesapeake Bay analysis, was supported by addition of five other toxigenic strains, i.e., two clinical isolates of *V. cholerae* O1, Classical biotype (RC3, RC215), and three clinical isolates of the El Tor biotype from Louisiana (UM2681, UM2683, and UM2684). The three Louisiana isolates had the same fingerprint pattern and strain UM2683 is shown in Fig. 5.5. Other strains clustering in the A3 subcluster were non-toxigenic environmental isolates, O14 serogroup (RC66) isolate from Brazil, four Chesapeake Bay isolates of serogroups O1, O6 and O24 isolated by J. Kaper in 1977-1978 (UM943, UM944, UM972, and UM993), and five isolates from the current Chesapeake Bay collection identified earlier (represented by RC414) that included two serogroups, O81 and O149. One of the isolates (UM943) is atypical *V. cholerae* since it does not ferment sucrose. From the analysis, strains closely related to RC2, the type strain and the 6<sup>th</sup> pandemic strain, but not toxigenic, belong to diverse serogroups and the phenotypes are persistent in the coastal waters of Chesapeake Bay and Brazil.

In Cluster A4, A5 and A6, there were no additional strains except strains from the Chesapeake Bay. The additional Cluster A7 is a tight cluster, consisting of three environmental, non-toxigenic isolates from three different geographical locations and belonging to different serogroups [non-O1 from Bangladesh (ZB125), O1 from Brazil (RC28), and O28 from Chesapeake Bay (RC360)]. Cluster A8 was made of two strains from Chesapeake Bay and an environmental isolate from the ballast water of a cargo ship.

The groups of strains designated A9 and A10 in Fig. 5.5 were not supported by consensus. However, close relatedness of these strains was implied by the possession of

cholera toxin gene and other factors. In Group A10, all isolates were toxigenic and clinical, with five of six belonging to the O139 serogroup. RC4 is shown as a representative of the four O139 isolates from various locations that gave identical ERIC fingerprints, i.e., RC12 and RC20 from Thailand, RC30 from Bangladesh, and RC33 from India. The sixth member of the A10 cluster was RC145, the 7<sup>th</sup> pandemic O1 El Tor strain, and endemic in Bangladesh.

Group A9 consisted of twelve strains, nine of which were toxigenic *V. cholerae* O1. RC773 is shown in Fig. 5.5 as the representative of the four El Tor biotypes (RC105, RC11, RC775, and RC90) isolated from clinical cases in three different countries (India, Chile, and Peru). Four other toxigenic strains in this group were an O1 El Tor from Louisiana (RC 290), two O1 Classical biotypes, and one O37 serogroup. Among the three non-toxigenic isolates of the A9 group, RC301 is an environmental *V. cholerae* O1 from Mexico, RC98 is a clinical isolate belonging to serogroup O5 and from Mexico, and AT3 is the previous type strain of *V. albensis*, now included in *V. cholerae*. Although RC301 and RC98 were not toxigenic, i.e. CTX negative, their serotype group and source O1, and from a clinical case, were similar to the toxigenic strains, considering this group to be a toxigenic lineage.

**Intermediate clusters and sub-clusters in Cluster B and M.** In Fig. 5.5, intermediate and subclusters were determined, according to branching order within Cluster B and M, due to discrete internal clustering structures in the two primary clusters. Cluster B intermediate and subclusters were consistent with those of the Chesapeake Bay collection. P30, an isolate from water off the coast of Peru, was included in the cluster. Consistent with the characteristics of the main cluster in lacking the gene with high

homology to *toxR* of *V. cholerae*, PCR amplification specific for the *toxR* gene also failed to yield positive results. Although a minority of the *V. cholerae* population, compared with the number of isolates in Cluster A, strains in Cluster B are present in Chesapeake Bay, and in at least one other geographical location. There is the possibility that members of Cluster B persist as a minority population in other locations, since there are reports of *toxR* negative strains from diverse geographic locations (161).

### **C Inference of the population structure**

In summary, the population of *V. cholerae* in the extended collection was structured as three levels of differentiation: (1) primary clustering which separated three major phylogenetic lineages; (2) intermediate clustering, which differentiated populations closely related, with relatively small genome deviations; and (3) subclusters which differentiated intermediate clusters with very little genomic or phenetic deviation.

The difference in branch length in Cluster B and Cluster M were in contrast to Cluster A, supporting the primary clustering and contributing a different evolutionary path for each of the lineages.

In the subcluster, fixation of variable phenotypic characteristics was observed for subclusters in Cluster A (Fig. 5.1). This can be interpreted as the genome having relatively small deviations among the strains, with strong fixation of gene expression. Several subclusters consisted of strains from different geographical locations, as well as different time of isolation more than 20 years apart. The same clone of *V. cholerae* was distributed globally, which implies a niche-specific selective pressure for persistence of specific clones.

The intermediate clustering, to a certain degree, reflects differences based on the observation that traits fixed in intermediate clusters are life style specific. For example, the majority of strains in Cluster A3, and the groups A9 and A10, were linked to epidemics of cholera. No other strain in any of the other clusters was related to cholera cases. Another example is that A1 and A2 clusters consisted mostly of luminescent strains, while other intermediate clusters were not. Presence of multiple intermediate clusters was evident in the A1 cluster of luminescent strains. It is interpreted as the micro-habitat differentiating strains into different intermediate clusters.

### **5.3.3 Inference of population structure from DNA-DNA hybridization**

For strains in the extended collection, DNA-DNA hybridization (DDH) was performed with three probe strains i.e., RC466, RC145 and RC395, representing Cluster A1, Cluster A10, and Cluster B, respectively (see Chapter 4). The resulting relative binding ratio (RBR) values were used to elucidate the population structure of the strains in the extended collection and to determine concordance of the inferred structure with that of ERIC fingerprinting.

Fig. 5.7 illustrates the distribution of strains according to relative relatedness to probe strains representing Cluster A on the X-axis and Cluster B on the Y-axis. Each strain was examined for phylogenetic standing by similarity of genomic DNA with that of the probe strains. The primary clusters, elucidated by ERIC fingerprinting were almost completely identical to the DDH clusters. In the upper panel of Fig. 5.7, the probe strain of Cluster A is represented by RC145, the reference strain of the 7<sup>th</sup> pandemic. Members assigned to Cluster A formed an elliptical cluster and were completely separate from strains of the other clusters. Although the border between Cluster B and Cluster M

was not clear (upper panel), the lower panel where another probe of Cluster A, RC466, representing the luminescent cluster A1, was used, showed clear separation of B and M clusters. Thus the primary clusters based on ERIC fingerprinting were confirmed by genomic relatedness.

The intermediate clusters and subclusters described above were examined by plotting strains according to RBR values to the most closely related probe strains. When Cluster A1 to A10 were plotted against two probe strains of Cluster A1 and Cluster A10, respectively (Fig. 5.8), the strains were distributed widely along both axes, except for those clusters with less than four strains. The range of RBR within the intermediate clusters was more than 50% and the magnitude was comparable to RBR of the primary cluster. Subclusters with two to six strains are shown in Fig. 5.9. The range of strains belonging to subcluster A1 $\gamma$ , A1 $\theta$ , B1 $\alpha$  and B1 $\beta$  was as large as that of the primary or intermediate clusters. Thus, concordance of the population structure by ERIC fingerprinting and DDH was not observed below the level of the primary clusters. The failure to detect phylogenetic sub-structures in genome-genome similarity among closely related strains can be due to two factors: (1) less precision of the DDH method or (2) very high plasticity of the bacterial genomes. The former was addressed previously in Chapter 4. From values shown in Table 4.2 (Chapter 4), RBR from DDH has a higher measurement error than ERIC fingerprinting (5 times in  $s_e/s_A$  ratio). However, the high measurement error can not account for the wide range of RBR (>50%) because the range covered by random error alone (i.e., LSD value of RBR) is as small as 2.25% (Table 4.2). In fact, the coefficient of intraclass correlation ( $r_I$ ) revealed that the genome divergences is more responsible for the RBR range (i.e., the range by 174 strains

analyzed in Table 4.2) than the random measurement error, with the ratio 70:30. Therefore, the wide range of RBR in Figs. 5.8 and 5.9 cannot be the result of low precision of the method alone. The high plasticity of the bacterial genome is believed to be the cause of the low genome similarity between strains of the intermediate and subclusters. Horizontal transfer of non-essential genes, which contain less phylogenetic information, may be the source of this plasticity. Supporting this conclusion is the occurrence of various serotypes in clusters A3, A3 $\alpha$ , A9, and A10. Loss of genes for LPS polysaccharide synthesis, which can be as large as 50 kb in the case of O139 LPS, and gaining a new analogous gene can result in similarity change of 100 kb (179).

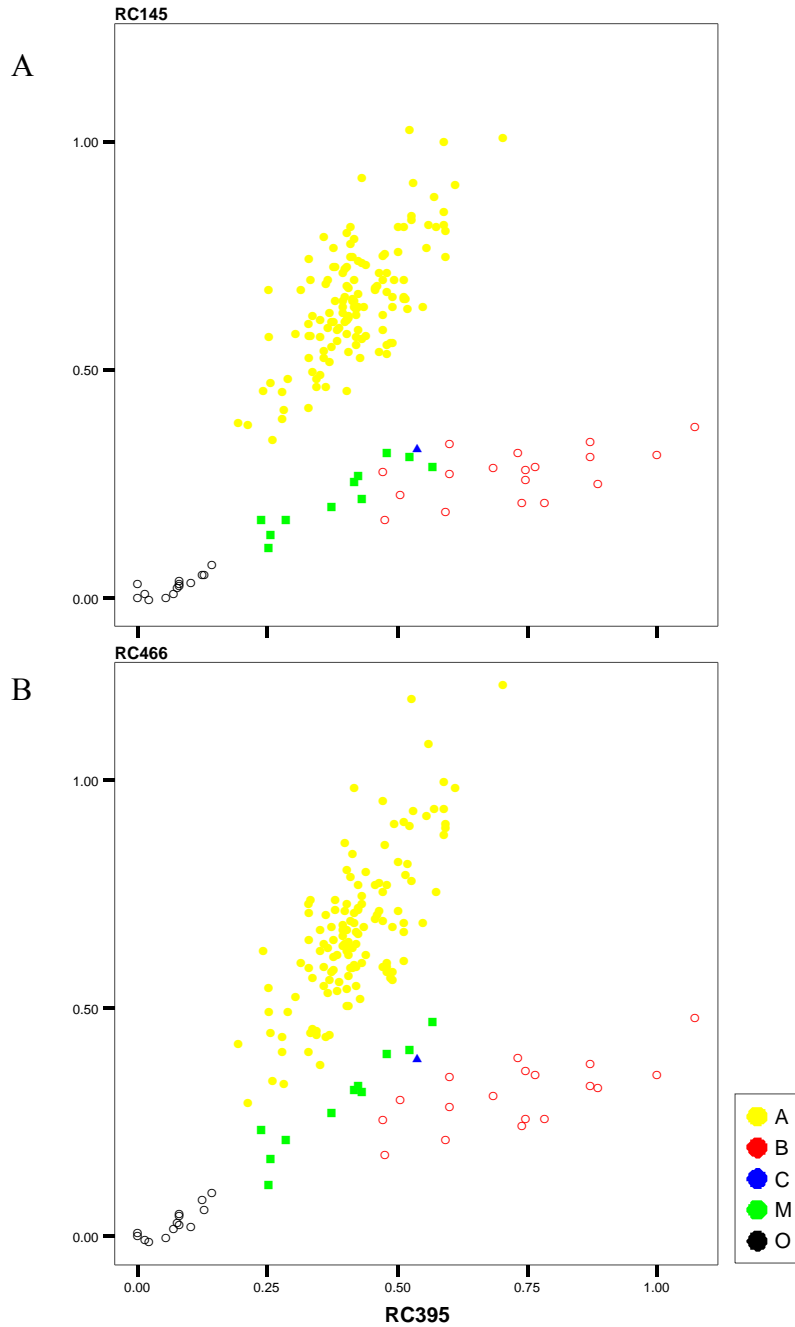


Figure 5.7. Distribution of strains by DNA-DNA hybridization. Relative binding ratio (RBR) obtained from two different probes were plotted against each other. RBR of strains to probe RC395 are plotted on the x-axis. RBR of strains to probe RC145 and RC466 are plotted on the y-axis of panel A and B, respectively. A, B, C, and M represent strains in primary Cluster A, B, C, and M respectively; O represents outgroup strains.



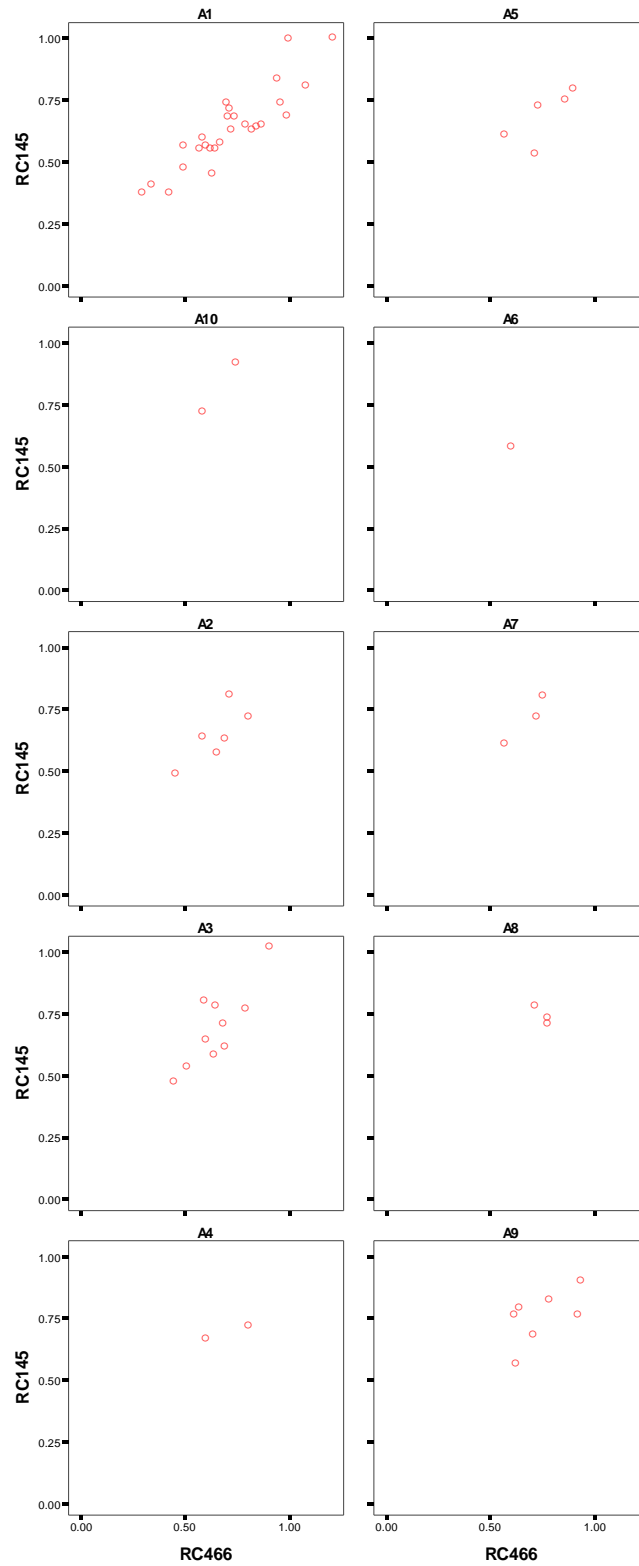


Figure 5.8. Distribution of RBR values for strains in intermediate clusters against two Cluster A probes (RC466 and RC145).

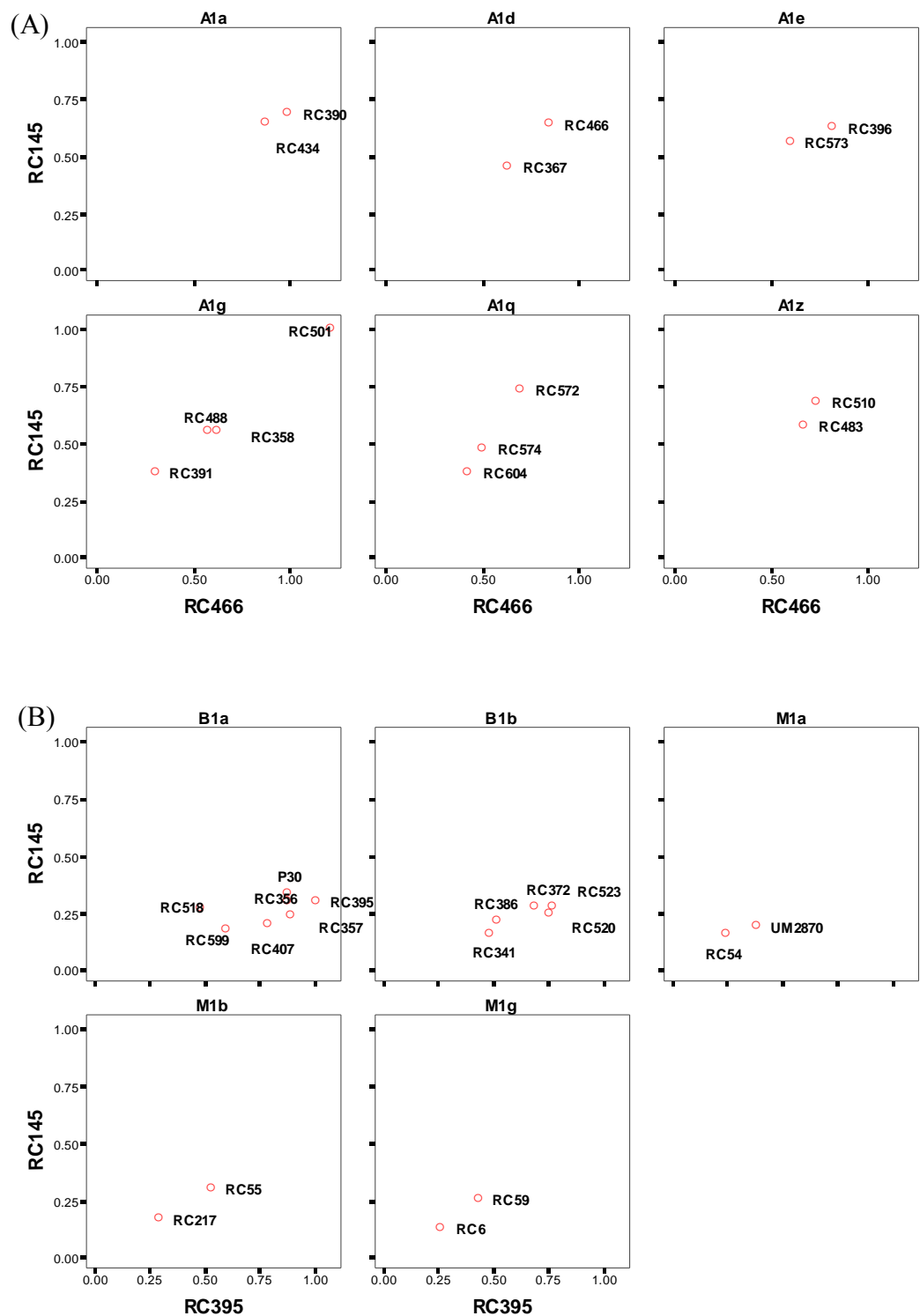


Figure 5.9. Distribution of RBR values for strains in subclusters against probes RC145 and RC466 (panel A) or RC145 and RC395 (panel B).

#### 5.3.4 Identification of Clusters B and C and *V. mimicus*

Strains of Clusters B and C possessed many characteristics that deviated from *V. cholerae* and some of the characteristic resembled those of *V. mimicus*. The results from both the clustering by ERIC fingerprinting (Fig. 5.5 and Fig. 5.6) and DDH (Fig. 5.7) showed that, while the position of Cluster C was uncertain, Cluster B had a genomic distance from Cluster A that was similar to that of *V. mimicus* (Cluster M) to Cluster A. Therefore, these clusters are concluded to be either new species or new branches of *V. cholerae* or *V. mimicus*.

Cluster B phenotypic characteristics show a hybrid-like profile (Table 5.1). Lack of luminescence and % positive reaction for the VP test are similar to *V. mimicus*. According to the sucrose test, the key feature used to discriminate *V. cholerae* and *V. mimicus*, Cluster B strains should be classified as *V. cholerae*. Thus, taken together with growth on TCBS and 0% NaCl broth, Cluster B strains are concluded to be *V. cholerae*, following the traditional phenotypic identification.

RBR of DDH and the 16S rRNA sequence are commonly used to identify a species. DDH results shown in Fig. 5.7 suggest Cluster B has low (<50%) DNA homology to both of the probes of Cluster A, with a mean of  $28 \pm 7$  % to RC145, and  $32 \pm 8$  % to RC466 probes (Table 5.4). Similar low levels of homology were observed for *V. mimicus* with  $22 \pm 7$  % and  $30 \pm 11$  % homology to probe RC145 and RC466, respectively. This low level of genomic homology casts doubt on including Cluster B within the species *V. cholerae*. RBR values of Cluster B probe RC395 with Cluster M was used to assess relationship between the probe strain to *V. mimicus*. The results showed maximum value of 57% to RC219, a *V. mimicus* environmental isolate from

Bangladesh. However, mean homology of the probe to *V. mimicus* was only  $39 \pm 11$  % (Table 5.4). Therefore, if the traditionally accepted value of ca. 70% or greater DNA relatedness for a species determination is used (189), Cluster B would be a closely related, but separate species from *V. cholerae* and *V. mimicus*. However, as suggested by the Ad Hoc committee on Reconciliation of Approaches to Bacterial Systematics (189), phenotypic characteristics should be taken into consideration and these provide agreement in identification.

Table 5.4. Percentage Relative Binding Ratios (RBR) from DNA-DNA Hybridizations among the primary clusters

| Cluster | Probe strain |       |            |       |           |       |
|---------|--------------|-------|------------|-------|-----------|-------|
|         | RC145 (A10)  |       | RC466 (A1) |       | RC395 (B) |       |
|         | Mean         | SD    | Mean       | SD    | Mean      | SD    |
| A       | 65.72        | 14.06 | 67.43      | 17.00 | 42.05     | 9.05  |
| B       | 27.89        | 6.76  | 31.69      | 7.74  | 72.61     | 16.41 |
| C       | 32.86        | ND    | 39.25      | ND    | 53.78     | ND    |
| M       | 22.20        | 7.04  | 29.55      | 10.78 | 38.62     | 11.45 |

Results of 16S rDNA analyses are shown in Fig. 5.10. Representative strains from the three clusters (A, B and M) showed little divergence (Fig. 5.10), except for one Cluster B strain, RC518, and Cluster C, RC586, which had 16S rDNA sequences with unique divergence in the 41-bp region at position 993 (Fig. 5.11). The exact sequence of the 41 bp region was also found in the 16S rRNA genes of *V. gazogens*, *V. coralliilyticus*, *V. neptunius*, and *Brenneria nigrifluens*, and similar copies are also found in *Actinobacillus* and *Colwellia* spp. Excluding this region, 16S rDNA of strains from clusters A, B, C, and M showed only 0-6 base differences for the 1494-bp sequence (>99.99% homology). Therefore, the 993-1033 bp region of RC518 and RC586 can be regarded as a chimera of the two 16S rRNA genes.

With the exception of the chimera RC518-RC586 and the “d” copy of 16S rDNA sequence from the RC145 database (N16961; full sequence (78)), when the full length of the 1535 bp 16S rDNA was compared for the clusters, the number of different bases ranged from 0-13, with a mean of 3.8. These values indicate <1% divergence in sequence. Bootstrap test results also were in agreement that these copies of the 16S genes cannot be discriminated. Therefore, with the exception of the chimera 16S found in the 993 bp region in the single representative of Cluster C (RC586) and one strain in Cluster B (RC518), the 16S rDNA sequences of Cluster A, B, *V. cholerae*, and *V. mimicus* are concluded to be virtually identical. The results suggest that, based on 16S rDNA sequence, Clusters A, B and M are *V. cholerae*.

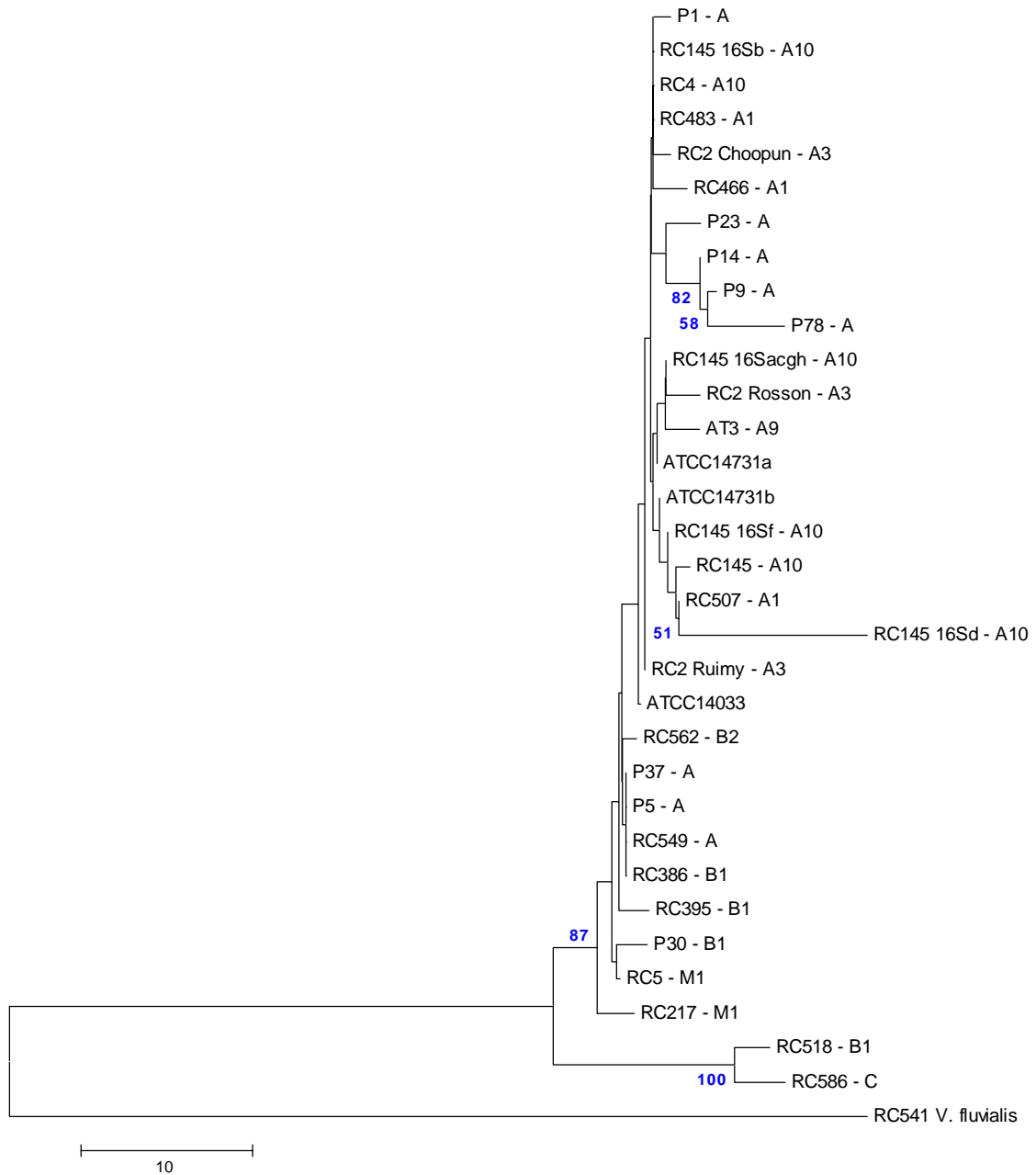


Figure 5.10. NJ tree based on the number of differences in the 16S rDNA sequences for *V. cholerae* and *V. mimicus* (scale bar = number of base differences; cluster name to which the strain belongs is given after the hyphen; the numbers next to branches are bootstrap support (support values less than 50 are not shown)).

| Base No.                     | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   | 21   | 22   | 23   | 24   | 25   | 26   | 27   | 28   | 29   | 30   | 31   | 32   | 33   | 34   | 35   | 36   | 37   | 38   | 39   | 40   | 41   | No. of difference<br>in other regions |   |
|------------------------------|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---------------------------------------|---|
| RC2 - A3                     | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T                                     | 2 |
| RC4 - A10                    | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T                                     | 1 |
| P1 - A                       | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T                                     | 2 |
| RC466 - A1                   | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 1                                     |   |
| RC549 - A                    | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 3                                     |   |
| RC386 - B1                   | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 3                                     |   |
| RC518 - B1                   | T   | C   | A   | G   | A   | A   | A   | A    | G    | A    | A    | C    | T    | T    | G    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 5                                     |   |
| RC586 - C                    | T   | C   | A   | G   | A   | A   | A   | A    | G    | A    | A    | C    | T    | T    | G    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 6                                     |   |
| P30 - B1                     | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 5                                     |   |
| RC562 - B2                   | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 4                                     |   |
| RC395 - B1                   | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 5                                     |   |
| AT3 - A9                     | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 1                                     |   |
| RC145 16Sa - A10             | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 0                                     |   |
| ATCC14731a                   | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 2                                     |   |
| ATCC14033                    | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 4                                     |   |
| <i>V. mimicus</i> RC217 - M1 | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 6                                     |   |
| <i>V. mimicus</i> RC5 - M1   | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 4                                     |   |
| <i>V. fluvialis</i> RC541    | A   | G   | A   | G   | A   | A   | A   | A    | T    | T    | A    | G    | G    | G    | A    | A    | A    | A    | A    | C    | C    | T    | T    | A    | A    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | T    | 81                                    |   |
| Base No. in RC145 16Sa       | 993 | 994 | 995 | 996 | 997 | 998 | 999 | 1000 | 1001 | 1002 | 1003 | 1004 | 1005 | 1006 | 1007 | 1008 | 1009 | 1010 | 1011 | 1012 | 1013 | 1014 | 1015 | 1016 | 1017 | 1018 | 1019 | 1020 | 1021 | 1022 | 1023 | 1024 | 1025 | 1026 | 1027 | 1028 | 1029 | 1030 | 1031 | 1032 | 1033 |                                       |   |

Figure 5.11. The region of 16S rDNA sequences in *V. cholerae* and *V. mimicus* where RC518 and RC586 showed significant deviation. The last column indicates number of base differences in each sequence relative to RC145 16Sa (1494 bp fragments from 1535 bp) after removing the 993-1033 region.



For the single clone in Cluster C, the relative position from phylogeny by the ERIC fingerprint is not clear, but it is best explained as a separate primary cluster, as mentioned previously. By phenotypic characteristics, the profile is compatible with that of *V. mimicus* (Table 5.1), except that it is luminescent. The 16S rDNA sequence show chimerism with the 16S rDNA of other species, however, this is not unique to the cluster since one of the Cluster B members also possesses a similar chimeric 16S rDNA. From the 16S rDNA sequence and DDH, the cluster is concluded to be closely related to Clusters B and M. Therefore, they are the same species as Cluster A, B, and M.

### **5.3.5 Reassessment of the relationship between *V. mimicus* and *V. cholerae***

Separation of previously atypical *V. cholerae* as a different species, named *V. mimicus* by Davis *et al*, 1981 was based, for the most part, on DNA relatedness value measured against a limited number of isolates, i.e., two epidemic O1 *V. cholerae*, ATCC 14033, ATCC 14035<sup>T</sup>, and six strains of *V. mimicus* (46). A few biochemical tests were suggested to differentiate the two species, such as the Voges-Proskauer, lipase, and polymyxin B tests. However, many exceptions among the isolates were recognized. With the accumulated data and more diverse isolates found in the natural populations, separation into the two species differing only in a few tests becomes problematic. The main phenotypic characteristics used to separate the two species, sucrose fermentation, is the last to break down when sucrose negative variants of *V. cholerae* were reported. One sucrose negative *V. cholerae* (UM943) was also identified in this study, supporting the conclusion that the exception is greater than previously recognized.

Attempts to find molecular tools to differentiate the two species proved unsuccessful. The *V. cholerae* specific primer, designed specifically for this purpose

(28), from the highly variable parts of the bacterial genome i.e., intergenic spacer region, served very well as a tool to identify *V. cholerae*, but occasionally provided cross-reactivity with *V. mimicus*. As mentioned above, at least four *V. mimicus* in Cluster M were positive for the PCR primers.

The two species share large numbers of characteristics. For examples, most of the *V. mimicus* are typable with *V. cholerae* antisera (46, 170); several genes with identical sequences have been found in both species, including several toxin-related genes (18, 102, 169, 205). They share the same habitats, low requirement for NaCl for growth. *V. mimicus* also is considered to be a human pathogen and have been isolated from human diarrheal stools and ear infections.

Therefore, besides its low genomic relatedness, the accumulated evidence to date fails to support *V. mimicus* as a separate species. Since both *V. cholerae* and *V. mimicus* were shown to have a high level of horizontal gene transfer and have the gene capture system (29, 78, 117) VCR (Vibrio cholerae repeat) that can capture genes from the environment without being limited by relatedness in the sequence of the alien genes, the low homology in the DNA-DNA hybridization value observed among members of the two species may be the result of interference of these foreign DNAs to otherwise similar genomes. Therefore, considering all of the available information, separate species for *V. mimicus* is not supported.

## 5.4 Conclusion

In this study, the population structure of *V. cholerae* in Chesapeake Bay, a typical temperate estuarine environment, was revealed using the census culture collection and an additional culture collection that extended the temporal and geographical coverage.

The expanded coverage revealed that the population can be divided clearly into at least three primary clusters, Cluster A, containing the majority of *V. cholerae*, including epidemic strains and environmental luminescent strain, Cluster B, containing fewer isolates, but all of which have a low *toxR* gene similarity and some phenotypic traits similar to *V. mimicus*, and Cluster M, containing exclusively *V. mimicus* isolates. More primary clusters may exist, as shown by Cluster C which consists of clonal strains with various atypical characteristics, i.e., luminescence and negative sucrose fermentation. Under Cluster A, sub-structures, named intermediate clusters and subclusters in cascades, were observed. The intermediate clusters had key features, such as luminescence, NAG-ST, or CTX genes. Among strains in a subcluster, phenotypic traits, which are presumed to be most variable, could be nearly fixed, such as in the subclusters of Cluster A1. In those cases, subcluster-level clustering can represent naturally occurring clonal complexes. Therefore, the population comprised multi-layered (multi-nested) clonal subpopulations, each of which shared certain commonly inherited genetic components.

The strongest determinant of genome content of a strain was, therefore, the inheritance of genes that differentiated the primary clusters. However, the variability in genome content, the magnitude of which was comparable to the primary clusters, was observed within intermediate and subclusters. Therefore, it is concluded that a significant proportion of the *V. cholerae* genome is highly plastic, changing freely without affecting

the genetic content and functions that maintain clonality at the intermediate and subcluster levels.

With the accumulated evidence and exceptionally high homology of 16S rDNA sequences the four primary clusters, including Cluster M (the *V. mimicus* cluster), should be classified as the same species, *V. cholerae*. The low DNA-DNA relatedness among *V. mimicus* and *V. cholerae*, the only trait separating the two species, is believed to arise from high genome plasticity and a significant ratio of foreign DNA captured during evolution. Therefore, the population structure of the species *V. cholerae*, assessed in this study is shown to be highly diverse, composed of at least three major branches, A, B, and M, with greater variation in phenotype, including sucrose negative *V. mimicus*, which is proposed as a subpopulation of the highly diverse over-all *V. cholerae* population.

## Chapter 6 The pathogenic risk of non-toxigenic *V. cholerae*

### 6.1 Introduction

Cholera is a severe watery diarrheal disease affecting populations worldwide especially where drinking water and sewage treatment are lacking and public health is compromised. The bacterial agent of epidemic cholera is a subpopulation of *Vibrio cholerae* with the capability of colonizing the human intestine and producing cholera toxin. Historically, only *V. cholerae* serogroup O1 cause epidemic cholera. However, the emergence of epidemic serogroup O139 of *V. cholerae* first appearing in the Bay of Bengal in 1992 (26), the localized outbreaks of cholera caused by serogroup O5 in Czechoslovakia in 1965 (1) and by serogroup O37 in Sudan in 1968 (95), and the sporadic cholera cases caused by several other serogroups have served to emphasize the population dynamics of the bacterial agent and the importance of surveillance in assessing the pathogenic risk of *V. cholerae* as a species.

The emergence of toxigenic *V. cholerae* non-O1 was demonstrated to result from the replacement of the O1-antigen genes with those of other O-antigenic groups (109, 126, 179, 180). Besides changing the O-antigen determinant of toxigenic *V. cholerae*, the major virulence factors also being mobile may explain the appearance of new toxigenic strains deriving from non-toxigenic strains. Genes coding for major virulence factors, i.e., colonization of host intestine mediated by toxin coregulated pilli (TCP) and production of cholera toxin (CT), a potent enterotoxin responsible for cholera

pathogenesis, have been shown to be parts of two different filamentous bacteriophages, VPIΦ (99) and CTXΦ (188), respectively.

Although the status of VPIΦ as a self-transmissible phage is very much under debate (64), O'Shea and Boyd (133) showed that the VPI element can have other means of mobilization, namely generalized transduction via phage CP-T1.

On the other hand, the CTXΦ has been proven to be active and is able to infect several other strains of *V. cholerae*, as well as the closely related *V. mimicus* (59, 61, 62, 188). All infectious CTXΦ have been produced by El Tor strains. Although classical strains can produce CT, they do not produce infectious CTXΦ. This phenomenon was explained by the arrangement of the prophage in classical strains being solitary prophages or arrays of two truncated, fused prophages (44), instead of the tandem array that is required for production of CTXΦ (45). The missing RS1 element in classical strains may also contribute to the defect, RS1 being a satellite phage, the transmission of which depends on proteins produced from the helper phage, CTXΦ. The RstC protein encoded in RS1 is an antirepressor that counteracts activity of the CTXΦ repressor, RstR, and promotes transcription of genes required for production of both phages (43).

Genes coding for the major virulence factor, namely cholera toxin (CT), *ctxAB*, are located on the CTX genetic element. The structure of the CTX genetic element is composed of a 4.5 Kb core region flanked by one or more copies of a 2.5- 2.7 Kb repetitive sequence (RS element). The later contains sequences for site-specific recombination that allows insertion of the element into an 17-18 bp attRS sequence located on the chromosome of non-toxigenic *V. cholerae* strains (136). The core region is composed of at least 6 genes, including *cep* (encoding core-encoded pilin), *orfU*

(unknown function), *ace* (encoding accessory cholera enterotoxin), *zot* (encoding zonula occludens toxin), and *ctxAB* (encoding A and B subunits of CT), respectively. Waldor and Mekalanos (188) later showed that the CTX element is actually the genome of a single-stranded DNA, lysogenic filamentous phage (CTX $\Phi$ ) and the core gene products and gene organization closely resemble those of other filamentous phages. The *cep* gene resembles gene VIII of coliphage M13 that encodes the major virion capsid protein, *orfU* whose location and size closely match those of gene III of M13, that encodes the protein responsible for receptor binding, *ace*, a gene VI homolog and proposed to code for a small hydrophobic protein that assembles into the virion particle, and *zot*, a homolog of gene I of M13 and several other filamentous phages that encode for an inner membrane protein required for assembly of the filamentous phage (188). From the observation that the core genes function as parts of the bacteriophage, doubt is cast on the proposed role of the *ace* and *zot* gene products as additional toxins, unless their products have dual functions.

The capability of CTX $\Phi$  lysogen to produce infectious phage particles, either spontaneously or by induction using mitomycin C, or via sunlight, has been demonstrated (59, 60, 188). After infecting the recipient cell, CTX $\Phi$  can form a stable lysogen by integrating into the recipient genome at the attRS site or maintaining itself extra-chromosomally as a replicative form (RF) (188).

The receptor for CTX $\Phi$  infection was shown to be the other virulence factor, toxin co-regulated pili (TCP) (188), since strains expressing the TCP were readily transduced by CTX $\Phi$  phage and strains lacking TCP appear to be resistant. However, other mechanisms of transduction exist. Faruque *et al.* (61, 62) demonstrated that a non-

toxigenic environmental strain of *V. cholerae* lacking the TCP genes, as well as a closely related species, TCP negative *Vibrio mimicus*, were infected by a genetically marked derivative of the phage (CTX-km $\Phi$ ). This alternative means of transduction is, however, less efficient than the one mediated by the TCP receptor and is strain specific.

Although TCP is considered to be the main structure for colonization, the first step in pathogenesis, those strains lacking TCP may still be able to colonize the host by using alternate mechanisms. On several occasions, clinical cases were shown to be caused by *V. cholerae* strains lacking TCP (23, 155). Other structures, such as membrane proteins that act as adhesin and other pilli elaborated by *V. cholerae*, e.g., mannose-sensitive hemagglutinin (MSHA) and NAGV14, were shown to be involved in the colonization process (132, 174, 204).

### **6.1.1 Objectives of this study**

Since *V. cholerae* is natural inhabitant of the aquatic environment and is readily isolated from water samples collected in many parts of the world, the objective of this study was to assess the pathogenic risk of environmental *V. cholerae* isolates and their risk to public health. The majority of isolates were non-O1/non-O139 *V. cholerae* in the Chesapeake Bay collection at the University of Maryland. Isolates from other geographical locations, clinical isolates, closely related *V. mimicus* strains, as well as 11 other bacterial species were also included in this study as representatives of a broader population of vibrios.

Susceptibility to CTX $\Phi$  infection without the receptor TCP was the main focus of this study, namely to assess the potential of environmental *V. cholerae* to convert to toxigenic strains. In addition to susceptibility to CTX $\Phi$ , other characteristics, such as



possession of other toxin related genes and genetic lineage determined by genomic fingerprinting were used to assist in the assessment of etiological risk of each subgroup. From the results, it was determined that members in each cluster had a different level of pathogenic risk. Thus better understanding of the potential of the natural populations of *V. cholerae* to cause disease was obtained.

## **6.2 Materials and Methods**

### **6.2.1 Bacterial strains employed in this study**

*Test strains:* A total of 304 bacterial isolates were tested for their susceptibility to CTX $\Phi$ . These included 283 isolates of *V. cholerae* nonO1/nonO139, 8 isolates of *V. cholerae* O1, 10 isolates of *V. mimicus*, and 11 isolates of other vibrios (Table 6.1). All isolates were negative for the *tcpA* genes and CTX element, except for two *V. mimicus* isolates that were negative for TCP, but positive for the CTX element.

*Donor strains :* *V. cholerae* El Tor strain SM44 carrying a single copy of the CTX element, marked by the insertion of kanamycin resistance gene (Km<sup>R</sup>) in place of *ctxAB* (70, 188) was used initially as the source of the marked CTX $\Phi$  (CTX-Km $\Phi$ ). The CTX-Km $\Phi$  was then transduced to *V. cholerae* classical strain O395 which maintained the marked phage in replicative form (RF) O395 (pCTX-Km) (188) and produced a high titer phage filtrate used in the experiment. The strain O395 was readily transduced by the phage, therefore it was selected for phage titration to determine the number of phage particles and as a positive control throughout the study.

### **6.2.2 Preparation of the CTX-Km $\Phi$ particles**

All of the following steps were performed at 30°C. The donor strains, SM44 or O395 (pCTX-Km), were grown in 5 ml Luria-Bertani (LB) broth containing 50  $\mu$ g/ml kanamycin, with aeration, overnight. A few loopfuls of the overnight culture were then sub-cultured into 50 ml fresh LB broth (pH 6.5) and grown with good aeration for 4 hr. The cell-free filtrate was obtained by centrifugation of the culture medium at 3,500 x g twice to remove the bacterial cells and filtered through 0.22- $\mu$ m pore size syringe filters (Acrodisc®, GelmanSciences, Ann Arbor, MI). One hundred microliter of the filtrate was spread onto LB agar and incubated overnight to confirm that the filtrate did not contain any bacterial cells. Concentration of the phage particles was determined using a strain that is readily transduced by the phage, the classical strain O395. The recipient cells were incubated with serially diluted phage filtrate in LB pH 6.5 for 1 hr, after which the kanamycin-resistant colonies were counted.

### **6.2.3 Assay for CTX-Km $\Phi$ infection**

Overnight cultures of the recipient strains were sub-cultured into LB broth adjusted to pH 6.5 and incubated for 4 hr with good aeration. The recipient cells were collected by centrifugation at 10,000 x g, washed, and re-suspended to one-fourth of the original volume using fresh broth medium.

Equal volumes (0.5 ml:0.5 ml) of the phage filtrate and recipient cells were mixed, either in multiwell plates or in microcentrifuge tubes, and incubated overnight without shaking. The mixtures were then plated in triplicate onto LB agar containing 50  $\mu$ g/ml kanamycin (100  $\mu$ l) to select for colonies resistant to the antibiotic. Recipient cells that had been mixed with fresh medium, instead of phage filtrate, were run in parallel, as

a control, to assess the level of natural or spontaneous resistance to kanamycin in the recipient cell population. Efficiency of infection was calculated as the number of infected cells to number of recipient cells.

#### **6.2.4 Analysis of CTX-Km $\Phi$ infected cells.**

The kanamycin resistant colonies were confirmed to be recipient cells infected with the CTX-Km $\Phi$  by using multiplex PCR targeted for *zot*, *tcpA*, and intergenic spacer region of *V. cholerae* (VC-PCR). *zot* was used as the amplification target to confirm the presence of CTX-Km $\Phi$ , since CTX-Km $\Phi$  has the kanamycin resistance gene in place of the *ctx* genes. In addition to kanamycin resistance, presence of *zot*, but not the *tcpA* gene, indicated that the colonies originated from the recipient cells and not from contamination of the donor cells (since all recipients were TCP-negative and the donor was TCP-positive). Presence of the expected product from the intergenic spacer region primers by VC-PCR indicated that the colonies did not result from contamination by other bacteria.

The kanamycin resistant colonies were suspended in 20  $\mu$ l milliQ water and boiled for 10 min in a thermal cycler. One microliter of the boiled cells was used as the PCR template. The primers used were 0.4  $\mu$ M each of pZot-225F (TCGCTTAACGA TGGCGCGTTTT) and pZot-1129R (AACCCCGTTTCACTTCTACCCA) targeted to the *zot* gene with an expected amplicon size of 947 bp (151); 0.4  $\mu$ M each of pTcpA-72F (CACGATAAGAAAACCGGTCAAGAG), pTcpA-477R (CGAAAGCACCTTCTTT CACGTTG), and pTcpA-647R (TTACCAAATGCAACGCCGAATG) targeted to the *tcpA* gene of El Tor and classical strains with expected amplicon size of 451 and 620 bp, respectively (101); and 0.15  $\mu$ M each of pVC-F2 (TTAAGCSTTTTCRCTGAGAATG) and pVCM-R1 (AGTCACTTAACCATAACAACCCG) targeted to the intergenic spacer

region between the 16S and 23S-rRNA genes of *V. cholerae*, with expected amplicon size of 295-310 bp (28). The PCR reaction was done in 20 µl reaction volumes using the Takara *Ex Taq*<sup>TM</sup> DNA polymerase and buffer system (Takara Mirus Bio Corporation, Madison, WI) containing all the multiplex primers, 200 µM each of dNTPs, and 0.5 unit of the Taq enzyme. The amplification reaction was performed in a thermal cycler (MJ Research PTC-200 Peltier Thermal cycler) with the calculated control option. The cycle conditions were initial denaturation at 95°C for 5 min, followed by 30 cycles of 94°C for 30 s, 60°C for 45 s and 72°C for 1 min, and one final extension step at 72°C for 5 min. Ten microliters of the amplified product were separated by electrophoresis using 1% agarose gel and 0.5x TBE buffer. The amplified DNAs were visualized under UV light after staining with ethidium bromide solution.

## **6.3 Results**

### **6.3.1 Production of CTX-KmΦ and its stability**

Under the conditions employed in this study, the original donor of the marked CTX-KmΦ, SM 44, produced ca.  $10^4$  infectious particles/ml. When the phage was used to infect *V. cholerae* classical strain O395, the resulting O395 (pCTX-Km) produced higher titer of the phage, ca.  $1.8 \times 10^7$  particles/ml. This high-titer filtrate was used throughout the study. With strain O395 as positive control, infectivity of the phage filtrate was determined to have been maintained at the same high level throughout the study, showing that the phage filtrate can be stored at 4°C for at least 1 month, the duration of this study, without reduction in infectivity of the phage.

### **6.3.2 Natural occurrence of kanamycin resistance in recipient cells**

The level of kanamycin resistance of the recipient cells occurring naturally, i.e., without CTX-Km $\Phi$  infection (control), was assessed in parallel with the reaction mixture with the CTX-Km $\Phi$ . Of 304 isolates tested, 54% showed no resistance to the antibiotic, since no visible colonies appeared on the kanamycin-LB plates, 24% had a very low level of resistance, shown by occurrence of very tiny colonies, 14% had a low level of resistance, shown by a number of small colonies <0.5 mm., and 8% were moderately resistant, shown by either a large number of small colonies and/or bigger colonies, 1-2 mm in diameter. These naturally resistant colonies usually grew slower than colonies containing CTX-Km $\Phi$  and grew poorly when sub-cultured onto fresh kanamycin-LB agar. Therefore, cells infected with CTX-Km $\Phi$  can be identified primarily by bigger colonies being produced on plates incubated for at least 24 hr, especially if the recipient showed moderate resistance, when compared with the control plates.

### **6.3.3 Analysis of colonies containing CTX-Km $\Phi$**

Multiplex PCR targeted for *zot*, *tcpA*, and the intergenic spacer VC-PCR was used to confirm that suspect colonies were recipient cells containing CTX-Km $\Phi$ . Based on results from testing more than 1,000 colonies, it was confirmed that although some of the recipient cells were able to grow on kanamycin-LB media due to natural resistance, the colonies containing CTX-Km $\Phi$  were easily identified after incubation for 24 hr. The size difference became more obvious as the time of incubation increased, i.e. the infected colonies continue to grow and produce very large colonies, while the naturally resistant colonies usually stopped growing or grew at a much slower rate. All colonies bigger than

those appearing on the control plate were confirmed to harbor CTX-Km $\Phi$  by multiplex PCR (*zot* positive, *tcp* negative), while all of the smaller sized colonies, as well as colonies from the control plates, gave negative results (*zot* negative-*tcp* negative).

No product from the intergenic region was expected from *V. mimicus* isolates. However, two of the eight non-toxicogenic *V. mimicus* (RC54 and RC218) gave the VC-PCR expected size amplified product. The PCR was repeated and proved not to be contamination, since the same result was also obtained when pure cultures of the recipient cells were tested. Therefore, it is concluded that it was the intrinsic property of these *V. mimicus* isolates and that the VC-PCR primers cross reacts with some very closely related strains of *V. mimicus*. For the two toxicogenic *V. mimicus* (RC55 and RC56), in which the original recipient cells already harbored the CTX element, including *zot*, transduction and its efficiency were assessed using kanamycin resistance properties of infected colonies that were bigger than naturally resistant colonies on the control plates.

#### **6.3.4 Distribution and efficiency of isolates capable of CTX $\Phi$ infection**

Of 304 TCP-negative strains tested, a total of 37 strains (12%) were infected by CTX-Km $\Phi$ . Distribution and efficiency of the infected isolates are given in Table 6.1. Of the 283 *V. cholerae* strains, 32 isolates (11%) were infected and included 25% (2/8) non-toxicogenic O1-environmental isolates, 67% (2/3) clinical nonO1/nonO139 isolates, and 10% (28/272) environmental non O1/nonO139 isolates. The closely related species, *V. mimicus*, was also found to be susceptible to CTX-Km $\Phi$  infection, of which 50% (5/10) were infected. None of the 11 other bacterial species tested were susceptible to phage

infection. Efficiency of the CTX-Km $\Phi$  infection in the TCP-negative isolates was low ( $10^{-8}$ - $10^{-6}$ ) compared to the TCP-positive strain O395 (ca.  $10^{-2}$ ).

Table 6.1. (next page) Bacterial strains and efficiency of infection by CTX-KmΦ.

<sup>a</sup> Strain numbers of the positive strains are provided. Because of the large number of positive strains from CB, USA<sup>1</sup>, only two strains that showed high efficiency ( $10^{-6}$ ) are given.

<sup>b</sup> Place: CB, USA<sup>1</sup>, Chesapeake Bay, Maryland, USA. 1998-2000; CB, USA<sup>2</sup> Chesapeake Bay, Maryland, USA. 1976-1979. Albensis<sup>T</sup> : Type strain for *V. cholerae* biotype albensis (ATCC 14547), type strain for *V. mimicus* (ATCC 33653<sup>T</sup>).

<sup>c</sup> Efficiency of infection was calculated from the number of infected cells per ca.  $10^8$  recipient cells used in the transduction reaction.



Table 6.1.

| <i>Vibrio cholerae</i>      | Strain number <sup>a</sup> | Source                | Place <sup>b</sup>          | No.positive<br>( No.tested) | Percent<br>positive | Efficiency of<br>infection <sup>c</sup>     |   |                      |
|-----------------------------|----------------------------|-----------------------|-----------------------------|-----------------------------|---------------------|---|---|----------------------|
| O1                          | UM2711                     | Env                   | USA                         | 1 (2)                       | 50                  | $6.7 \times 10^{-8}$                        |   |                      |
|                             |                            |                       | Mexico                      | 0 (2)                       | 0                   |   |   |                      |
|                             |                            |                       | India                       | 0 (2)                       | 0                   |   |   |                      |
|                             |                            |                       | Brazil                      | 0 (1)                       | 0                   |   |   |                      |
|                             | nonO1/nonO139              | RC23                  |                             | Chile                       | 1 (1)               | 100   | $8.6 \times 10^{-6}$                        |                      |
|                             |                            | RC98                  | Clin                        | Mexico                      | 1 (2)               | 50  | $3.5 \times 10^{-6}$                        |                      |
|                             | nonO1/nonO139              | UM2269                |                             | USA                         | 1 (1)               | 100   | $3.3 \times 10^{-8}$                        |                      |
|                             |                            | RC432, 501            | Env                         | CB, USA <sup>1</sup>        | 23 (224)            | 10  | $3.3 \times 10^{-8}$ - $8.6 \times 10^{-6}$ |                      |
|                             |                            |                       |                             | CB, USA <sup>2</sup>        | 4 (13)              | 31  |   |                      |
|                             |                            |                       |                             | Peru                        | 0 (10)              | 0   |   |                      |
|                             |                            | ZB124                 |                             |                             | Bangladesh          | 1 (9)                                       | 11  | $3.3 \times 10^{-8}$ |
|                             |                            |                       |                             |                             | Italy               | 0 (4)                                       | 0   |                      |
|                             |                            |                       |                             |                             | Denmark             | 0 (2)                                       | 0   |                      |
|                             |                            |                       |                             |                             | Czechoslovakia      | 0 (2)                                       | 0   |                      |
|                             |                            |                       |                             |                             | Thailand            | 0 (2)                                       | 0   |                      |
|                             |                            |                       |                             |                             | Argentina           | 0 (2)                                       | 0   |                      |
|                             |                            |                       |                             |                             | Philippines         | 0 (1)                                       | 0   |                      |
|                             |                            |                       |                             |                             | Florida, USA        | 0 (1)                                       | 0   |                      |
|                             |                            | Ballast water         | 0 (1)                       | 0                           |                     |   |   |                      |
|                             |                            | Albensis <sup>T</sup> | 0 (1)                       | 0                           |                     |   |   |                      |
| <b>Total</b>                |                            |                       |                             | <b>32 (283)</b>             | <b>11</b>           |   |   |                      |
| <i>Vibrio mimicus</i>       |                            |                       |                             |                             |                     |   |   |                      |
| Non-toxigenic               | RC5                        |                       | Type strain                 | 1 (1)                       | 100                 | $2.6 \times 10^{-6}$                        |   |                      |
|                             |                            |                       | Bangladesh                  | 0 (3)                       | 0                   |   |   |                      |
|                             | RC54                       |                       | LA, USA                     | 1 (3)                       | 33                  | $1.0 \times 10^{-7}$                        |   |                      |
| Toxigenic                   | RC6                        |                       | Japan                       | 1 (1)                       | 100                 | $1.0 \times 10^{-7}$                        |   |                      |
|                             | RC55, 56                   |                       | LA, USA                     | 2 (2)                       | 100                 | $1.0 \times 10^{-7}$ - $4.7 \times 10^{-7}$ |   |                      |
|                             |                            |                       | <b>Total</b>                | <b>5 (10)</b>               | <b>50</b>           |   |   |                      |
| Other species : 11 isolates |                            |                       |                             |                             |                     |   |   |                      |
|                             |                            |                       | <i>Vibrio fluvialis</i>     | CB, USA <sup>1</sup>        | 0 (4)               | 0   |   |                      |
|                             |                            |                       | <i>Vibrio</i> spp.          | Deep sea vent               | 0 (2)               | 0   |   |                      |
|                             |                            |                       | <i>Vibrio harveyi</i>       | Deep sea vent               | 0 (1)               | 0   |   |                      |
|                             |                            |                       | <i>Shewanella alga</i>      | Deep sea vent               | 0 (1)               | 0   |   |                      |
|                             |                            |                       | <i>Shewanella</i> spp.      | CB, USA <sup>1</sup>        | 0 (1)               | 0   |   |                      |
|                             |                            |                       | <i>Aeromonas</i> spp.       | CB, USA <sup>1</sup>        | 0 (1)               | 0   |   |                      |
|                             |                            |                       | <i>Exiguobacterium</i> spp. | CB, USA <sup>1</sup>        | 0 (1)               | 0   |   |                      |
|                             |                            |                       | <b>Total</b>                | <b>0 (11)</b>               | <b>0</b>            |   |   |                      |
| <b>Total</b>                |                            |                       |                             | <b>37 (304)</b>             | <b>12</b>           |   |   |                      |

### 6.3.5 Assessment of pathogenic risk by cluster membership and other toxin-related genes

When the positive isolates were considered by cluster membership (determined by genomic fingerprinting, see Chapter 5), the CTX $\Phi$ -susceptible isolates were found to be distributed in all of the major clusters of *V. cholerae* and the *V. mimicus* cluster. Table 6.2 lists the pathogenic risk of the isolates, classified by cluster membership. With respect to the percent positive isolates susceptible to CTX $\Phi$ , the most susceptible cluster was cluster M, which included all of the *V. mimicus* (42%), followed by clusters associated with epidemic clones, A9 (33%) and A3 (18%). Cluster A5 yielded 17% positive, followed by Cluster A1, which included a majority of the luminescent *V. cholerae* non-O1/non-O139 (14%), and the rest of Cluster A (11%). Cluster B included all of the low *toxR* non-O1/non-O139 isolates gave lower percentage of susceptible isolates (4%). Cluster C, which comprised only three clonal isolates of luminescent, *toxR* negative, sucrose negative isolates was negative for the test (0%).

Table 6.2. Pathogenic risk of the isolates according to cluster membership

| Cluster <sup>a</sup> | Susceptible to CTX $\Phi$ |        |      | <i>stn</i> |        |      | Average intensity to gene probe <sup>b</sup> |             |             |             |
|----------------------|---------------------------|--------|------|------------|--------|------|--|-------------|-------------|-------------|
|                      | No.test                   | No.pos | %    | No.test    | No.pos | %    | <i>toxR</i>                                  | <i>hlyA</i> | <i>ompU</i> | <i>ompW</i> |
| A1                   | 113                       | 16     | 14.2 | 112        | 37     | 33.0 | 3.95   | 3.87        | 3.85        | 3.95        |
| A2                   | 12                        | 0      | 0.0  | 12         | 7      | 58.3 | 4.00   | 4.00        | 3.70        | 3.70        |
| A3                   | 11                        | 2      | 18.2 | 15         | 1      | 6.7  | 4.00   | 4.00        | 1.83        | 3.83        |
| A4                   | 8                         | 0      | 0.0  | 8          | 0      | 0.0  | 4.00   | 4.00        | 2.88        | 4.00        |
| A5                   | 6                         | 1      | 16.7 | 6          | 0      | 0.0  | 4.00   | 3.83        | 4.00        | 3.67        |
| A6                   | 3                         | 0      | 0.0  | 3          | 0      | 0.0  | 4.00   | 4.00        | 4.00        | 4.00        |
| A7                   | 3                         | 0      | 0.0  | 3          | 0      | 0.0  | 4.00   | 4.00        | 4.00        | 4.00        |
| A8                   | 3                         | 0      | 0.0  | 3          | 0      | 0.0  | ND   | ND          | ND          | ND          |
| A9                   | 3                         | 1      | 33.3 | 8          | 0      | 0.0  | ND   | ND          | ND          | ND          |
| A10                  | 0                         | 0      | ND   | 2          | 0      | 0.0  | ND   | ND          | ND          | ND          |
| Ax                   | 89                        | 11     | 12.4 | 81         | 5      | 6.2  | 3.98   | 3.78        | 3.37        | 3.78        |
| B                    | 27                        | 1      | 3.7  | 27         | 0      | 0.0  | 1.38   | 3.38        | 2.81        | 1.77        |
| C                    | 3                         | 0      | 0.0  | 3          | 0      | 0.0  | 0.00   | 1.00        | 2.67        | 2.00        |
| M                    | 12                        | 5      | 41.7 | 11         | 6      | 54.5 | 3.00   | 2.00        | 3.00        | 2.00        |

<sup>a</sup> Cluster membership was determined by ERIC-PCR genomic fingerprinting.

<sup>b</sup> Average signal intensity of toxin-related genes by gene probe hybridization, calculated from an intensity scale of 0-4; where 0 is negative and 4 is the strongest signal.

In addition to the ability of non-toxigenic *V. cholerae* to be infected by CTXΦ, information about other toxin-related genes in these isolates was used to assess the overall pathogenic risk of each cluster. As shown in Table 6.2, members of Cluster M (*V. mimicus* cluster) have the highest potential (42%) to acquire the cholera toxin gene via CTXΦ infection. As many as 55% of the members of this cluster also possessed the *stn* gene and the gene with high sequence similarity to *toxR*, the regulatory gene essential for expression of cholera toxin. Hence, members of cluster M are considered to be high risk.

Among the other non-toxigenic *V. cholerae*, members of cluster A3 and A9, the toxigenic cluster, were considered to be the highest etiological risk, because in addition to genome similarity to epidemic strains, they were susceptible to CTXΦ, and, in case of A3 cluster, 7% of the members also possess *stn* gene, the gene coding for non-O1 heat stable enterotoxin (NAG-ST). Cluster A1 and A2, which was composed of the majority of luminescent *V. cholerae*, was considered to be the next high risk. High percentage of isolates in A1 and A2 clusters possessed *stn* gene (33% and 58%, respectively). Isolates in A1 cluster were also susceptible to CTXΦ (14%) On average, the gene with high sequence similarity to all of the toxin-related genes tested (*toxR*, *hlyA*, *ompU*, and *ompW*) were present. The next highest risk group comprised the rest of Cluster A, a diverse group not forming a tight cluster. Potential risk from members of this group included 12% positive for infection with CTXΦ, 6% possessed the *stn* gene, and on average, genes with high sequence similarity to *toxR*, *hlyA*, *ompU*, and *ompW*. Cluster B was composed of all *V. cholerae* possessing genes with low sequence similarity to the *toxR* gene, suggesting lower etiological risk than members of the A cluster, with only 4% able to be infected by CTXΦ, none *stn* positive, lower signal intensity for the *ompU* and *ompW*

genes, and significantly lower signal intensity for the toxin regulatory gene, *toxR*.

Finally, the small cluster C, composed of only three isolates of the same clone showed the least pathogenic potential because the members did not infect with CTX $\Phi$ , did not possess the *stn* and *toxR* genes, and had genes with very low sequence similarity to *hlyA* and *ompW*.

Summary of the etiological risk by their characteristics and cluster membership is shown in Table 6.3.

Table 6.3. Estimation of etiological risk of non-toxigenic *V. cholerae* and *V. mimicus*

| Etiological risk | Group/Cluster         | Main characteristics        | Percent positive |                     |
|------------------|-----------------------|-----------------------------|------------------|---------------------|
|                  |                       |                             | CTXΦ infection   | <i>stn</i> positive |
| High<br>↓<br>Low | M                     | <i>V. mimicus</i>           | 42               | 55                  |
|                  | Clinical <sup>a</sup> | Non-O1/non-O139, clinical   | 67               | 0                   |
|                  | A9                    | Toxigenic cluster           | 33               | 0                   |
|                  | O1 <sup>a</sup>       | Serogroup O1, environmental | 25               | 0                   |
|                  | A3                    | Toxigenic cluster           | 18               | 7                   |
|                  | A5                    |                             | 17               | 0                   |
|                  | A1                    | Luminescence                | 14               | 33                  |
|                  | A2                    | Luminescence                | 0                | 58                  |
|                  | Ax                    | Other Cluster A             | 12               | 6                   |
|                  | B                     | Low <i>toxR</i> gene        | 4                | 0                   |
| Low              | C                     | sucrose-luminescence        | 0                | 0                   |

<sup>a</sup> The groups were identified by their special characteristics, i.e., from clinical sources or belongs to O1 serogroup. Two positives from clinical isolates were in Ax and A9 clusters and two positives from O1 serogroup were in Ax cluster.

Table 6.4 provides habitat information, as well as serogroup of the 37 isolates positive for CTX $\Phi$  infection. Although the number of positive isolates was too small to draw conclusions about a specific habitat associated with positive CTX $\Phi$  infection, this trait appears to be widely spread. The trait was found in isolates from all three sample types, six sampling stations in Chesapeake Bay, as well as from several different countries, in samples of pH range 7.6-8.0, temperatures of 12-28 °C, salinity from 2 to 13 ppt, and isolated from 1977 through 2000. The ability to be infected with CTX $\Phi$  was also found to be distributed among diverse serogroups, i.e., at least 15 of the serogroups in this study (Table 6.4).

Table 6.4. Environmental parameters and serogroups of strains susceptible to CTX $\Phi$ .

| ID     | Cluster | Species            | Serogroup <sup>1</sup> | Place      | Year | Site | Frac | pH  | Temp | Sal  |
|--------|---------|--------------------|------------------------|------------|------|------|------|-----|------|------|
| UM2269 | A       | <i>V. cholerae</i> |                        | Fl, USA    | 1980 |      |      |     |      |      |
| RC98   | A       | <i>V. cholerae</i> | O5                     | Mexico     |      |      |      |     |      |      |
| RC23   | A       | <i>V. cholerae</i> | O1                     | Chile      |      |      |      |     |      |      |
| UM2711 | A       | <i>V. cholerae</i> | O1                     | Fl, USA    | 1981 |      |      |     |      |      |
| ZB124  | A       | <i>V. cholerae</i> |                        | Bangladesh | 1997 |      |      |     |      |      |
| UM943  | A       | <i>V. cholerae</i> | O24                    | CB, USA    | 1977 |      |      |     |      |      |
| RC354  | A       | <i>V. cholerae</i> | O45                    | CB, USA    | 1998 | K    | P    | 7.9 | 21.5 | 2.0  |
| RC343  | A       | <i>V. cholerae</i> | O128                   | CB, USA    | 1998 | H    | W    | 8.0 | 24.0 | 5.0  |
| RC400  | A       | <i>V. cholerae</i> | O135                   | CB, USA    | 1998 | S    | W    | 8.1 | 27.0 | 6.0  |
| RC342  | A       | <i>V. cholerae</i> | O186                   | CB, USA    | 1998 | H    | W    | 8.0 | 24.0 | 5.0  |
| RC432  | A       | <i>V. cholerae</i> | O200                   | CB, USA    | 1998 | K    | W    | 8.0 | 25.0 | 6.0  |
| RC355  | A       | <i>V. cholerae</i> | R*                     | CB, USA    | 1998 | S    | W    | 8.1 | 21.0 | 2.0  |
| UM993  | A       | <i>V. cholerae</i> |                        | CB, USA    | 1978 |      |      |     |      |      |
| RC521  | A       | <i>V. cholerae</i> |                        | CB, USA    | 1999 | B    | W    | 7.9 | 12.5 | 5.0  |
| RC601  | A       | <i>V. cholerae</i> |                        | CB, USA    | 2000 | 9    | W    |     | 23.6 | 8.7  |
| UM947  | A1      | <i>V. cholerae</i> | O2                     | CB, USA    | 1977 |      |      |     |      |      |
| UM987  | A1      | <i>V. cholerae</i> | O19                    | CB, USA    | 1977 |      |      |     |      |      |
| RC457  | A1      | <i>V. cholerae</i> | O58                    | CB, USA    | 1998 | S    | P    | 8.0 | 26.0 | 6.0  |
| RC394  | A1      | <i>V. cholerae</i> | O62                    | CB, USA    | 1998 | K    | P    |     | 25.0 | 7.0  |
| RC509  | A1      | <i>V. cholerae</i> | O62                    | CB, USA    | 1998 | K    | Z    |     | 25.0 | 7.0  |
| RC430  | A1      | <i>V. cholerae</i> | O62                    | CB, USA    | 1998 | K    | W    | 8.0 | 25.0 | 6.0  |
| RC380  | A1      | <i>V. cholerae</i> | O135                   | CB, USA    | 1998 | H    | W    | 7.6 | 28.0 | 5.0  |
| RC385  | A1      | <i>V. cholerae</i> | O135                   | CB, USA    | 1998 | H    | P    | 7.6 | 28.0 | 5.0  |
| RC388  | A1      | <i>V. cholerae</i> | O201                   | CB, USA    | 1998 | H    | Z    | 7.6 | 28.0 | 5.0  |
| RC455  | A1      | <i>V. cholerae</i> | O201                   | CB, USA    | 1998 | S    | P    | 8.0 | 26.0 | 6.0  |
| RC438  | A1      | <i>V. cholerae</i> | R*                     | CB, USA    | 1998 | K    | P    | 8.0 | 25.0 | 6.0  |
| RC501  | A1      | <i>V. cholerae</i> |                        | CB, USA    | 1998 | K    | W    |     | 25.0 | 7.0  |
| RC484  | A1      | <i>V. cholerae</i> |                        | CB, USA    | 1998 | S    | P    | 7.7 | 28.0 | 9.0  |
| RC475  | A1      | <i>V. cholerae</i> |                        | CB, USA    | 1998 | H    | P    | 7.3 | 28.0 | 9.5  |
| RC594  | A1      | <i>V. cholerae</i> |                        | CB, USA    | 2000 | 9    | W    |     | 23.7 | 4.7  |
| RC606  | A1      | <i>V. cholerae</i> |                        | CB, USA    | 2000 | 3    | W    |     | 24.6 | 12.7 |
| RC341  | B       | <i>V. cholerae</i> | O153                   | CB, USA    | 1998 | H    | W    | 8.0 | 24.0 | 5.0  |
| RC54   | M       | <i>V. mimicus</i>  | O101                   | LA, USA    |      |      |      |     |      |      |
| RC55   | M       | <i>V. mimicus</i>  | O115                   | LA, USA    |      |      |      |     |      |      |
| RC56   | M       | <i>V. mimicus</i>  | O115                   | LA, USA    |      |      |      |     |      |      |
| RC5    | M       | <i>V. mimicus</i>  | O135                   | ATCC       |      |      |      |     |      |      |
| RC6    | M       | <i>V. mimicus</i>  | R*                     | Japan      |      |      |      |     |      |      |

<sup>1</sup> Partial results from serotyping performed by Dr. Eiji Arakawa and provided courtesy of Dr. Haruo Watanabe, Department of Bacteriology, National Institute of Infectious Diseases, Japan.



## 6.4 Discussion

### 6.4.1 Occurrence of spontaneous kanamycin resistance in recipient cells

Because a high concentration of recipient cells was used in the transduction experiments, it was expected that there might be a small proportion of cells resistant to the antibiotic, complicating interpretation of the results. Therefore, a control, in which only recipient cells without the phage filtrate were plated on the kanamycin-LB plate, was run in parallel. The results showed that the majority of the isolates were highly susceptible to kanamycin. That is, 8% of the isolates were moderately resistance and produced 1-2 mm diameter colonies. However, the naturally resistant colonies were less capable of proliferating compared to those colonies infected with CTX-Km $\Phi$ , and could be differentiated by comparing colony size after subculturing on fresh kanamycin-LB agar or for a longer incubation time.

Kanamycin is an antibiotic of the aminoglycoside family that acts on the 30S ribosomal subunit and inhibits protein synthesis by bacterial cells. Resistance to the drug can result from inactivation of the drug by aminoglycoside-modifying enzymes (AMEs), such as that encoded in the Km-resistance cassette of CTX-Km $\Phi$ , ribosomal alterations that prevent the drug from binding to its binding site, or loss of permeability of the cell membrane to the drug (125). Partial resistance to the drug shown by uninfected cells was probably due to one or more of these mechanisms. However, their efficiency was lower than that of the CTX-Km $\Phi$  infected cells, resulting in a slower growth rate, especially when subcultured onto fresh Km-LB agar.

#### 6.4.2 Mechanism of CTX-Km $\Phi$ infection in TCP-negative isolates

From the results shown in Table 6.1, the efficiency of CTX-Km $\Phi$  infection of TCP-negative isolates was very low (ca.  $10^{-8}$  -  $10^{-6}$ ), compared with the TCP-positive control strain (ca.  $10^{-2}$ ). The role of TCP as CTX $\Phi$  receptor has been described by Waldor and Mekalanos (188). The strains expressing TCP structure were highly susceptible to infection (as high as 50% of the cells can be infected), while the same strains under condition not suitable for TCP expression were completely resistant or were infected at a much lower rate (ca.  $0$ - $10^{-9}$ ). Faruque *et al.* (61) also found that all TCP-positive isolates were susceptible to CTX $\Phi$  infection with an efficiency of ca.  $10^{-6}$  *in vitro*, and as high as  $10^{-1}$ - $10^0$  in the mouse intestine, a condition known to promote expression of TCP. However, their study also reported a rare (1 of 136) TCP-negative strain that was infected by CTX $\Phi$  at low efficiency (ca.  $10^{-7}$ ), whether *in vivo* or *in vitro*. It was proposed that, in addition to the TCP-mediated mechanism, there may be an alternate mechanism for CTX $\Phi$  infection.

A study conducted by Heilpern and Waldor (81) showed that *tolQRA* gene products were necessary for infection of CTX $\Phi$  into *V. cholerae*, possibly in a parallel mechanism as the infection of other filamentous phages that infect *Escherichia coli*. The efficient entry of CTX $\Phi$  into *V. cholerae* requires both a TCP pilus and *tolQRA* gene products. It was proposed that the TCP receptor enables effective binding of the phage at a significant distance from the cell surface and directs the phage to the outer membrane, where it can interact with the TolQRA complex, which facilitates its traversal through the periplasmic space. Without TCP, however, the phage particles can occasionally, although less efficiently, come in contact with the outer membrane and interact with the

TolQRA complex, resulting in the low efficiency of infection observed for TCP-negative strains. Therefore, our results are in agreement with the TCP independent, low efficiency mode of infection. However, the percent of strains capable of being infected by CTX $\Phi$  was higher than previously reported (61). In our studies, as high as 11% of *V. cholerae* lacking TCP could be infected with CTX $\Phi$ .

Although the CTX $\Phi$  infection of TCP-negative strains was less efficient and the high concentration of recipient cells and phage particles used in this study may be considered less likely to be encountered in the natural environment, the possibility of conversion to a toxigenic strain cannot be overlooked. By attaching and multiplying on particulate matter or in biofilms, localization of high concentrations of susceptible cells and donor cells producing infectious CTX $\Phi$  is possible, and thereby contributes to the emergence of a new strain containing the major virulence factor, namely, the CTX genes.

#### **6.4.3 Other toxin-related genes related to pathogenic risk.**

In addition to the ability to be infected by the CTX $\Phi$ , possession of two toxin genes, *stn* and *hlyA*, the regulatory gene *toxR*, and two outer membrane proteins *ompU* and *ompW*, were considered in assessing the pathogenic risk of the isolates.

*Heat stable enterotoxin (NAG-ST)*: The *stn* gene product, a heat stable enterotoxin (NAG-ST), is one of the virulence factors proposed to explain clinical manifestations of non-O1 gastroenteritis. This toxin is a well-studied virulence factor of *V. cholerae*. A human volunteer study showed that, in the presence of adequate colonization factors, it can caused diarrhea of a severity comparable to that of CT-producing strains (127).

Several investigators reported detection of the ST gene in both environmental and clinical isolates of *V. cholerae*, ranging from 2-10% in environmental isolates (42, 134, 159) to 7-33% in clinical isolates (8, 75, 83). Our analysis of the *stn* gene among *V. cholerae* from Chesapeake Bay showed that the overall frequency of isolates possessing this gene was 20.5% (see Chapter 3). However, the gene is not evenly distributed within the species. The luminescent cluster, A1 and A2, had the highest percent positive (33% and 58%, respectively), while the low or no *toxR* gene clusters (Cluster B and C) were negative for the gene.

The ST gene, although less frequent, was also found in the O1 serogroup (*sto* gene). The nucleotide sequence of *sto* is very similar to that of *stn* (98% homology) and the predicted amino acid sequence has only four mismatches in the N –terminus of the toxin precursor (131). Both genes were flanked by 123 base pair direct repeats (131). The repeat and the gene are part of VCRs (*Vibrio cholerae* repeated sequences), a large integron providing a system for acquisition of heterologous genes (gene capture system) found in *V. cholerae* and in some strains of *V. mimicus*, *V. metschnikovii*, and *V. parahaemolyticus* (120). This finding suggests a mobile nature of the ST gene in *V. cholerae*.

*El Tor Hemolysin*: Another well characterized toxin in *V. cholerae* is a hemolysin, called El Tor hemolysin since it is a characteristic of the biotype El Tor in contrast to the classical biotype that does not lyse red blood cells. However, the gene encoding the hemolysin is found in most *V. cholerae*, including classical, El Tor and non O1 isolates. The purified hemolysin was shown to be cytolytic for a variety of erythrocytes and mammalian cells in culture and lethal for mice (84), capable of causing

bloody fluid, with mucus-fluid accumulation, in ligated rabbit ileal loops which is different from the watery fluid cause by CT (89). In classical strains, there is 11-bp deletion, leading to a truncated protein of 27 kDa instead of the full size 65 kDa active cytolysin. However, the truncated protein is still capable of lysing chicken and rabbit but not sheep erythrocytes (148). Therefore, it was proposed that the enterotoxic activity lies in the N-terminal within the truncated protein.

Despite the clear enterotoxicity of the purified hemolysin, a direct role of the hemolysin gene in *V. cholerae* pathogenicity in humans has not been established. In contrast, Morris *et al.*(127) described a non-O1 *V. cholerae* containing the hemolysin gene (as well as producing a Shiga-like toxin, but no CT or NAG-ST) and the ability to lyse rabbit RBCs, causing diarrhea and death in animal models but failing to cause diarrhea in any of the 13 human volunteers at an inoculum size as high as  $10^9$ . From the results of the volunteer study and the universal distribution of the *hlyA* gene among environmental isolates, the role of the hemolysin in human pathogenesis is questionable, possibly because of a lack of gene expression or the absence of the receptor for the toxin in the human host.

The regulatory gene, *toxR*, is a central regulator, coordinating with environmental signals and regulating expression of several genes (137), including the *ctx*, *tcp*, and *ompU* genes. ToxR has been shown to be essential for virulence of *V. cholerae* (50, 82). The gene is universally found in a majority of *V. cholerae* (cluster A), as well as *V. mimicus*, indicating a capability to express CT when the *ctx* gene is present. For the low and no *toxR* cluster B and C, a lower risk was concluded since expression of CT may not be possible in these isolates.

The outer membrane protein, 38 kDa porin OmpU is a major outer membrane protein of *V. cholerae* regulated directly by ToxR. OmpU has been shown to be critical for *V. cholerae* bile resistance (141, 142), intestinal colonization (173, 174), and also to play a role in organic acid tolerance response (ATR) (123). These properties imply an importance of the *ompU* gene in adaptation of *V. cholerae* to survive in the environment inside the host.

Another membrane protein, OmpW, is a 22 kDa, highly immunogenic protein, expressed in minor amounts under normal laboratory conditions (116). The function of this protein is currently unknown but it has been suggested to be a porin-like protein (116) and has a high sequence similarity to a colicin S4 receptor, OmpW in *E. coli* (138). Parts of the OmpW amino acid sequence were shown to be identical to a partial sequence of a 20-kDa pilus protein produced by a diarrheogenic strain of non-O1/non-O139 *Vibrio cholerae* (165). This pilus protein was shown to have a function in intestinal colonization.

Consistent with other toxin related genes, the majority of *V. cholerae* (Cluster A) possess genes with high sequence similarity to the two outer membrane proteins, while strains of Cluster B and C had genes with less similarity, indicating that the B and C cluster has less adaptive ability to colonize and survive inside the host.

#### **6.4.4 Levels of pathogenic risk of non-toxigenic *V. cholerae***

As shown in Table 6.3, there were different levels of etiological risk concluded to be associated with each group of non-toxigenic *V. cholerae*. Clinical isolates as well as strains in toxigenic cluster A9 and A3, and *V. cholerae* O1, were in the highest risk group with a greater possibility to be positive for CTX $\Phi$  infection. This finding emphasizes the

importance of domestic waste management to prevent contamination of water reservoirs from clinical strains, since the strains have a high potential to become more virulence if exposed to the phage. The non-toxigenic O1 serogroup, showed as many as 25% of the isolates could be infected with CTX $\Phi$ . Notably, most of epidemic strains also belonged to the O1 serogroup. Therefore, expression of the O1 antigen may have a positive relationship to CTX $\Phi$  infection and may facilitate conversion of this serogroup to toxigenicity.

To the non-O1 environmental *V. cholerae*, which was a majority of the *V. cholerae* population in the natural habitat, a moderate to low risk was assigned, depending on the cluster to which the strain belonged. Strains in cluster A1 and A2 were shown to have relatively high etiological risk because of the high percent of strains susceptible to CTX $\Phi$  and possessing *stn* gene.

Strains in the A1-A2 clusters, which had the distinctive characteristic that most of the members were luminescent or contained luminescent genes, should be closely monitored as a risk, in addition to the high risk O1 serogroup. Unfortunately, luminescence was not included in routine identification schemes for *V. cholerae*, hence overlooked with no clinical significance attached to luminescent *V. cholerae*.

Isolates in cluster B and C were lowest risk groups since they were rarely infected with CTX $\Phi$ . In addition, they did not possess the *toxR* gene (or had the gene with very low similarity to the *toxR* gene), a gene required for CT expression, no *stn* gene, and lower sequence similarity to the two outer membrane proteins, *ompU* and *ompW*.

#### **6.4.5 Habitat and serotype of isolates susceptible to CTX $\Phi$ infection**

Besides etiological risk associated with specific populations of *V. cholerae*, a specific habitat associated with the high risk factors could not be identified. The high risk isolates were found in diverse habitats and persisted in the natural environment. The currently available serotyping data also indicated no relationship of high risk traits with serogroup, since at least fifteen different serogroups of *V. cholerae* and three different serogroups of *V. mimicus* were susceptible to CTX $\Phi$  infection. A complete serotyping of all the tested isolates is currently underway. The results will provide additional information with respect to prevalence of each serogroup, their association, if any, with specific clusters, and associated etiological risk.

#### **6.4.6 *V. mimicus*, a reservoir for toxin genes**

*V. mimicus*, a closest relative of *V. cholerae* and possibly the same species is also considered a human pathogen and a natural inhabitant of aquatic environments. Significance of this species as a human pathogen and its role as reservoir for toxin genes has been implied from accumulated evidence of toxin related genes, similar to those of *V. cholerae*, in *V. mimicus*, such as CT, VPI, heat-stable enterotoxin ST, hemolysin, and *toxR* (18, 169, 205).

In this study, a high percent of *V. mimicus*, higher than *V. cholerae*, was susceptible to infection by CTX $\Phi$  and harbored the heat-stable enterotoxin, *stn* gene. The role of *V. mimicus* as a reservoir for CTX genes was supported by results of a study reported by Faruque *et al.* (62) showing that infected *V. mimicus* harboring CTX $\Phi$  was able to produce large number of infectious phage particles.



In another study, Arita *et al.* (5) reported that *V. mimicus* was able to produce a heat-stable enterotoxin (VM-ST) identical to NAG-ST produced by *V. cholerae*. Subsequence studies suggested that the gene coding for this toxin is widely distributed in *V. mimicus* (185, 205). In an extensive study including 1109 strains of bacteria belonging to 17 *Vibrio* spp., the ST gene was found in only two species, *V. cholerae* and *V. mimicus* and it was proposed, from the high frequency of the strains harboring the gene, that *V. mimicus* is the reservoir of the *stn* gene (205).

Based on this information, *V. mimicus* populations in the environment should be closely monitored, considering the organism has a high risk to become toxigenic and serves as the reservoir for mobile toxin genes.

## 6.5 Conclusions

A total of 11% of TCP-negative *V. cholerae*, and 50% of *V. mimicus* were susceptible to CTX $\Phi$  infection, whereas none of the isolates of other species tested in this study were susceptible. No specific habitat or serogroup could be associated with strains susceptible to the phage. Different levels of etiological risk associated with members of different genetic lineages were determined, mainly from their susceptibility to CTX $\Phi$ , possession of other toxin *stn* genes, regulatory gene *toxR*, and the outer membrane proteins, *ompU* and *ompW*, with possible roles in colonization and adaptation of the organism inside host environment. Among all TCP-negative strains tested, members of *V. mimicus* cluster were in the highest risk and should be closely monitored. *V. cholerae* clinical isolates, O1 serogroups, and strains belonged to toxigenic cluster A9 and A3, presented the next highest risk. There were also high risk associated with luminescent cluster A1 and A2, followed with other members in Cluster A. In contrast, members of

clusters B and C with low or no *toxR* gene homology presented the lowest risk with very low percentage of susceptible isolates and no *stn* gene positive strains.

Results of this study were used to establish the pathogenic risk of natural populations of *V. cholerae* and *V. mimicus* in the environment. Although most of the environmental isolates were TCP negative, they were susceptible to CTX $\Phi$  infection and pose health risk if CTX $\Phi$  is present and in contact with the bacterial populations. *V. mimicus* is proposed as a high risk group and a reservoir for toxin genes, CTX elements and *stn* gene. The luminescent property of *V. cholerae* should be recognized as being associated with etiological risk, although the environmental role these genes in *V. cholerae* may play needs to be elucidated.

## Chapter 7 Summary

The objective of this research was to determine the population structure of *Vibrio cholerae*. To achieve this objective, a collection of *V. cholerae* isolated from throughout the Chesapeake Bay was extensively studied. In addition, a set of strains from both clinical and environmental sources, representing diverse geographical areas as well as closely related species, such as *V. mimicus*, was also included to define the structure of the overall population of *V. cholerae*.

As a first step, a simple and effective method was developed to isolate and identify *V. cholerae* from environmental samples. Several biochemical tests that are conventionally used to identify *V. cholerae* were studied to assess their effectiveness in differentiating *V. cholerae* from other bacterial species growing on TCBS agar. A simple procedure that employed a limited number of biochemical tests was found to be highly effective in identifying *V. cholerae*, providing 100% sensitivity and 99% specificity, compared with a molecular method employing *V. cholerae*- specific gene probe/primers. The procedure included two tests, arginine dihydrolase and esculin hydrolysis. An evaluation of the simple scheme with additional environmental isolates from Peru, was carried out and strains that did not hydrolyze esculin, but blackened the medium were observed, thus, the simple scheme for *V. cholerae* identification was modified to recommend interpreting esculin hydrolysis by screening the test medium with a UV lamp.

Two methods for isolating *V. cholerae*, enrichment and direct plating, were compared. Enrichment was concluded to be more sensitive, yielding many different *V. cholerae* clones, but also “artificial clones”, and did not detect cells tightly bound to plankton. Direct plating using a non-selective medium combined with colony hybridization presented problems in determination of positive colonies and had low sensitivity and, therefore, was not optimal for isolation and enumeration. However, atypical strains, i.e., sucrose negative *V. cholerae*, were isolated by this method. A new method of isolation, taking advantages of both methods, by enrichment of dilutions of homogenized samples, followed with detection by colony hybridization, is proposed.

Another method developed in this study is genomic fingerprinting using long-range enterobacterial repetitive intergenic consensus (ERIC) PCR. The optimized protocol has an advantage of an improved long-range Taq polymerase enzyme to produce a highly efficient PCR reaction and larger amplified products. Gel electrophoresis conditions providing high resolution of the fingerprint patterns was defined and the protocol yielded reproducible and discriminative fingerprint patterns for the *V. cholerae* genome. The closely related *V. cholerae* serogroups O1 Classical, O1 El Tor, and O139 could be differentiated by visual inspection of the bands at positions 1.0 and 1.4 kb. Similarity was measured for the large number of fingerprints by using a curve-based analysis based on densitometric profiles of the banding patterns. This calculation proved to be superior to band-based analysis using binary data representing presence or absence of bands. The similarity values obtained using the ERIC fingerprints yielded good correlation with the similarity calculations based on DNA-DNA hybridization. The results showed higher resolution, with less measurement error, in the case of similarity

measured from the ERIC PCR fingerprint than by the relative binding ratio of DNA-DNA hybridization. Therefore, ERIC PCR fingerprinting, as described in this study, was concluded to be the preferable method, because it is rapid and simple and yielded high throughput, with robust results. It allows assessment of clonality and genetic relatedness and was very effective for *V. cholerae*.

By using ERIC PCR fingerprinting to identify clonality of the isolates, artificial clones from enrichment were identified and removed from the strain set, allowing the creation of a non-redundant collection of 224 *V. cholerae* isolates from the Chesapeake Bay. Seasonal as well as annual variation was observed on the isolation of *V. cholerae*. In the case of seasonal variation, *V. cholerae* were isolated from spring until fall, with a peak in number in the summer, with 87% of the strains in the collection were isolated during this time. Although no isolates were obtained at other times of the year, *V. cholerae* DNA was detected in the water throughout the year, with the only exception being in the coldest months of winter when the water was frozen (0-4°C) and the following month (5-8°C). Annual variation was observed and was found to be linked to fresh water influx into the Chesapeake Bay that affected salinity and nutrient concentration in Chesapeake Bay. The year 1998, with high fresh water influx, yielded a more than 5 fold larger number of isolates than the low influx year of 1999.

Phenotypic and genotypic analyses of the stains from the Chesapeake Bay collection were also done. Most of the isolates were able to degrade macromolecules, such as starch, gelatin, chitin, and lipid, suggesting that *V. cholerae* is ubiquitous to the aquatic environment and that it plays a role in nutrient cycling of these macromolecules. None of the isolates possessed the gene coding for the major *V. cholerae* toxin CT or

colonization factor TCP. However, most possessed the gene coding for hemolysin, suggesting this gene to be ecologically significant. Another toxin gene *stn*, coding for NAG-ST, was found in relatively high frequency (21%) among the strains. It is interesting to note that most of the *stn* positive strains were also luminous, suggesting a possible role of the two genes in the association of *V. cholerae* with higher organisms, such as zooplankton and aquatic animals. A higher than ever before reported incidence of luminous *V. cholerae*, as high as 50%, was observed, suggesting a possible selective advantage of luminescent strains in the Chesapeake Bay. And it should be noted that the luminescence trait is common in some populations of *V. cholerae*.

Cluster analysis of the 224 Chesapeake Bay isolates, as well as 80 additional isolates from diverse geographical locations (from 13 countries) including environmental and clinical isolates, and 10 strains of *V. mimicus*, were analyzed based on distance estimation using ERIC PCR fingerprints. Three clustering methods, Neighbor-Joining (NJ), Unweighted Pair Group Method with Arithmetic Mean (UPGMA), and Fitch-Margoliash, were used to assess significance and reliability of the branching structure. Two primary clusters were found among the *V. cholerae* isolates, designated Clusters A and B. Cluster A comprised the majority (and diverse) *V. cholerae*, with several levels of subclusters identified. Several unique subclusters were described, e.g., toxigenic and luminescence clusters. Another primary cluster of *V. cholerae* was Cluster B, comprising strains with some characteristics different from those of the major group. Notably, for example, in contrast to strains in Cluster A, all strains in Cluster B possessed a gene with very low similarity to the *toxR* gene. Some strains in this cluster presented a phenotype similar to *V. mimicus*, especially in the Voges-Proskauer (VP) test; however, they were

all sucrose positive. All *V. mimicus* formed a separate cluster, designated Cluster M. From the fact that clusters B and M give similar distance estimations with Cluster A, the question of identity or species designation of strains in clusters B and M was investigated. Genomic similarity measured by DNA-DNA hybridization to three probes generated from strains from clusters A and B supported separation of the A, B, and M clusters, as well as similar distance of clusters B and C from Cluster A. The 16S rDNA sequences of several representative strains of the three clusters revealed >99% identity. Additional information on *V. cholerae* isolates in Cluster B and another small clonal cluster with the luminescent and sucrose negative traits of Cluster C, expanded the diversity range of the *V. cholerae* population, especially in terms of genomic homology. Low overall genomic homology among strains of the different clusters may be due to high horizontal gene transfer and/or foreign genes captured by strains in each cluster. This additional information argues strongly against separation of *V. mimicus* since it had the same phylogenetic status as *V. cholerae* Cluster B.

The significance of the *V. cholerae* population structure was demonstrated by the pathogenic risk of strains in each cluster. The pathogenic risk of non-toxigenic isolates was shown by their ability to acquire the major toxin gene, CTX, via CTX $\Phi$  infection. In overall, a surprisingly high percent of TCP-negative strains could be infected with CTX $\Phi$  (11%), although at a lower efficiency than for TCP-positive strains. Among the TCP-negative strains tested, members of *V. mimicus* cluster presented the highest pathogenic risk, with respect to the other clusters because of the very high percent of strains susceptible to phage infection (42%) and *stn* positive strains (55%). The next highest risk group were strains with characteristics related to toxigenic strains (despite the lack of

CTX and TCP), i.e., strains isolated from clinical cases, strains in serogroup O1, and strains in toxigenic clusters, all showed high percent of strains susceptible to the CTX $\Phi$  (18-67%). Higher risk was also associated with the luminescent clusters because in addition to susceptibility to the CTX $\Phi$  infection (14% in clusterA1), high percent of strains in luminescent clusters also possess other toxin gene, *stn*, (33-58%). Strains in Clusters B and C presented a lower risk, with 4 % and 0% susceptible strains in each cluster, respectively, no *stn* positive strains, and a lack of the toxin regulatory gene *toxR*.

The population structure of *V. cholerae*, as shown in this study, is concluded to yield better understanding of natural populations of this species in the environment, expanding the species definition to include *V. mimicus*, and showing diversity and allowing estimation of potential public health risk associated with each of its subpopulations. For the future, specific micro-habitats, function, and association among strains of different cluster could provide even better understanding of the roles of each subpopulation in contributing to the emergence of new toxigenic strains.



## REFERENCES

1. Aldova, E., K. Laznicko, E. Stepanko, and J. Lietava. 1968. Isolation of nonagglutinable vibrios from an enteritis outbreak in Czechoslovakia. *J. Infect. Dis.* **118**:25-31.
2. Alm, R. A., U. H. Stroehler, and P. A. Manning. 1988. Extracellular proteins of *Vibrio cholerae*: nucleotide sequence of the structural gene (*hlyA*) for the haemolysin of the haemolytic El Tor strain O17 and characterization of the *hlyA* mutation in the non-haemolytic classical strain 569B. *Mol. Microbiol.* **2**:481-488.
3. Amann, R. I., W. Ludwig, and K. H. Schleifer. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**:143-169.
4. Antonellis, A., S. C. Curtis, M. G. Pezzolesi, D. K. Moczulski, and A. S. Krolewski. 1999. Rapid identification of polymorphisms in genomic DNA using long range PCR and transposon based sequencing. *Am. J. Hum. Genet.* **65**:A220-A220.
5. Arita, M., T. Honda, T. Miwatani, T. Takeda, T. Takao, and Y. Shimonishi. 1991. Purification and characterization of a heat-stable enterotoxin of *Vibrio mimicus*. *FEMS Microbiol. Lett.* **79**:105-110.
6. Arita, M., T. Takeda, T. Honda, and T. Miwatani. 1986. Purification and characterization of *Vibrio cholerae* non-O1 heat-stable enterotoxin. *Infect. Immun.* **52**:45-49.
7. Attridge, S. R., E. Voss, and P. A. Manning. 1993. The role of toxin co-regulated pili in the pathogenesis of *Vibrio cholerae* O1 El Tor. *Microb. Pathog.* **15**:421-431.
8. Bagchi, K., P. Echeverria, J. D. Arthur, O. Sethabutr, O. Serichantalergs, and C. W. Hoge. 1993. Epidemic of diarrhea caused by *Vibrio cholerae* non-O1 that produced heat-stable toxin among Khmers in a camp in Thailand. *J. Clin. Microbiol.* **31**:1315-1317.
9. Barrett, T. J., and P. A. Blake. 1981. Epidemiological usefulness of changes in hemolytic activity of *Vibrio cholerae* biotype El Tor during the seventh pandemic. *J. Clin. Microbiol.* **13**:126-129.
10. Barua, D. 1992. History of cholera, p. 1-35. *In* W. B. Greenough (ed.), *Cholera*. Plenum Publishing, New York.
11. Barua, D., and W. B. Greenough III. 1992. *Cholera*. Plenum Publishing Corp, New York, NY.

12. Baudry, B., A. Fasano, J. Ketley, and J. B. Kaper. 1992. Cloning of a gene (*zot*) encoding a new toxin produced by *Vibrio cholerae*. *Infect. Immun.* **60**:428-434.
13. Baumann, P., and R. H. W. Schubert. 1984. Family II. *Vibrionaceae*, p. 516-550. *In* J. G. Holt (ed.), *Bergey's Manual of Systematic Bacteriology*, vol. 1. Williams & Wilkins, Baltimore.
14. Beltran, P., G. Delgado, A. Navarro, F. Trujilto, R. K. Selander, and A. Cravioto. 1999. Genetic diversity and population structure of *Vibrio cholerae*. *J. Clin. Microbiol.* **37**:581-590.
15. Bik, E. M., A. E. Bunschoten, R. D. Gouw, and F. R. Mooi. 1995. Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. *EMBO J.* **14**:209-216.
16. Bik, E. M., A. E. Bunschoten, R. J. L. Willems, A. C. Y. Chang, and F. R. Mooi. 1996. Genetic organization and functional analysis of the *otn* DNA essential for cell-wall polysaccharide synthesis in *Vibrio cholerae* O139. *Mol. Microbiol.* **20**:799-811.
17. Blake, P. A. 1994. Historical perspectives on pandemic cholera, p. 293-295. *In* O. Olsvik (ed.), *Vibrio cholerae* and Cholera: molecular to global perspectives. American Society for Microbiology Press, Washington, D. C.
18. Boyd, E. F., K. E. Moyer, L. Shi, and M. K. Waldor. 2000. Infectious CTX $\Phi$  and the vibrio pathogenicity island prophage in *Vibrio mimicus*: evidence for recent horizontal transfer between *V. mimicus* and *V. cholerae*. *Infect. Immun.* **68**:1507-1513.
19. Brenner, D. J., L. W. Mayer, G. M. Carlone, L. H. Harrison, W. F. Bibb, *et al.* 1988. Biochemical, genetic, and epidemiologic characterization of *Hemophilus influenzae* biogroup aegyptius (*Haemophilus aegyptius*) strains associated with Brazilian purpuric fever. *J. Clin. Microbiol.* **26**:1524-1534.
20. Brumlik, M. J., U. Szymajda, D. Zakowska, X. D. Liang, R. J. Redkar, *et al.* 2001. Use of long-range repetitive element polymorphism-PCR to differentiate *Bacillus anthracis* strains. *Appl. Environ. Microbiol.* **67**:3021-3028.
21. Byun, R., L. D. H. Elbourne, R. T. Lan, and P. R. Reeves. 1999. Evolutionary relationships of pathogenic clones of *Vibrio cholerae* by sequence analysis of four housekeeping genes. *Infect. Immun.* **67**:1116-1124.
22. Cameron, D. N., F. M. Khambaty, I. K. Wachsmuth, R. V. Tauxe, and T. J. Barrett. 1994. Molecular characterization of *Vibrio cholerae* O1 strains by pulsed-field gel electrophoresis. *J. Clin. Microbiol.* **32**:1685-1690.
23. Chakraborty, S., P. Garg, T. Ramamurthy, M. Thungapathra, J. K. Gautam, *et al.* 2001. Comparison of antibiogram, virulence genes, ribotypes and DNA

- fingerprints of *Vibrio cholerae* of matching serogroups isolated from hospitalised diarrhoea cases and from the environment during 1997-1998 in Calcutta, India. *J. Med. Microbiol.* **50**:879-888.
24. Chakraborty, S., A. K. Mukhopadhyay, R. K. Bhadra, A. N. Ghosh, R. Mitra, *et al.* 2000. Virulence genes in environmental strains of *Vibrio cholerae*. *Appl. Environ. Microbiol.* **66**:4022-8.
  25. Cho, J. C., and J. M. Tiedje. 2000. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66**:5448-5456.
  26. Cholera Working Group, I. C. D. D. R. B., M. J. Albert, M. Ansaruzzaman, P. K. Bardhan, A. S. G. Faruque, *et al.* 1993. Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. *Lancet* **342**:387-390.
  27. Choopun, N., V. Louis, A. Huq, and R. R. Colwell. 2002. Simple procedure for rapid identification of *Vibrio cholerae* from the aquatic environment. *Appl. Environ. Microbiol.* **68**:995-998.
  28. Chun, J., A. Huq, and R. R. Colwell. 1999. Analysis of 16S-23S rRNA intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. *Appl. Environ. Microbiol.* **65**:2202-2208.
  29. Clark, C. A., L. Purins, P. Kaewrakon, T. Focareta, and P. A. Manning. 2000. The *Vibrio cholerae* O1 chromosomal integron. *Microbiology* **146**:2605-2612.
  30. Clark, C. A., L. Purins, P. Kaewrakon, and P. A. Manning. 1997. VCR repetitive sequence elements in the *Vibrio cholerae* chromosome constitute a mega-integron. *Mol. Microbiol.* **26**:1137-1138.
  31. Clerc, A., C. Manceau, and X. Nesme. 1998. Comparison of randomly amplified polymorphic DNA with amplified fragment length polymorphism to assess genetic diversity and genetic relatedness within genospecies III of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* **64**:1180-1187.
  32. Colombo, M. M., S. Mastrandrea, F. Leite, A. Santona, S. Uzzau, *et al.* 1997. Tracking of clinical and environmental *Vibrio cholerae* O1 strains by combined analysis of the presence of the toxin cassette, plasmid content and ERIC PCR. *FEMS Immunol. Med. Microbiol.* **19**:33-45.
  33. Colwell, R. R. 1973. Genetic and phenetic classification of bacteria. *Adv. Appl. Microbiol.* **16**:137-195.
  34. Colwell, R. R., J. Kaper, and S. W. Joseph. 1977. *Vibrio cholerae*, *Vibrio parahaemolyticus*, and other vibrios: occurrence and distribution in Chesapeake Bay. *Science* **198**:394-396.

35. Colwell, R. R., and W. M. Spira. 1992. The ecology of *V. cholerae*, p. 107-127. In D. B. a. W. B. G. III (ed.), *Cholera*. Plenum Medical Book Co., New York, N.Y.
36. Coyne, V. E., and L. Al-Harthi. 1992. Induction of Melanin Biosynthesis in *Vibrio cholerae*. *Appl. Environ. Microbiol* **58**:2861-2865.
37. Czyz, A., K. Plata, and G. Wegrzyn. 2003. Stimulation of DNA repair as an evolutionary drive for bacterial luminescence. *Luminescence* **18**:140-144.
38. Czyz, A., B. Wrobel, and G. Wegrzyn. 2000. *Vibrio harveyi* bioluminescence plays a role in stimulation of DNA repair. *Microbiology* **146**:283-288.
39. Dabora, S. L., D. Franz, S. Jozwiak, and D. J. Kwiatkowski. 1999. Identification and characterization of 4 large (1.4 kb-34 kb) deletions in TSC2 using long range PCR suggests diverse deletion mechanisms. *Am. J. Hum. Genet.* **65**:A108.
40. Dalsgaard, A., M. J. Albert, D. N. Taylor, T. Shimada, R. Meza, *et al.* 1995. Characterization of *Vibrio cholerae* non-O1 serogroups obtained from an outbreak of diarrhea in Lima, Peru. *J. Clin. Microbiol.* **33**:2715-2722.
41. Dalsgaard, A., O. Serichantalergs, A. Forslund, W. Lin, J. Mekalanos, *et al.* 2001. Clinical and environmental isolates of *Vibrio cholerae* serogroup O141 carry the CTX phage and the genes encoding the toxin-coregulated pili. *J. Clin. Microbiol.* **39**:4086-4092.
42. Dalsgaard, A., O. Serichantalergs, T. Shimada, O. Sethabutr, and P. Echeverria. 1995. Prevalence of *Vibrio cholerae* with heat-stable enterotoxin (NAG-ST) and cholera toxin genes; restriction fragment length polymorphisms of NAG-ST genes among *V. cholerae* O serogroups from a major shrimp production area in Thailand. *J. Med. Microbiol.* **43**:216-220.
43. Davis, B. M., H. H. Kimsey, A. V. Kane, and M. K. Waldor. 2002. A satellite phage-encoded antirepressor induces repressor aggregation and cholera toxin gene transfer. *EMBO J.* **21**:4240-4249.
44. Davis, B. M., K. E. Moyer, E. F. Boyd, and M. K. Waldor. 2000. CTX prophages in classical biotype *Vibrio cholerae*: functional phage genes but dysfunctional phage genomes. *J. Bacteriol.* **182**:6992-6998.
45. Davis, B. M., and M. K. Waldor. 2000. CTX $\Phi$  contains a hybrid genome derived from tandemly integrated elements. *Proc. Natl. Acad. Sci. U. S. A.* **97**:8572-8577.
46. Davis, B. R., R. Fanning, J. M. Madden, A. G. Steigerwalt, H. B. J. Bradford, *et al.* 1981. Characterization of biochemically atypical *Vibrio cholerae* strains and designation of a new pathogenic species, *Vibrio mimicus*. *J. Clin. Microbiol.* **14**:631-639.

47. de Bruijn, F. J. 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl. Environ. Microbiol.* **58**:2180-2187.
48. DePaola, A., M. L. Motes, D. W. Cook, J. Veazey, W. E. Garthright, and R. Blodgett. 1997. Evaluation of an alkaline phosphatase-labeled DNA probe for enumeration of *Vibrio vulnificus* in Gulf Coast oysters. *J. Microbiol. Methods* **29**:115-120.
49. DePaola, A., J. L. Nordstrom, J. C. Bowers, J. G. Wells, and D. W. Cook. 2003. Seasonal abundance of total and pathogenic *Vibrio parahaemolyticus* in Alabama oysters. *Appl. Environ. Microbiol.* **69**:1521-1526.
50. DiRita, V. J. 1992. Co-ordinate expression of virulence genes by ToxR in *Vibrio cholerae*. *Mol. Microbiol.* **6**:451-458.
51. Duby, A., K. A. Jacobs, and A. Celeste. 1988. Using synthetic oligonucleotides as probes, p. 6.4.1-6.4.10. *In* K. Struhl (ed.), *Current protocols in molecular biology*, vol. 1. John Wiley & Sons, New York.
52. Ellison, R. K., E. Malnati, A. DePaola, J. Bowers, and G. E. Rodrick. 2001. Populations of *Vibrio parahaemolyticus* in retail oysters from Florida using two methods. *J. Food Prot.* **64**:682-686.
53. Farfan, M., D. Minana, M. C. Fuste, and J. G. Loren. 2000. Genetic relationships between clinical and environmental *Vibrio cholerae* isolates based on multilocus enzyme electrophoresis. *Microbiology* **146**:2613-2626.
54. Farfan, M., D. Minana-Galbis, M. C. Fuste, and J. G. Loren. 2002. Allelic diversity and population structure in *Vibrio cholerae* O139 Bengal based on nucleotide sequence analysis. *J. Bacteriol.* **184**:1304-1313.
55. Farmer III, J. J., F. W. Hickman-Brenner, and M. T. Kelly. 1985. *Vibrio*, p. 282-301. *In* H. J. Shadomy (ed.), *Manual of Clinical Microbiology*, 4 th ed. American Society for Microbiology, Washington, D.C.
56. Faruque, S. M., K. M. Ahmed, A. R. Abdul Alim, F. Qadri, A. K. Siddique, and M. J. Albert. 1997. Emergence of a new clone of toxigenic *Vibrio cholerae* O1 biotype El Tor displacing *V. cholerae* O139 Bengal in Bangladesh. *J. Clin. Microbiol.* **35**:624-630.
57. Faruque, S. M., K. M. Ahmed, A. K. Siddique, K. Zaman, A. R. Alim, and M. J. Albert. 1997. Molecular analysis of toxigenic *Vibrio cholerae* O139 Bengal strains isolated in Bangladesh between 1993 and 1996: evidence for emergence of a new clone of the Bengal vibrios. *J. Clin. Microbiol.* **35**:2299-2306.

58. Faruque, S. M., M. J. Albert, and J. J. Mekalanos. 1998. Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.* **62**:1301-1314.
59. Faruque, S. M., Asadulghani, A. R. Alim, M. J. Albert, K. M. Islam, and J. J. Mekalanos. 1998. Induction of the lysogenic phage encoding cholera toxin in naturally occurring strains of toxigenic *Vibrio cholerae* O1 and O139. *Infect. Immun.* **66**:3752-3757.
60. Faruque, S. M., Asadulghani, M. M. Rahman, M. K. Waldor, and D. A. Sack. 2000. Sunlight-induced propagation of the lysogenic phage encoding cholera toxin. *Infect. Immun.* **68**:4795-801.
61. Faruque, S. M., Asadulghani, M. N. Saha, A. R. Alim, M. J. Albert, *et al.* 1998. Analysis of clinical and environmental strains of nontoxigenic *Vibrio cholerae* for susceptibility to CTX $\Phi$ : molecular basis for origination of new strains with epidemic potential. *Infect. Immun.* **66**:5819-5825.
62. Faruque, S. M., M. M. Rahman, Asadulghani, K. M. Nasirul Islam, and J. J. Mekalanos. 1999. Lysogenic conversion of environmental *Vibrio mimicus* strains by CTX $\Phi$ . *Infect. Immun.* **67**:5723-5729.
63. Faruque, S. M., S. K. Roy, A. R. Alim, A. K. Siddique, and M. J. Albert. 1995. Molecular epidemiology of toxigenic *Vibrio cholerae* in Bangladesh studied by numerical analysis of rRNA gene restriction patterns. *J. Clin. Microbiol.* **33**:2833-2838.
64. Faruque, S. M., J. Zhu, Asadulghani, M. Kamruzzaman, and J. J. Mekalanos. 2003. Examination of diverse toxin-coregulated pilus-positive *Vibrio cholerae* strains fails to demonstrate evidence for vibrio pathogenicity island phage. *Infect. Immun.* **71**:2993-2999.
65. Fasano, A., B. Baudry, D. W. Pumphlin, S. S. Wasserman, B. D. Tall, *et al.* 1991. *Vibrio cholerae* produces a second enterotoxin, which affects intestinal tight junctions. *Proc. Natl. Acad. Sci. U. S. A.* **88**:5242-5246.
66. Felsenstein, J. 2004. PHYLIP version 3.6. Department of Genome Sciences, University of Washington, Seattle, WA.  
<http://evolution.genetics.washington.edu/phylip.html> [Online.]
67. Fields, P. I., T. Popovic, K. Wachsmuth, and O. Olsvik. 1992. Use of polymerase chain reaction for detection of toxigenic *Vibrio cholerae* O1 strains from the Latin American cholera epidemic. *J. Clin. Microbiol.* **30**:2118-2121.
68. Fitch, W. M., and M. Margoliash. 1967. Construction of phylogenetic trees. *Science* **155**:279-284.

69. Garg, P., A. Aydanian, D. Smith, J. G. Morris, G. B. Nair, and O. C. Stine. 2003. Molecular epidemiology of O139 *Vibrio cholerae*: mutation, lateral gene transfer, and founder flush. *Emerg. Infect. Dis.* **9**:810-814.
70. Goldberg, I., and J. J. Mekalanos. 1986. Effect of a *recA* mutation on cholera toxin gene amplification and deletion events. *J. Bacteriol.* **165**:723-731.
71. Gooch, J. A., A. DePaola, J. Bowers, and D. L. Marshall. 2002. Growth and survival of *Vibrio parahaemolyticus* in postharvest American oysters. *J. Food Prot.* **65**:970-974.
72. Gooch, J. A., A. DePaola, C. A. Kaysner, and D. L. Marshall. 2001. Evaluation of two direct plating methods using nonradioactive probes for enumeration of *Vibrio parahaemolyticus* in oysters. *Appl. Environ. Microbiol.* **67**:721-724.
73. Grim, C. J., N. Choopun, A. Huq, and R. R. Colwell. 2003. A new method for identifying bioluminescent *Vibrio cholerae*. Presented at the American Society for Microbiology, 103 th General Meeting, Washington, D. C.
74. Grimont, F., and P. A. D. Grimont. 1986. Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. *Ann. Inst. Pasteur Microbiol.* **137B**:165-175.
75. Guglielmetti, P., L. Bravo, A. Zanchi, R. Monte, G. Lombardi, and G. M. Rossolini. 1994. Detection of the *Vibrio cholerae* heat-stable enterotoxin gene by polymerase chain reaction. *Mol. Cell. Probes* **8**:39-44.
76. Hammer, B. K., and B. L. Bassler. 2003. Quorum sensing controls biofilm formation in *Vibrio cholerae*. *Mol. Microbiol.* **50**:101-114.
77. Hane, B. G., K. Jager, and H. G. Drexler. 1993. The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. *Electrophoresis* **14**:967-972.
78. Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, *et al.* 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477-483.
79. Heidelberg, J. F., K. B. Heidelberg, and R. R. Colwell. 2002. Bacteria of the  $\gamma$ -subclass proteobacteria associated with zooplankton in Chesapeake Bay. *Appl. Environ. Microbiol.* **68**:5498-5507.
80. Heidelberg, J. F., K. B. Heidelberg, and R. R. Colwell. 2002. Seasonality of Chesapeake Bay bacterioplankton species. *Appl. Environ. Microbiol.* **68**:5488-5497.
81. Heilpern, A. J., and M. K. Waldor. 2000. CTX $\Phi$  infection of *Vibrio cholerae* requires the *tolQRA* gene products. *J. Bacteriol.* **182**:1739-1747.

82. Herrington, D. A., R. H. Hall, G. Losonsky, J. J. Mekalanos, R. K. Taylor, and M. M. Levine. 1988. Toxin, toxin-coregulated pili, and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J. Exp. Med.* **168**:1487-1492.
83. Hoge, C. W., O. Sethabutr, L. Bodhidatta, P. Echeverria, D. C. Robertson, and J. G. Morris. 1990. Use of a synthetic oligonucleotide probe to detect strains of nonserovar O1 *Vibrio cholerae* carrying the gene for heat-stable enterotoxin (Nag-St). *J. Clin. Microbiol.* **28**:1473-1476.
84. Honda, T., and R. A. Finkelstein. 1979. Purification and characterization of a hemolysin produced by *Vibrio cholerae* biotype El Tor: another toxic-substance produced by cholera vibrios. *Infect. Immun.* **26**:1020-1027.
85. Hulton, C. S. J., C. F. Higgins, and P. M. Sharp. 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol. Microbiol.* **5**:825-834.
86. Huq, A., E. B. Small, P. A. West, and R. R. Colwell. 1984. The role of planktonic copepods in the survival and multiplication of *Vibrio cholerae* in the aquatic environment, p. 521-534. *In* R. R. Colwell (ed.), *Vibrios in the environment*. John Wiley & Sons, New York.
87. Huq, A., E. B. Small, P. A. West, M. I. Huq, R. Rahman, and R. R. Colwell. 1984. Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. *Appl. Environ. Microbiol.* **48**:420-424.
88. Huys, G., R. Coopman, P. Janssen, and K. Kersters. 1996. High-resolution genotypic analysis of the genus *Aeromonas* by AFLP fingerprinting. *Int. J. Syst. Bacteriol.* **46**:572-580.
89. Ichinose, Y., K. Yamamoto, N. Nakasone, M. J. Tanabe, T. Takeda, *et al.* 1987. Enterotoxicity of El Tor-like hemolysin of non-O1 *Vibrio cholerae*. *Infect. Immun.* **55**:1090-1093.
90. Ivins, B. E., and R. K. Holmes. 1980. Isolation and characterization of melanin-producing (*mel*) mutants of *Vibrio cholerae*. *Infect. Immun.* **27**:721-729.
91. Janssen, P., R. Coopman, G. Huys, J. Swings, M. Bleeker, *et al.* 1996. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142**:1881-1893.
92. Jiang, S. C., and W. Fu. 2001. Seasonal abundance and distribution of *Vibrio cholerae* in coastal waters quantified by a 16S-23S intergenic spacer probe. *Microb. Ecol.* **42**:540-548.
93. Jiang, S. C., M. Matte, G. Matte, A. Huq, and R. R. Colwell. 2000. Genetic diversity of clinical and environmental isolates of *Vibrio cholerae* determined by



- amplified fragment length polymorphism fingerprinting. *Appl. Environ. Microbiol.* **66**:148-153.
94. Jones, C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, *et al.* 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed.* **3**:381-390.
  95. Kamal, A. M. 1971. Outbreak of gastroenteritis by non-agglutinable (NAG) vibrios in the republic of the Sudan. *J. Egypt. Public Health Assoc.* **46**:125-133.
  96. Kaper, J., H. Lockman, R. R. Colwell, and S. W. Joseph. 1979. Ecology, serology, and enterotoxin production of *Vibrio cholerae* in Chesapeake Bay. *Appl. Environ. Microbiol.* **37**:91-103.
  97. Kaper, J. B., J. G. Morris, and M. M. Levine. 1995. Cholera. *Clin. Microbiol. Rev.* **8**:48-86.
  98. Karaolis, D. K. R., R. T. Lan, and P. R. Reeves. 1994. Molecular evolution of the seventh-pandemic clone of *Vibrio cholerae* and its relationship to other pandemic and epidemic *V. cholerae* isolates. *J. Bacteriol.* **176**:6199-6206.
  99. Karaolis, D. K. R., S. Somara, D. R. Maneval, J. A. Johnson, and J. B. Kaper. 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**:375-379.
  100. Kay, B. A., C. A. Bopp, and J. G. Wells. 1994. Isolation and identification of *Vibrio cholerae* O1 from fecal specimens, p. 3-20. *In* O. Olsvik (ed.), *Vibrio cholerae* and Cholera: molecular to global perspectives. ASM Press, Washington, D. C.
  101. Keasler, S. P., and R. H. Hall. 1993. Detecting and biotyping *Vibrio cholerae* O1 with multiplex polymerase chain reaction. *Lancet* **341**:1661-1661.
  102. Kim, G. T., J. Y. Lee, S. H. Huh, J. H. Yu, and I. S. Kong. 1997. Nucleotide sequence of the *vmhA* gene encoding hemolysin from *Vibrio mimicus*. *Biochim. Biophys. Acta.* **1360**:102-104.
  103. Kim, J. 1993. Improving the accuracy of phylogenetic estimation by combining different methods. *Syst. Biol.* **42**:331-340.
  104. Koblavi, S., F. Grimont, and P. A. D. Grimont. 1990. Clonal diversity of *Vibrio cholerae* O1 evidenced by rRNA gene restriction patterns. *Res. Microbiol.* **141**:645-657.
  105. Koch, R. 1884. An address on cholera and its bacillus. *Br. Med. J.* **2**:403-407.
  106. Kotob, S. I., S. L. Coon, E. J. Quintero, and R. M. Weiner. 1995. Homogentisic acid is the primary precursor of melanin synthesis in *Vibrio cholerae*, a

- Hyphomonas* strain, and *Shewanella colwelliana*. Appl. Environ. Microbiol. **61**:1620-1622.
107. Legendre, P., and L. Legendre. 1998. Numerical Ecology, 2nd English ed. Elsevier Science BV, Amsterdam.
  108. Lennette, E. H., A. Balows, W. J. J. Hausler, and H. J. Shadomy. 1985. Manual of Clinical Microbiology, 4th ed. American Society for Microbiology, Washington, D.C.
  109. Li, M. R., T. Shimada, J. G. Morris, A. Sulakvelidze, and S. Sozhamannan. 2002. Evidence for the emergence of non-O1 and non-O139 *Vibrio cholerae* strains with pathogenic potential by exchange of O-antigen biosynthesis regions. Infect. Immun. **70**:2441-2453.
  110. Lipp, E. K., I. N. G. Rivera, A. Gil, E. M. Espeland, N. Choopun, *et al.* 2003. Direct detection of *Vibrio cholerae* and *ctxA* in Peruvian coastal water and plankton by PCR. Appl. Environ. Microbiol. **69**:3676-3680.
  111. Louis, V. R., E. Russek-Cohen, N. Choopun, I. N. G. Rivera, B. Gangle, *et al.* 2003. Predictability of *Vibrio cholerae* in Chesapeake Bay. Appl. Environ. Microbiol. **69**:2773-2785.
  112. Louws, F. J., J. Bell, C. M. Medina-Mora, C. D. Smart, D. Opgenorth, *et al.* 1998. rep-PCR-mediated genomic fingerprinting: a rapid and effective method to identify *Clavibacter michiganensis*. Phytopathology **88**:862-868.
  113. MacFaddin, J. F. 1980. Biochemical tests for identification of medical bacteria. Williams & Wilkins, Baltimore.
  114. MacPherson, J. M., P. E. Eckstein, G. J. Scoles, and A. A. Gajadhar. 1993. Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. Mol. Cell. Probes **7**:293-299.
  115. Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, *et al.* 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. **95**:3140-3145.
  116. Manning, P. A., E. J. Bartowsky, D. I. Leavesly, J. A. Hackett, and M. W. Heuzenroeder. 1985. Molecular cloning using immune sera of a 22-kDal minor outer membrane protein of *Vibrio cholerae*. Gene **34**:95-103.
  117. Manning, P. A., C. A. Clark, and T. Focareta. 1999. Gene capture in *Vibrio cholerae*. Trends Microbiol. **7**:93-95.

118. Marshall, S., C. G. Clark, G. Wang, M. Mulvey, M. T. Kelly, and W. M. Johnson. 1999. Comparison of molecular methods for typing *Vibrio parahaemolyticus*. J. Clin. Microbiol. **37**:2473-2478.
119. Massad, G., and J. D. Oliver. 1987. New selective and differential medium for *Vibrio cholerae* and *Vibrio vulnificus*. Appl. Environ. Microbiol. **53**:2262-2264.
120. Mazel, D., B. Dychinco, V. A. Webb, and J. Davies. 1998. A distinctive class of intergron in the *Vibrio cholerae* genome. Science **280**:605-608.
121. McDonald, O. G., E. Y. Krynetski, and W. E. Evans. 2002. Molecular haplotyping of genomic DNA for multiple single-nucleotide polymorphisms located kilobases apart using long-range polymerase chain reaction and intramolecular ligation. Pharmacogenetics **12**:93-99.
122. McNicol, L. A., S. P. DE, J. B. Kaper, P. A. West, and R. R. Colwell. 1983. Numerical taxonomy of *Vibrio cholerae* and related species isolated from areas that are endemic and nonendemic for cholera. J. Clin. Microbiol. **17**:1102-1113.
123. Merrell, D. S., C. Bailey, J. B. Kaper, and A. Camilli. 2001. The ToxR-mediated organic acid tolerance response of *Vibrio cholerae* requires OmpU. J. Bacteriol. **183**:2746-2754.
124. Meunier, J. R., and P. A. D. Grimont. 1993. Factors affecting reproducibility of random amplified polymorphic DNA-fingerprinting. Res. Microbiol. **144**:373-379.
125. Mingeot-Leclercq, M. P., Y. Glupczynski, and P. M. Tulkens. 1999. Aminoglycosides: activity and resistance. Antimicrob. Agents Chemother. **43**:727-737.
126. Mooi, F. R., and E. M. Bik. 1997. The evolution of epidemic *Vibrio cholerae* strains. Trends Microbiol. **5**:161-165.
127. Morris, J. G., T. Takeda, B. D. Tall, G. A. Losonsky, S. K. Bhattacharya, *et al.* 1990. Experimental non-O group 1 *Vibrio cholerae* gastroenteritis in humans. J. Clin. Invest. **85**:697-705.
128. Musser, J. M., P. M. Schlievert, A. W. Chow, P. Ewan, B. N. Kreiswirth, *et al.* 1990. A single clone of *Staphylococcus aureus* causes the majority of cases of toxic shock syndrome. Proc. Natl. Acad. Sci. U. S. A. **87**:225-229.
129. Nandi, B., R. K. Nandy, S. Mukhopadhyay, G. B. Nair, T. Shimada, and A. C. Ghose. 2000. Rapid method for species-specific identification of *Vibrio cholerae* using primers targeted to the gene of outer membrane protein OmpW. J. Clin. Microbiol. **38**:4145-4151.

130. Ogawa, A., J. Kato, H. Watanabe, B. G. Nair, and T. Takeda. 1990. Cloning and nucleotide sequence of a heat-stable enterotoxin gene from *Vibrio cholerae* non-O1 isolated from a patient with traveler's diarrhea. *Infect. Immun.* **58**:3325-3329.
131. Ogawa, A., and T. Takeda. 1993. The gene encoding the heat-stable enterotoxin of *Vibrio cholerae* is flanked by 123-base pair direct repeats. *Microbiol. Immunol.* **37**:607-616.
132. Osek, J., A. M. Svennerholm, and J. Holmgren. 1992. Protection against *Vibrio cholerae* El Tor infection by specific antibodies against mannose-binding hemagglutinin pili. *Infect. Immun.* **60**:4961-4964.
133. O'Shea, Y. A., and E. F. Boyd. 2002. Mobilization of the *Vibrio* pathogenicity island between *Vibrio cholerae* isolates mediated by CP-T1 generalized transduction. *Fems Microbiol. Lett.* **214**:153-157.
134. Pal, A., T. Ramamurthy, R. K. Bhadra, T. Takeda, T. Shimada, *et al.* 1992. Reassessment of the prevalence of heat-stable enterotoxin (NAG-ST) among environmental *Vibrio cholerae* non-O1 strains isolated from Calcutta, India, by using a NAG-ST DNA probe. *Appl. Environ. Microbiol.* **58**:2485-2489.
135. Palmer, L. M., and R. R. Colwell. 1991. Detection of luciferase gene sequence in nonluminescent *Vibrio cholerae* by colony hybridization and polymerase chain reaction. *Appl. Environ. Microbiol.* **57**:1286-1293.
136. Pearson, G. D. N., A. Woods, S. L. Chiang, and J. J. Mekalanos. 1993. CTX genetic element encodes a site-specific recombination system and an intestinal colonization factor. *Proc. Natl. Acad. Sci. U. S. A.* **90**:3750-3754.
137. Peterson, K. M., and J. J. Mekalanos. 1988. Characterization of the *Vibrio cholerae* ToxR regulon: identification of novel genes involved in intestinal colonization. *Infect. Immun.* **56**:2822-2829.
138. Pils, H., D. Smajs, and V. Braun. 1999. Characterization of colicin S4 and its receptor, OmpW, a minor protein of the *Escherichia coli* outer membrane. *J. Bacteriol.* **181**:3578-3581.
139. Politzer, R. 1959. Cholera. World Health Organization, Geneva.
140. Popovic, T., C. Bopp, O. Olsvik, and K. Wachsmuth. 1993. Epidemiologic application of a standardized ribotype scheme for *Vibrio cholerae* O1. *J. Clin. Microbiol.* **31**:2474-2482.
141. Provenzano, D., and K. E. Klose. 2000. Altered expression of the ToxR-regulated porins OmpU and OmpT diminishes *Vibrio cholerae* bile resistance, virulence factor expression, and intestinal colonization. *Proc. Natl. Acad. Sci. U. S. A.* **97**:10220-10224.

142. Provenzano, D., D. A. Schuhmacher, J. L. Barker, and K. E. Klose. 2000. The virulence regulatory protein ToxR mediates enhanced bile resistance in *Vibrio cholerae* and other pathogenic *Vibrio* species. *Infect. Immun.* **68**:1491-1497.
143. Rademaker, J. L. W., B. Hoste, F. J. Louws, K. Kersters, J. Swings, *et al.* 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *Int. J. Syst. Evol. Microbiol.* **50**:665-677.
144. Rademaker, J. L. W., F. J. Louws, and F. J. de Bruijn. 1998. Characterization of the diversity of ecologically important microbes by rep-PCR fingerprinting, p. 1-26. *In* F. J. de Bruijn (ed.), *Molecular microbial ecology manual*, suppl.3. Kluwer Academic Publishers, Dordrecht, The Netherlands.
145. Ramaiah, N., J. Chun, J. Ravel, W. L. Straube, R. T. Hill, and R. R. Colwell. 2000. Detection of luciferase gene sequences in nonluminescent bacteria from the Chesapeake Bay. *FEMS microbiol. Ecol.* **33**:27-34.
146. Ramamurthy, T., S. Garg, R. Sharma, S. K. Bhattacharya, G. B. Nair, *et al.* 1993. Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. *Lancet* **341**:703-704.
147. Reeves, P. R., M. Hobbs, M. A. Valvano, M. Skurnik, C. Whitfield, *et al.* 1996. Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol.* **4**:495-503.
148. Richardson, K., J. Michalski, and J. B. Kaper. 1986. Hemolysin production and cloning of two hemolysin determinants from classical *Vibrio cholerae*. *Infect. Immun.* **54**:415-420.
149. Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, *et al.* 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* **308**:681-685.
150. Rivera, I. G., M. A. Chowdhury, A. Huq, D. Jacobs, M. T. Martins, and R. R. Colwell. 1995. Enterobacterial repetitive intergenic consensus sequences and the PCR to generate fingerprints of genomic DNAs from *Vibrio cholerae* O1, O139, and non-O1 strains. *Appl. Environ. Microbiol.* **61**:2898-2904.
151. Rivera, I. N. G., J. Chun, A. Huq, R. B. Sack, and R. R. Colwell. 2001. Genotypes associated with virulence in environmental isolates of *Vibrio cholerae*. *Appl. Environ. Microbiol.* **67**:2421-2429.
152. Rivera, I. N. G., E. K. Lipp, A. Gil, N. Choopun, A. Huq, and R. R. Colwell. 2003. Method of DNA extraction and application of multiplex polymerase chain reaction to detect toxigenic *Vibrio cholerae* O1 and O139 from aquatic ecosystems. *Environ. Microbiol.* **5**:599-606.

153. Robert-Pillot, A., S. Baron, J. Lesne, J. M. Fournier, and M. L. Quilici. 2002. Improved specific detection of *Vibrio cholerae* in environmental water samples by culture on selective medium and colony hybridization assay with an oligonucleotide probe. *FEMS Microbiol. Ecol.* **40**:39-46.
154. Roberts, N. C., R. J. Siebeling, J. B. Kaper, and H. B. Bradford. 1982. Vibrios in the Louisiana Gulf Coast environment. *Microbiol. Ecol.* **8**:299-312.
155. Rodrigue, D. C., T. Popovic, and I. K. Wachsmuth. 1994. Nontoxigenic *Vibrio cholerae* O1 infections in the United States, p. 69-76. *In* O. Olsvik (ed.), *Vibrio cholerae* and cholera: molecular to global perspectives. ASM Press, Washington, D.C.
156. Rosenberg, C. E. 1962. The cholera years, the United States in 1832, 1849, and 1866. University of Chicago Press, Chicago.
157. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406-425.
158. Sambrook, J., and E. F. Fritsch. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, New York.
159. Sarkar, B., T. Bhattacharya, T. Ramamurthy, T. Shimada, Y. Takeda, and G. B. Nair. 2002. Preferential association of the heat-stable enterotoxin gene (*stn*) with environmental strains of *Vibrio cholerae* belonging to the O14 serogroup. *Epidemiol. Infect.* **129**:245-251.
160. Sea Grant, M. 2003. Biofilms and biodiversity: How to calculate biodiversity? Sea Grant, Maryland. <http://www.mdsg.umd.edu/Education/biofilm/diverse.htm> [Online.]
161. Sechi, L. A., I. Dupre, A. Deriu, G. Fadda, and S. Zanetti. 2000. Distribution of *Vibrio cholerae* virulence genes among different *Vibrio* species isolated in Sardinia, Italy. *J. Appl. Microbiol.* **88**:475-481.
162. Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol.* **51**:873-884.
163. Selander, R. K., D. A. Caugant, and T. S. Whittam. 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625-1648. *In* H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
164. Selander, R. K., J. Li, and K. Nelson. 1996. Evolutionary genetics of *Salmonella enterica*, p. 2691-2707. *In* H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, 2nd ed, vol. 2. ASM Press, Washington, D.C.

165. Sengupta, T. K., R. K. Nandy, S. Mukhopadhyay, R. H. Hall, V. Sathyamoorthy, and A. C. Ghose. 1998. Characterization of a 20-kDa pilus protein expressed by a diarrheogenic strain of non-O1/non-O139 *Vibrio cholerae*. FEMS Microbiol. Lett. **160**:183-189.
166. Shangkuan, Y. H., Y. S. Show, and T. M. Wang. 1995. Multiplex polymerase chain reaction to detect toxigenic *Vibrio cholerae* and to biotype *Vibrio cholerae* O1. J. Appl. Bacteriol. **79**:264-273.
167. Shangkuan, Y. H., C. M. Tsao, and H. C. Lin. 1997. Comparison of *Vibrio cholerae* O1 isolates by polymerase chain reaction fingerprinting and ribotyping. J. Med. Microbiol. **46**:941-948.
168. Sharma, C., M. Thungapathra, A. Ghosh, A. K. Mukhopadhyay, A. Basu, *et al.* 1998. Molecular analysis of non-O1, non-O139 *Vibrio cholerae* associated with an unusual upsurge in the incidence of cholera-like disease in Calcutta, India. J. Clin. Microbiol. **36**:756-763.
169. Shi, L., S. Miyoshi, M. Hiura, K. I. Tomochika, T. Shimada, and S. Shinoda. 1998. Detection of genes encoding cholera toxin (CT), zonula occludens toxin (ZOT), accessory cholera enterotoxin (ACE) and heat-stable enterotoxin (ST) in *Vibrio mimicus* clinical strains. Microbiol. Immunol. **42**:823-828.
170. Shimada, T., E. Arakawa, K. Itoh, T. Okitsu, A. Matsushima, *et al.* 1994. Extended serotyping scheme for *Vibrio cholerae*. Curr. Microbiol. **28**:175-178.
171. Sneath, P. H. A., and R. R. Sokal. 1973. Numerical Taxonomy. W. H. Freeman and Company, San Francisco.
172. Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research, 3rd ed. W. H. Freeman and Co., New York.
173. Sperandio, V., C. Bailey, J. A. Giron, V. J. DiRita, W. D. Silveira, *et al.* 1996. Cloning and characterization of the gene encoding the OmpU outer membrane protein of *Vibrio cholerae*. Infect. Immun. **64**:5406-5409.
174. Sperandio, V., J. A. Giron, W. D. Silveira, and J. B. Kaper. 1995. The OmpU outer membrane protein, a potential adherence factor of *Vibrio cholerae*. Infect. Immun. **63**:4433-4438.
175. Spratt, B. G. 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Curr. Opin. Microbiol. **2**:312-316.
176. Spratt, B. G., and M. C. J. Maiden. 1999. Bacterial population genetics, evolution and epidemiology. Philos. Trans. R. Soc. Lond., B, Biol. Sci. **354**:701-710.

177. Stackebrandt, E., and B. M. Goebel. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species in bacteriology. *Int. J. Syst. Bacteriol.* **44**:846-849.
178. Stine, O. C., S. Sozhamannan, Q. Gou, S. Q. Zheng, J. G. Morris, and J. A. Johnson. 2000. Phylogeny of *Vibrio cholerae* based on *recA* sequence. *Infect. Immun.* **68**:7180-7185.
179. Stroehler, U. H., K. E. Jedani, and P. A. Manning. 1998. Genetic organization of the regions associated with surface polysaccharide synthesis in *Vibrio cholerae* O1, O139 and *Vibrio anguillarum* O1 and O2: a review. *Gene* **223**:269-282.
180. Stroehler, U. H., and P. A. Manning. 1997. *Vibrio cholerae* serotype O139: swapping genes for surface polysaccharide biosynthesis. *Trends Microbiol.* **5**:178-180.
181. Szpilewska, H., A. Czyz, and G. Wegrzyn. 2003. Experimental evidence for the physiological role of bacterial luciferase in the protection of cells against oxidative stress. *Curr. Microbiol.* **47**:379-382.
182. Takeda, T., Y. Peina, A. Ogawa, S. Dohi, H. Abe, *et al.* 1991. Detection of heat-stable enterotoxin in a cholera toxin gene-positive strain of *Vibrio cholerae* O1. *FEMS Microbiol. Lett.* **80**:23-27.
183. Thornley, M. J. 1960. The differentiation of *Pseudomonas* from other gram-negative bacteria on the basis of arginine metabolism. *J. Appl. Bacteriol.* **23**:37-52.
184. Versalovic, J., T. Koeuth, and J. R. Lupski. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res.* **19**:6823-6831.
185. Vicente, A. C. P., A. M. Coelho, and C. A. Salles. 1997. Detection of *Vibrio cholerae* and *V. mimicus* heat-stable toxin gene sequence by PCR. *J. Med. Microbiol.* **46**:398-402.
186. Vinuesa, P., J. L. W. Rademaker, F. J. de Bruijn, and D. Werner. 1998. Genotypic characterization of *Bradyrhizobium* strains nodulating endemic woody legumes of the Canary Islands by PCR-restriction fragment length polymorphism analysis of genes encoding 16S rRNA (16S rDNA) and 16S-23S rDNA intergenic spacers, repetitive extragenic palindromic PCR genomic fingerprinting, and partial 16S rDNA sequencing. *Appl. Environ. Microbiol.* **64**:2096-2104.
187. Waldor, M. K., and J. J. Mekalanos. 1994. *Vibrio cholerae* O139 specific gene sequences. *Lancet* **343**:1366.
188. Waldor, M. K., and J. J. Mekalanos. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**:1910-1914.



189. Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, *et al.* 1987. Report of the Ad Hoc Committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* **37**:463-464.
190. West, P. A., P. R. Brayton, T. N. Bryant, and R. R. Colwell. 1986. Numerical taxonomy of *Vibrios* isolated from aquatic environment. *Int. J. Syst. Bacteriol.* **36**:531-543.
191. West, P. A., and R. R. Colwell. 1983. Identification and classification of *Vibrionaceae*-an overview, p. 285-341. *In* M. B. Hatem (ed.), *Vibrios* in the environment. John Wiley & Sons, New York.
192. West, P. A., and J. V. Lee. 1982. Ecology of *Vibrio* spp. including *V. cholerae* in natural waters of Kent, England. *J. Appl. Bacteriol.* **52**:435-448.
193. West, P. A., E. Russek-Cohen, P. R. Brayton, and R. R. Colwell. 1982. Statistical evaluation of a quality control method for isolation of pathogenic *Vibrio* species on selected thiosulfate-citrate-bile salts-sucrose agars. *J. Clin. Microbiol.* **16**:1110-1116.
194. WHO. 1995. Meeting on the potential role of new cholera vaccines in the prevention and control of cholera outbreaks during acute emergencies. Document CDR/GPV/95.1. World Health Organization.
195. WHO. 2003. Disease outbreak news, Cholera in Liberia. Communicable Disease Surveillance & Response (CSR), World Health Organization. [http://www.who.int/csr/don/2003\\_09\\_30/en/](http://www.who.int/csr/don/2003_09_30/en/) [Online.]
196. WHO. 2004. Health crisis in Liberia-the long road to recovery. World Health Organization. <http://www.who.int/features/2003/09b/en/print.html> [Online.]
197. Wong, H. C., and C. H. Lin. 2001. Evaluation of typing of *Vibrio parahaemolyticus* by three PCR methods using specific primers. *J. Clin. Microbiol.* **39**:4233-4240.
198. Wright, A. C., R. T. Hill, J. A. Johnson, M. C. Roghman, R. R. Colwell, and J. G. Morris, Jr. 1996. Distribution of *Vibrio vulnificus* in the Chesapeake Bay. *Appl. Environ. Microbiol.* **62**:717-24.
199. Wright, A. C., G. A. Miceli, W. L. Landry, J. B. Christy, W. D. Watkins, and J. G. Morris. 1993. Rapid identification of *Vibrio vulnificus* on nonselective media with an alkaline phosphatase-labeled oligonucleotide probe. *Appl. Environ. Microbiol.* **59**:541-546.
200. Xu, H. S., N. C. Roberts, F. L. Singleton, R. W. Attwell, D. J. Grimes, and R. R. Colwell. 1982. Survival and viability of nonculturable *Escherichia coli* and *Vibrio cholerae* in the estuarine and marine environment. *Microb. Ecol.* **8**:313-323.

201. Yamai, S., T. Okitsu, T. Shimada, and Y. Katsube. 1997. Distribution of serogroups of *Vibrio cholerae* non-O1 non-O139 with specific reference to their ability to produce cholera toxin, and addition of novel serogroups. *Kansenshogaku Zasshi* **71**:1037-1045.
202. Yamamoto, K., Y. Ichinose, N. Nakasone, M. Tanabe, M. Nagahama, *et al.* 1986. Identity of hemolysins produced by *Vibrio cholerae* non-O1 and *Vibrio cholerae* O1, biotype El Tor. *Infect. Immun.* **51**:927-931.
203. Yamasaki, S., T. Shimizu, K. Hoshino, S. T. Ho, T. Shimada, *et al.* 1999. The genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are closely related to those of *Vibrio cholerae* O22. *Gene* **237**:321-332.
204. Yamashiro, T., and M. Iwanaga. 1996. Purification and characterization of a pilus of a *Vibrio cholerae* strain: a possible colonization factor. *Infect. Immun.* **64**:5233-5238.
205. Yuan, P., A. Ogawa, T. Ramamurthy, G. B. Nair, T. Shimada, *et al.* 1994. *Vibrio mimicus* are the reservoirs of the heat stable enterotoxin gene (nag-st) among species of the genus *Vibrio*. *World J. Microbiol. Biotechnol.* **10**:59-63.
206. Zhu, J., and J. J. Mekalanos. 2003. Quorum sensing-dependent biofilms enhance colonization in *Vibrio cholerae*. *Dev. Cell* **5**:647-656.
207. Zhu, J., M. B. Miller, R. E. Vance, M. Dziejman, B. L. Bassler, and J. J. Mekalanos. 2002. Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.* **99**:3129-3134.
208. Zlatkin, I. V., M. Schneider, F. J. deBruijn, and L. J. Forney. 1996. Diversity among bacteria isolated from the deep subsurface. *J. Indust. Microbiol.* **17**:219-227.
209. Zo, Y. G., I. N. G. Rivera, E. Russek-Cohen, M. S. Islam, A. K. Siddique, *et al.* 2002. Genomic profiles of clinical and environmental isolates of *Vibrio cholerae* O1 in cholera endemic areas of Bangladesh. *Proc. Natl. Acad. Sci. U. S. A.* **99**:12409-12414.