Comparative Genomics of *Vibrio cholerae*

Jongsik Chun[1,2,3 *], Yoon-Seong Jeon[2], Je Hee Lee[1], Seon Young Choi[1], Dong Wook Kim[2], Christopher J. Grim[4], Nur A. Hasan[4], and Rita R. Colwell[4]

[1]*School of Biological Sciences and Institute of Microbiology, Seoul National University, Seoul 151-742*
[2] *International Vaccine Institute, Seoul, 151-818*
[3] *Center of Bioinformatics and Computational Biology, University of Maryland Institute of Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA*
[4] *Maryland Pathogen Research Institute, University of Maryland, College Park, MD 20742, USA*

Vibrio cholerae, a bacterium autochthonous to the aquatic environment, is the causative agent of cholera, a severe, watery, life-threatening diarrheal disease. Historically, cholera bacteria have been serogrouped based on their somatic O antigens, with more than 200 serogroups identified to date. Although strains from many of the serogroups of *V. cholerae* have caused either individual cases of mild gastroenteritis or local outbreaks of gastroenteritis, only the toxigenic strains of serogroups O1 and O139 have been identified as agents of cholera epidemics. Genes coding for cholera toxin, *ctx*AB, and other virulence factors have been shown to reside in bacteriophages and various mobile genetic elements. In addition, *V. cholerae* serogroup O1 is differentiated into two biotypes, classical and El Tor, by a combination of biochemical traits and sensitivity to specific bacteriophages.

V. cholerae serogroup O1 is responsible for the previous two cholera pandemics, in which classical and El Tor biotypes were dominant in the 6th and the current 7th pandemics, respectively. Cholera researchers continually face newly emerging and re-emerging pathogenic clones carrying combinations of new serogroups as well as of phenotypic and genotypic properties. These genotype and phenotype changes have hampered control of the disease. Here we compare the complete genome sequences of 23 strains of *V. cholerae* isolated from a variety of sources and geographical locations over the past 98 years in an effort to elucidate the evolutionary mechanisms governing genetic diversity and genesis of new pathogenic clones. The genome-based phylogeny revealed 12 distinct *V. cholerae* phyletic lineages, of which one, designated the *V. cholerae* core genome (CG), comprises both O1 classical and El Tor biotypes. All 7th pandemic clones share nearly identical gene content, i.e., the same genome backbone. The transition from 6th to 7th pandemic strains is defined here as a "shift" between pathogenic clones belonging to the same O1 serogroup, but from significantly different phyletic lineages within the CG clade. In contrast, transition among clones during the present 7th pandemic period can be characterized as a "drift" between clones, differentiated mainly by varying composition of laterally transferred genomic islands, resulting in emergence of variants, exemplified by *V.cholerae* serogroup O139 and *V.cholerae* O1 El Tor hybrid clones that produce cholera toxin of classical biotype. Based on the comprehensive comparative genomics presented in this study it is concluded that *V. cholerae* undergoes extensive genetic recombination *via* lateral gene transfer, and, therefore, genome assortment, not serogroup, should be used to define pathogenic *V. cholerae* clones.