DISTRIBUTION OF UNIVERSAL STRESS PROTEINS ENCODED IN VIBRIO, ESCHERICHIA, AND BURKHOLDERIA GENOMES

Jennifer N. Sims, Raphael D. Isokpehi, Wellington K. Ayensu and Hari H.P. Cohly

Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson MS 39217, USA

Abstract: The availability of complete genome sequences provides the possibility to identify the genes that encode for proteins that contain the universal stress protein (USP) domain (PF00582). In this study, we have utilized Escherichia and Burkholderia genomes to provide a comparative analysis of USPs encoded in genomes of Vibrio species. We hypothesize that the amino acid length variation and gene count will provide the status of expansion versus contraction will reveal some unique signature with respect to pathogenicity. The objective was to determine the distribution and compare the length of universal stress proteins encoded in Vibrio, Escherichia, and Burkholderia genomes. The Integrated Microbial Genome System was used to determine the distribution and the number and types of amino acids making up the universal stress proteins among sequenced Vibrio, Escherichia, and Burkholderia genomes. We further investigated length differences among Vibrio species by performing a multiple sequence alignment using ClustalW. Moreover, we investigated amino acid sequence differences revealed by ClustalW using Phyre to determine the secondary structure prediction and disorder. In genomes of Vibrio species, a minimum of three and a maximum of seven USP genes were observed. The USP gene counts of Escherichia USP were 2, 6, 7, and 8 while for Burkholderia, the range was from 7 to 22. The protein lengths with highest frequency in the three genera were 141, 142 and 144 for Vibrio, Escherichia and Burkholderia, respectively. We further generated a multiple sequence alignment for V. vulnificus since we observed a USP gene (VV0075) in V. vulnificus YJ016 with a different length from the other V. vulnificus genes including those that share the same chromosomal context and have similar size range (about 141 aa). This alignment uncovered an N-terminal insertion, which accounted for the increase in the amino acid length. The function of this insertion is unknown; however, it has been shown that N-terminal extensions confers a selective advantage and is said to be involved in quaternary structure. The disordered prediction of the N-terminal insertion of VV0075 revealed that most of the amino acids are disordered. Disordered regions at the N- and/or C- termini can cause problems in protein expression, purification, and crystallization and are the molecular basis of diseases. The amino acid length variation reflects ecological diversity in the three genera and the gene count for USP provides evidence of expansion and contraction of USP genes. In conclusion, the unique finding of the presence of the extended N-terminus in a V. vulnificus USP gene needs to be further studied in the context of strain-specific pathogenicity.

Keywords: Biological databases, Genome Analysis, Protein Domain, Vibrionaceae

Acknowledgements: National Science Foundation [EPSCoR and URM] (EPS-0903787; DBI-0958179); Pittsburgh Supercomputing Center's National Resource for Biomedical Supercomputing (T36 GM008789); U.S. Department of Homeland Security Science & Technology Directorate ((2007-ST-104-000007; 2009- ST-062-000014; 2009-ST-104-000021); and Research Centers in Minority Institutions (RCMI) – Center for Environmental Health (NIH-NCRR G12RR13459). Disclaimer: “The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.