SHORT COMMUNICATION

Virulence Factor Database (VFDB): Future in Dental Genomics

Swati Das[1], Vivek Kumar Bains[1], Rajesh Jhingran[1], Rohit Madan[1], Vivek Gupta[1] and Iram Rizvi[1]

ABSTRACT

Aim: To understand and analyze the potential application of the virulence factor database (VFDB) in dentistry.

Summary: Virulence factor database (VFDB) offer a platform for scientists to easy and rapid access to current knowledge about various Virulence Factors (VFs) from different bacterial pathogens to significantly expand knowledge about the varied mechanism of the disease processes at the molecular level. The core database on VFs, namely VFDB (http://www.mgc.ac.cn/VFs/) would be helpful for researchers to illuminate pathogenic mechanisms in various bacterial/infectious diseases that involve further instant investigations for the knowledge and henceforth development of different approaches for the disease treatment and prevention.

Keywords: Genomics, *S. mutans*, virulence factors (VFs), VFDB.

INTRODUCTION

"Look to thy mouth, disease enters here." The truthfulness of the phrase is based on the fact that there are approximately thirty different genera and 250 different cultivable species in the oral cavity. Various microorganisms contribute to the destructive process in the oral cavity by secreting toxins or virulent by products and further leads to the shift to pathogenic microflora thereby predisposing the sites for disease. Virulence by definition is defined as the degree of pathogenicity within a group or species of parasites as designated by case fatality rates and/or the ability of an organism to invade the tissues of the host, and it is determined by its virulence factors (VFs). Virulence derives from the adjective virulent, which can be describe either disease severity or a pathogen's infectivity.[1] The word "virulent" derives from "virulentus", which is a Latin word meaning "full of poison" or "a poisoned wound".[2,3]

Dr. Swati Das completed her graduation (BDS) from Meenakshi Ammal Dental College & Hospital, Chennai (TN) in the year 2010. Currently, she is pursuing her postgraduation (MDS) in Periodontology at Saraswati Dental College & Hospital, Lucknow (UP), India.

[1]Departments of Periodontology, Saraswati Dental College & Hospital, Lucknow (UP), India.
Address for Correspondence:
Dr. Swati Das, Department of Periodontology, Saraswati Dental College & Hospital, 233 Tiwari Ganj, Faizabad Road, Chinhat, Lucknow (UP), India.
Contact: +91 8948351317, E-mail: swatidas1214@gmail.com
Date of Submission: 02-09-2012
Reviews Completed: 19-09-2012
Date of Acceptance: 27-09-2012

Bacterial VFs are intriguing as they produce disease and survive in an intimidating environment by large virulence mechanisms, interact with host cells and modulate their functions. Understanding the mechanisms of action of virulence factors will enlighten new avenues for identifying promising approaches to disease prevention and therapy. Virulence factor database (VFDB), since its inception in 2004 (VFDB, http://www.mgc.ac.cn/VFs/) provide a source for scientists to fast access to the current acquaintance of VFs from different bacterial pathogens to greatly expand our understanding about the disease mechanisms at the molecular level.[4]

VFDB: virulence factor database

Molecular information obtained by genome analysis of microorganisms provides information for virulence; evolution and host adaptation.[5] The Joint Genome Institute operated by the University of California for the U.S. Department of Energy reported the completion of high quality draft nucleic acid sequences of 15 different microbial genomes in one month (www.jgi.doe.gov/tempweb/News/news). Till 28th Feb, 2001, forty-one microbial genomes had been completed, another thirteen was being annotated, and other 89 microbial genomes were "in progress",[6] however, this database is not available. Functional genomics of oral microbial pathogens (virus, bacteria, fungi, or yeast) use DNA microarrays to explore gene expression, gene-gene interactions, and multiple gene-environment interactions. VFDB is comprehensive and user-friendly and has provided the largest and most comprehensive up-to-date facts regarding bacterial VFs (such as secreted effectors or cell-associated products, such as outer membrane proteins or capsular polysaccharides and extracellular products, such as enzymes and toxins).[7] This core dataset of VFDB was initially developed in 2004 by Chen *et al*.[7] and later updated by them in 2012.[8,9] Based on experimentally

established VFs from around twenty-four genera bacterial pathogens of medical significance, the current dataset was developed by the researchers. Among them VFs in humanoid pathogens have several homologies present in plant/animal pathogens and also non pathogens. The most functional endorsed VFs are in fragments which encode the genomic encoding thereby, widespread literature screening along with various expert review were conducted and finally concluded on more than twelve hundred VFs and their basic information by the researchers. VFs collected from seventy-five genera of bacteria were included in VFBD at present which were further divided into four super-families and 31 subclasses. In the database only the representative sequence from individual species was collected for the comprehensive record in VFDB as more than one sequence was available for individual species.[9] VFDB provides a potent search engine for users to obtain information from the databank quickly by browsing each genus or by typing keywords or through three different ways, i.e., by text search, by BLAST search (Basic Local Alignment Search Tool) and VFs function category search. In order to search, store, retrieve and update information about VFs from different bacterial pathogens, VFDB provides a unified gateway (Fig. 1).[4,8]

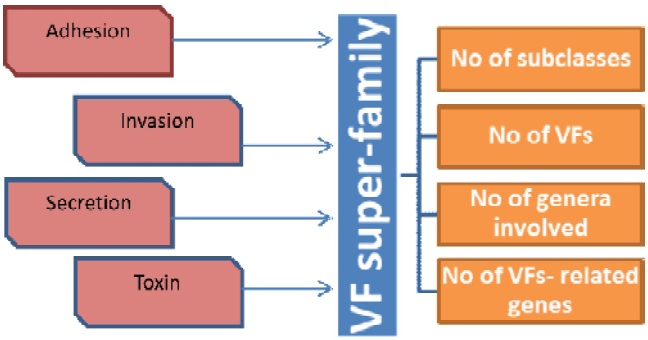

Figure 1: Data summary of newly released contents for the diversity and evolution analyses of VFs as of September 2011 (Modified from Chen *et al.*[7])

In the core data set by Chen *et al.*,[7] the amino-acid and nucleotide sequences of VF-encoding genes and associated annotation information were obtained from different GenBank records using ad-hoc BioPerl scripts. By local protein family (Pfam) search using the HMMER3 program (http://hmmer.org/), the conserved domain(s) of each protein were recognized, and the associated protein structure information was obtained from the protein data bank (PDB) database via batch BLAST search trailled by manual curation. Using reciprocal BLAST on different data sets of each class, homologue groups were determined, and the results were further created based on preserved synteny. The matrix global alignment tool (MatGAT program) and DaliLite server were utilised to obtain pairwise sequence as well as structural similarities (if any), respectively, among each group. For largely divergent proteins, respective conserved domain(s) segments were used instead of complete sequences for producing reliable alignments. The overall data-processing procedure is shown in Fig 2. The growing variety of VFs has incited numerous efforts to develop classification schemes and untangle the evolutionary origins of VFs. For instance, 6-major fimbrial clades of usher/chaperone systems and 7-different families of type III secretion system (T3SS) were established.[9]

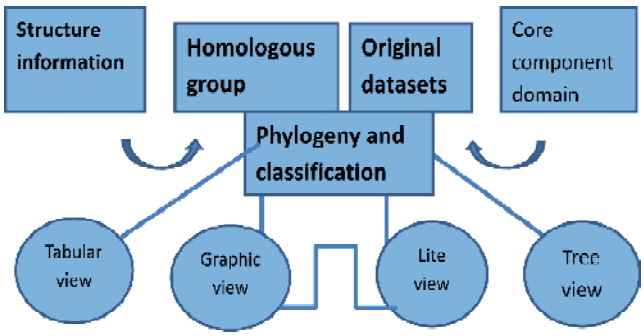

Figure 2: The overall data-processing procedure (Modified from Chen *et al.*[7])

VFDB in Oral Health and Disease

A specialized central repository and analysis platform for the corynebacterial research community is needed to host the fast-growing amount of genomic data and facilitate the analysis of these data. CoryneBase is a genomic database for *Corynebacterium* with varied functionality for the analysis of genomes aimed to provide: access to comprehensive *Corynebacterium* data through the use of advanced web technologies for interactive web interfaces; annotated genome arrangements of *Corynebacterium* where 4,180 RNAs and 165,918 coding sequences can be found in 27 species; and advanced bioinformatic investigation tools containing standard BLAST for homology search, VFDB BLAST for sequence homology search against VFDB, Pairwise Genome Comparison (PGC) tool for comparative genomic analysis, and a newly designed Pathogenomics Profiling Tool (PathoProT) for comparative pathogenomic analysis. CoryneBase offers the access of wide-range of *Corynebacterium* genomic resources as well as investigational tools for comparative genomics and pathogenomics. It is publicly available at http://corynebacterium.um.edu.my/.[10]

Another genomic database was for *Streptococcus*, Comparison of pathogenomics organization of *Streptococcus* was made out of total 28 genomes available. Several strains of *Streptococcus* were included out of which, strain of our interest for oral health provider is *S. mutans* strain UA159. The virulence factors to define the coding sequence of the *Streptococcus mutans* was by the adhesion Antigen I/II (SpaP), in the immune evasion , the capsule and the C3-degrading protease (Fig 3(a),(b) & (c)].[11]

Other Microbial virulence database

The bio-defense research community has been under constant surge for a system that could provide all the necessary evidence to identify the genomic signature and furthermore pronounce the functional signature of the pathogen. Under this context, many database were acquainted viz. protein toxins (Tox-Prot), a subset of the Swiss-Prot protein knowledgebase,[12] the SCORPION molecular database of scorpion toxins,[13] the PRINTS database of virulence factors,[14] the ARGO database of vancomycin and b-lactam antibiotic resistance genes,[15] the virus database (VIDA) of animal virus, consisting of five virus families,[16] Microbial database of protein toxins (MvirDB), a data granary of participating organism, classification, keyword, attribute, cross-references and sequence from all of these databanks,[17] and the TVFac toxin and VFDB at Los Alamos National Laboratory, the Islander database of genomic island.[18]

The knowledge of VFDB and its co-relation with the virulent pathogenic organism gives an insight to a better and programmed approach to various pathogenic condition caused by these microbes.

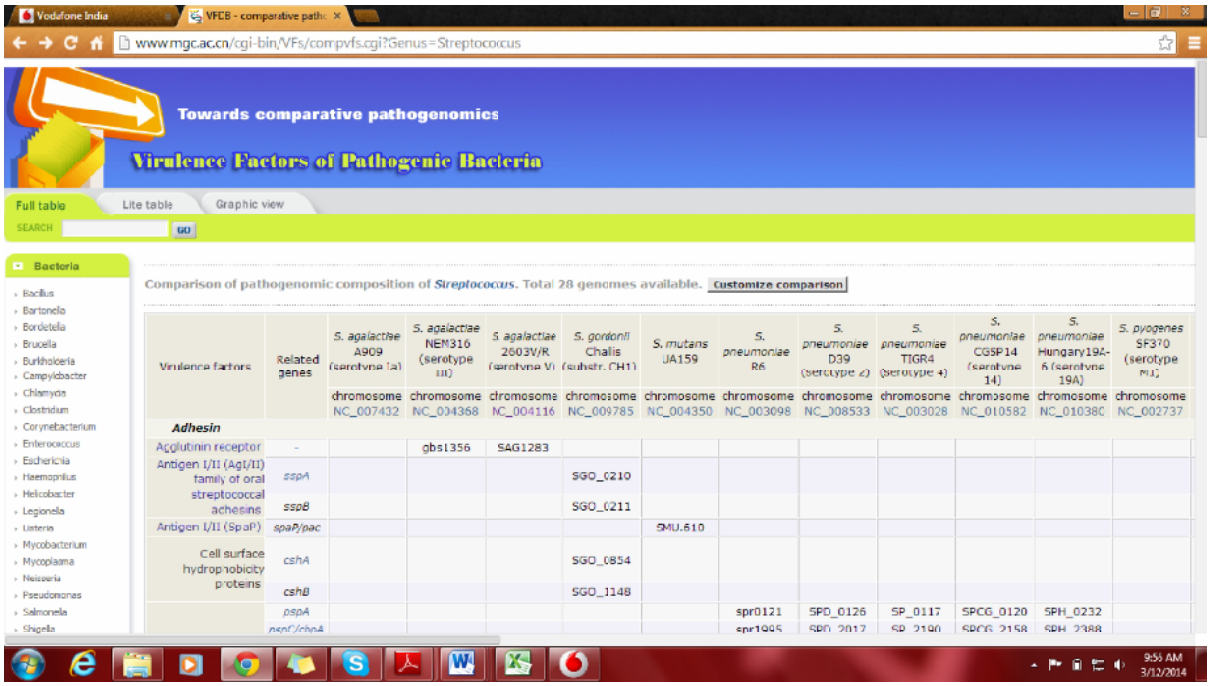

Figure 3(a): The main page contents of the genus *Streptococcus* (*www.ncbi.nlm.nih.gov/PMGifs/ Genomes/bact.html 02/28/01*)

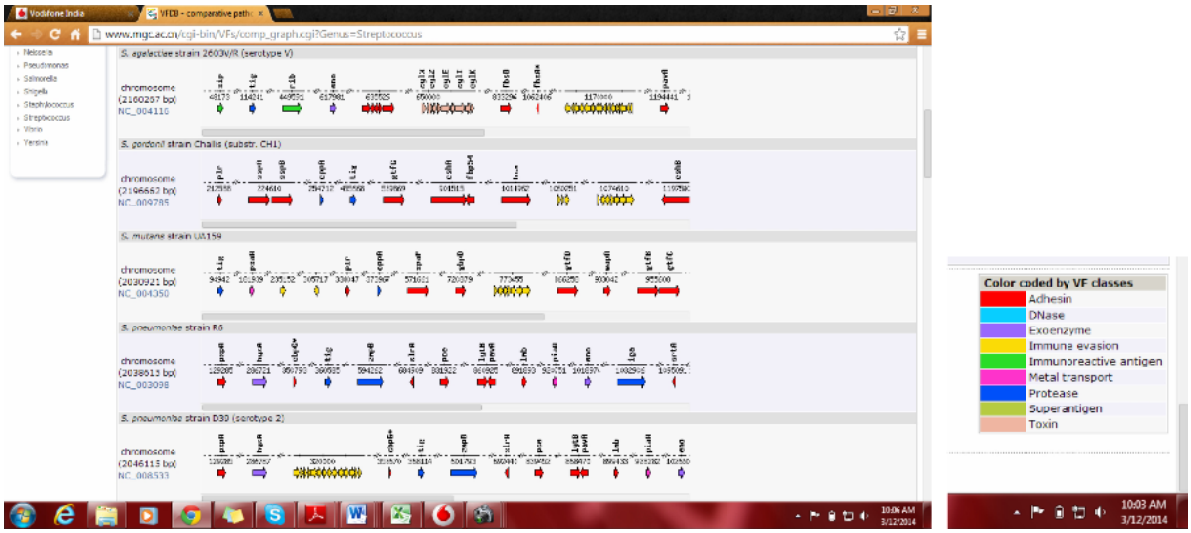

Figure 3(b): Lite table view of *S. mutans* (*www.ncbi.nlm.nih.gov/PMGifs/ Genomes/bact.html 02/28/01*)

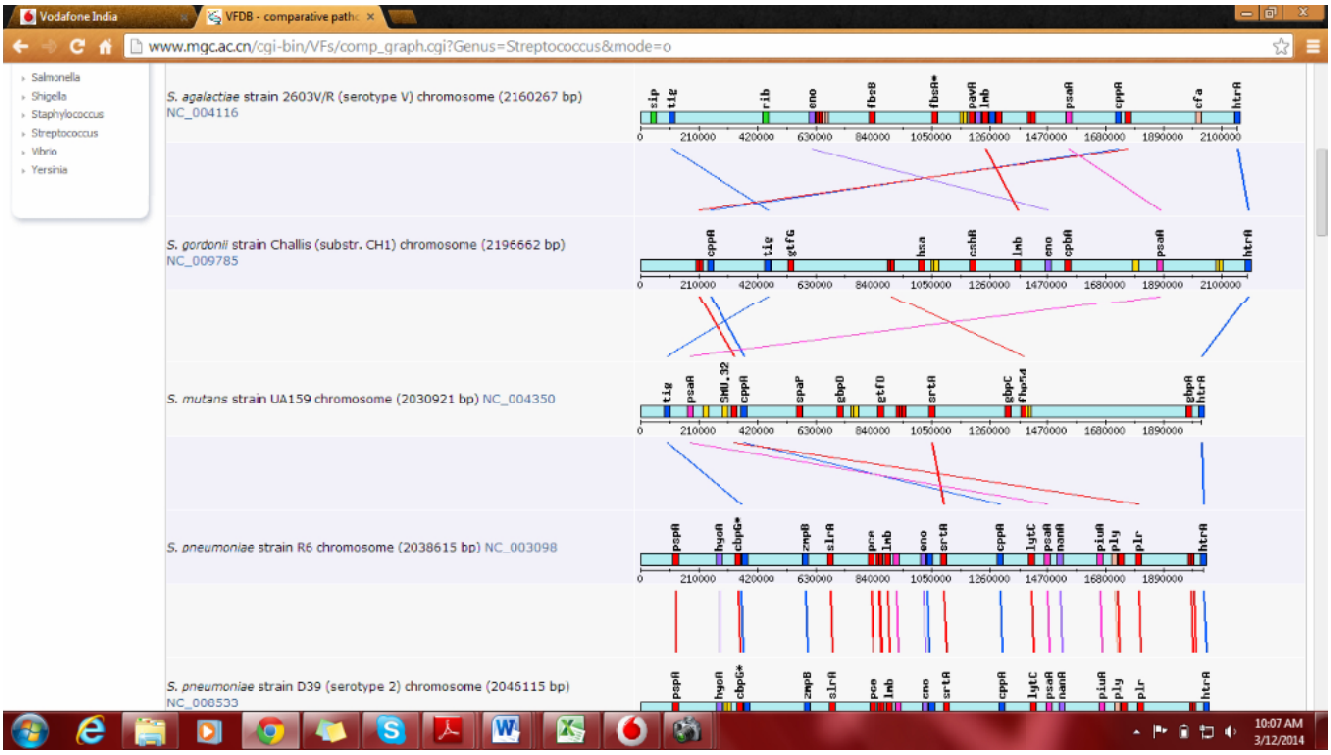

Figure 3(c): Graphic view of *S. mutans* (*www.ncbi.nlm.nih.gov/PMGifs/ Genomes/bact.html 02/28/01*)

CONCLUSION

The objective of the developing virulence factor database (VFDB) [7] was to provide a detailed summary of the important VFs from various bacterial pathogens by the functional and structural biology and immunology at its best. The systematic and comprehensive VFDB would not only be helpful for researchers to explain the mechanism of the pathogenicity which require innovative and novel approach to the treatment plan and its prevention, but also surge for more VFs of growing dynasty of micro-organisms.

REFERENCE

1. MeSH - Medical Subject Headings, Karolinska Institute, 13 April 2010.
2. Compact Oxford English Dictionary virulent.
3. A Latin Dictionary virulentus.
4. Weiss RA Virulence and pathogenesis. Trends Microbiol, 2002; 10: 314–7.
5. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. Simultaneous identification of bacterial virulence genes by negative selection. Science 1995; 269: 400-3.
6. Bumol TF, Watanabe AM. Genetic information, genomic technologies and the future of drug discovery. JAMA 2001; 285:551.
7. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 2005; 33: D325-8.
8. Yang J, Chen L, Sun L, Yu J, Jin Q. VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. Nucleic Acids Res 2008; 36: D539–42.
9. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res 2012; 40: D 641–45.
10. Heydari H, Siow CC, Tan MF, Jakubovics NS, Wee WY, Mutha NV, *et al*. CoryneBase: Corynebacterium genomic resources and analysis tools at your fingertips. PLoS One 2014; 17; 9: e86318.
11. Available online : www.ncbi.nlm.nih.gov/PMGifs/ Genomes/ bact.html 02/28/01
12. Jungo F, Bairoch A. Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein knowledgebase. Toxicon 2005; 45: 293-301.
13. Bergeron ZL, Bingham JP. Scorpion toxins specific for potassium (K+) channels: a historical overview of peptide bioengineering. Toxins (Basel) 2012; 4: 1082-119.
14. Attwood TK, Bradley P, Flower DR, Gaulton A, Maudling N, Mitchell AL, *et al*. PRINTS and its automatic supplement, prePRINTS. Nucleic Acids Res 2003;31:400-3.
15. Scaria J, Ghandramouli U, Verma SK. Antibiotic resistance genes online (ARGO): a database on vancomycin and b-lactam resistance genes Bioinformation 2005; 17: 5-7.
16. Alba MM, Lee D, Pearl FMG, Shepherd AJ, Martin N, Orengo CA, *et al*. VIDA: a virus database system for the organization of animal virus genome open reading frames Nucleic Acids Res 2001; 29: 133–6.
17. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. MvirDB- a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res 2007; 35: D 391-4.
18. Mantri Y, Williams KP. Islander: a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities Nucleic Acids Res 2004; 32: D55-8.