Bioinformation
www.bioinformation.net
Volume 5
Issue 1
open access
Database

CHIKVPRO – a protein sequence annotation database for chikungunya virus

Akaash Kumar Mishra[1], Chakresh Kumar Jain[1], Apurva Agrawal, Saransh Jain Kumar Sambhav Jain, Namrata Dudha, Kapila Kumar, Sanjeev K. Sharma, Sanjay Gupta*

Department of Biotechnology, Jaypee Institute of Information Technology, NOIDA 201307, India; Sanjay Gupta - E.mail: sjay1908@gmail.com, *Corresponding author

Received February 21, 2010; accepted April 06, 2010; published June 05, 2010

Abstract:
In the recent past, there has been a resurgence of interest in Chikungunya virus (CHIKV) attributed to massive outbreaks of Chikungunya fever in the South-East Asia Region. This has reflected in substantial increase in submission of CHIKV genome sequences to NCBI (National Center for Biotechnology Information) database. Hereby we submit a database "CHIKVPRO" containing structural and functional annotation of Chikungunya virus proteins (25 strains) submitted in the NCBI repository. The CHIKV genome encodes for 9 proteins: 4 non-structural and 5 structural. The CHIKVPRO database aims to provide the virology community with a single accession authoritative resource for CHIKV proteome- with reference to physiochemical and molecular properties, proteolytic cleavage sites, hydrophobicity, transmembrane prediction, and classification into functional families using SVM-Prot and other Expasy tools.

Availability: The database is freely available at http://www.chikvpro.info/

Keywords: CHIKVPRO, Chikungunya virus, CHIKV, database, protein sequence

Background:
The Chikungunya disease burden is highly associated with the large-scale morbidity. The disease manifests itself with an acute febrile illness that lasts for 2-5 days, followed by prolonged joint pain which may persist for weeks, months or even years **[1].** The causative agent Chikungunya virus was first isolated in Tanzania in 1953 and since then it was associated with a number of epidemic outbreaks in Africa, Southeast Asia and India **[2].** Aedes mosquitoes, primarily Aedes aegypti and Aedes albopictus have been identified as the vectors responsible for disease transmission.

In the 2006 epidemic, nearly 1.38 million people were reported with symptomatic disease across several states of India **[3].** Confirmed cases have been reported even in 2009 from various Indian states. There is no vaccine or any specific treatment for the disease and prevention by vector control is still the best policy. Most studies in literature are related to epidemiological aspects of CHIKV. The causative agent Chikungunya virus is a positive ssRNA virus belonging to the family Togaviridae and genus Alphavirus. The viral genome (approximately 11.8 kb) encodes for two polyproteins – the non structural polyprotein consisting of four proteins (nsP1, nsP2, nsP3 & nsP4) and the structural polyprotein consisting of five proteins (Capsid, E3, E2, 6K & E1) **[2].**

Development of new tools for diagnosis, prevention and treatment of Chikungunya need to be supported. There is a need to characterize the proteome of various CHIKV strains isolated during different outbreaks. CHIKVPRO, a database with computational analysis and integrated information of different CHIKV strains is an attempt in this direction to provide proteomic data to those involved in virology research.

Methodology:
Dataset:
The complete genome sequences of 25 Chikungunya strains (submitted till October 2009) were downloaded from NCBI nucleotide database **[4].** The translated protein sequence for each strain was extracted from NCBI protein database in two sets – the non-structural polyprotein sequence and the structural-polyprotein. For all strains, the non structural polyprotein sequence was divided into 4 protein sequences – non structural proteins nsP1, nsP2, nsP3, and nsP4; and the structural polyprotein sequence into 5 protein sequences – capsid, 6K, envelope protein E1, E2 and E3. This fragmentation was done computationally using parameters such as amino acid positions and sequence alignment in reference to the strains 37997 and African S27 characterized in Uniprot database **[5].**

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 5(1): 4-6 (2010)

Bioinformation **Volume 5**
www.bioinformation.net Issue 1 **Database**
open access

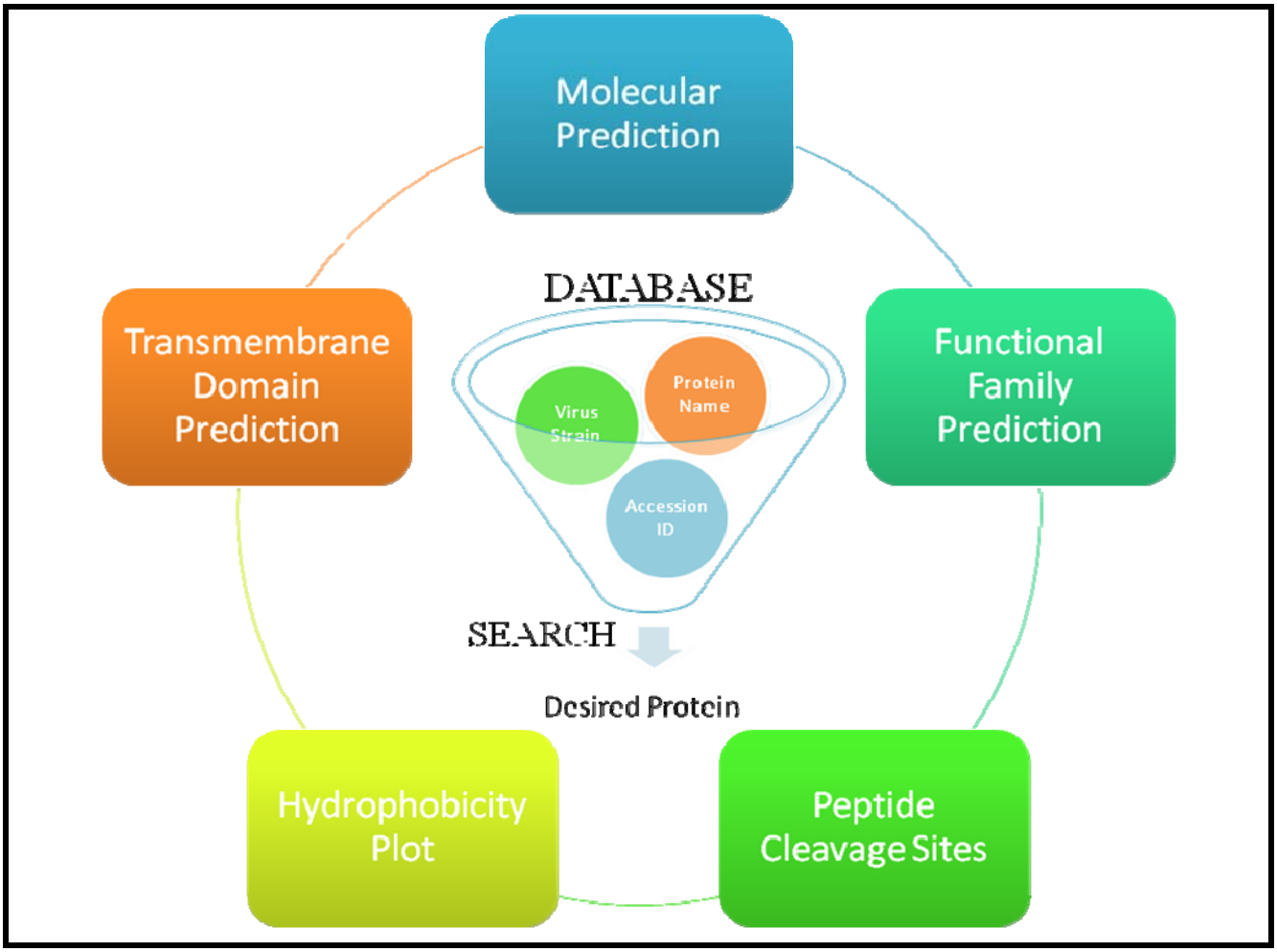

Figure 1: Database Structure of CHIKVPRO: The inner circles represent the search criteria to select a desired protein sequence and the outer rectangles represent different types of predictions that can be performed by the user.

Database:
MySQL was used to construct a relational database to store information of viral proteins in the form of related tables. Data consistency and non redundancy was maintained by using normalisation techniques. HTML and PHP were used to provide a dynamic web interface which was appropriately connected with the database. The database is freely available to view and download data.

Database features:
CHIKVPRO provides its users to search a protein through its name, accession ID and virus strain **(Figure 1).** After selecting a particular protein, the user can perform the following: (1) Molecular prediction – gives molecular weight, amino acid composition, atomic composition and physiochemical properties like theoretical pI, instability index, aliphatic index and grand average of hydrophobicity (GRAVY). The tool used for this analysis was Protparam **[6].** (2) Peptide cleavage prediction – provides proteases cleavage sites for the selected protein analysed by the tool Peptide cutter **[6].** (3) Hydrophobicity plot – provides hydrophobicity plot for each protein sequence analysed by ProtScale tool **[6].** (4) Transmembrane domain prediction – shows potential trans-membrane domains for each protein predicted by DAS tool **[6].** (5) Functional family prediction – classifies a protein into related functional families by using SVM Prot software **[7].**

Database utility:
The in silico analysis gives us a brief idea about the role of each protein and helps us to understand its biological significance. Such information might help virologists better understand the mode of virus replication, its mechanism of pathogenesis, strain specific variations and to develop potential anti-viral agents **[8].** Since earlier studies have shown that a single mutation in the virus affects vector specificity, severity and epidemic potential, this database may also provide useful information for determining the virulence of the new isolates **[9, 10].** The database provides its users a web-interface which is highly user-friendly and easy to access **(Figure 1).**

Conclusions and future aspects:
CHIKVPRO, a protein sequence annotation database of Chikungunya virus was designed to provide an easy access to the large and growing volume of data. The database provides useful resource of information on viral proteins for molecular biologists and virologists working in related areas. In the submitted version, CHIKVPRO provides information on physiochemical, molecular and functional properties of each protein. It also illustrates the various peptide cleavage sites, transmembrane and other functional domains present in each protein. We plan to further include additional features such as higher structural conformations, post translational modifications, etc. and also upload the data for the remaining strains in our advanced version of CHIKVPRO. The database shall be updated timely as more data on viral proteins is generated through our ongoing experimental analysis and future sequence submissions.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 5(1): 4-6 (2010)

Bioinformation **Volume 5**
www.bioinformation.net Issue 1 **Database**
open access

References:
[1] V Ravi, *Indian J Med Microbiol.* **24**: 83 (2006) [PMID 16687855]
[2] SP Kalantri *et al. Natl Med J India* **19**: 315 (2006) [PMID 17343016]
[3] AH Khan *et al. J Gen Virol.* **83**: 3075 (2002) [PMID 12466484]
[4] http://www.ncbi.nlm.nih.gov/
[5] http://www.uniprot.org/
[6] http://www.expasy.org/tools/
[7] http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi
[8] MR Dikhit *et al. Bioinformation* **3**: 299 (2009) [PMID 19293996]
[9] NP Kumar *et al. J Gen Virol.* **89**: 1945 (2008) [PMID 18632966]
[10] KA Tsetsarkin *et al. PLoS Pathog.* **3**: e201 (2007) [PMID 18069894]

Edited by P. Kangueane
Citation: Mishra *et al.* Bioinformation 5(1): 4-6 (2010)

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 5(1): 4-6 (2010)