JEVBase: An Interactive resource for protein annotation of JE Virus

Manas Ranjan Dikhit manasranjandikhit@gmail.com
Scientist, BioMedical Informatics Division
Rajendra Memorial Research Institute
of Medical Sciences
Agam Kuan, Patna, India-800007

Ganesh Chandra Sahoo ganeshiitkgp@gmail.com
Scientist & Head, BioMedical Informatics Division
Rajendra Memorial Research Institute
of Medical Sciences
Agam Kuan, Patna, India-800007

Pradeep Das drpradeep.das@gmail.com
Director
Rajendra Memorial Research Institute
of Medical Sciences
Agam Kuan, Patna, India-800007

Abstract

Databases containing proteomic information have become indispensable for virology related studies. Rajendra Memorial Research Institute of Medical Sciences (RMRIMS) has compiled and maintained a functional and molecular annotation database (http://www.jevbase.biomedinformri.org) commonly referred to as JEVBase. This database facilitates significant relationship between molecular analysis, cleavage sites and possible protein functional families assigned to different proteins of Japanese encephalitis virus (JEV). Identification of different protein functions and molecular analysis facilitates a mechanistic understanding of (JEV) infection and opens novel means for drug development. JEVBase database aims to be a resource for scientists working on JE virus.

Keywords: JEVBase, *Japanese encephalitis,* SVMProt, Protein annotation database, Functional database

1. INTRODUCTION

Japanese encephalitis virus (JEV) is the most common agent of viral encephalitis in the world, causing an estimated 45,000 cases and 10,000 deaths annually [1]. Epidemic form of Japanese encephalitis has been known since 1924 when 4,000 human deaths were recorded in Japan [2]. JE virus infection is also wide spread in southern states of India.

JEV contains a single positive sense RNA strand with about 11Kb nucleotides [3]. A single precursor polyprotein derived from JEV genome is subsequently processed by the host and viral protease to produce three structural proteins(Capsid (C), membrane(prM/M) and envelope (E)) and seven nonstructural proteins(NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [4].

These three structural proteins are synthesized in the order of C, M and E from the 5' half of a single long open reading frames of the flavivirus genome. The glycosylated preM (precursor of M protein) and E proteins appear to be released from the nascent polyprotein following co-translational cleavage by signal peptidases. Late in virion maturation, preM is cleaved to M [5].

Different protein functions and molecular analysis facilitates for finding potential anti-viral inhibitors. Knowledge about protein function is essential in the understanding of biological processes [6]. The presence of a shared domain within a group of proteins does not necessarily imply that these proteins perform the same function [7]. One approach for function prediction is to classify a protein into functional family. Support vector machine (SVM) is a useful method for such classification, which may involve proteins with diverse sequence distribution [8]. Cloning and expression of different proteins practiced by molecular biologist will be helped by in silico restriction site analysis in the database.

In virology research, virus-related databases and bioinformatics analysis tools are essential for discerning relationships within complex datasets about viruses [9]. Chemoinformatics is the use of informatics methods to solve chemical problem[10]. Computational analysis and chemical information on *Japanese encephalitis* viruses involves the general tasks related to the analysis of any novel sequences, such as molecular analysis, functional annotation, and analysis of cleavage sites of the sequences. Support vector machines (SVM), useful for predicting the functional class of distantly related proteins, is employed to ascribe a possible functional class to *Japanese encephalitis* virus protein . Novel JEV protein functions have been analyzed through SVMProt have been earlier reported [11].

As the gap between the amount of sequence information and functional characterization widens, increasing efforts are being directed to the development of databases. For virologist, it is therefore desirable to have a single data collection point which integrates research related data from different domains. JEVBase is our effort to provide virologist such a one-step information center. We describe herein the creation of JEVBase, a new database that integrates information of different proteins in to a single resource. For basic curation of protein information, the database relies on features from other selected databases, servers and related papers.

2. MOTIVATION FOR JEVBase

Virology was slower to embrace bioinformatics [12]. No computational functional analysis of different proteins of JE virus is available till date. Protein identification and analysis software performs a central role in the investigation of proteins from two-dimensional (2-D) gels and mass spectrometry. For protein annotation, the user matches certain empirically acquired information against a protein database to define a protein function as already known or as novel. For protein analysis, information in protein databases can be used to predict certain properties about a protein, which can be useful for its empirical investigation.

All these in silico analysis give us an idea concerning the role of different proteins of JEV in replication, survival and spread of JEV in the host. Considering the biological significance of JEV protein and with the aim of providing easy access to the large and growing volume of data, we have developed JEVBase, a repository of all known JEV protein. JEVBase is the first known web resources, which provide the sequences as well as annotation information. The JEV protein have been analyzed, organized and integrated to develop a user friendly database and analysis system. The web interface enables the user to execute a quick and efficient search on JEVBase data. The database can be queried comprehensively through arguments such as National Center for Biotechnology Information **(**NCBI**)** Locus number, different protein name, different predicted functional family, stability etc. JEVBase will be an extremely useful resource for computational and experimental biologist working in this and related areas.

3. DATA PROCUREMENT AND REFINEMENT

The large scale of protein sequences have been reported in the NCBI protein database and supplementary data in the published literature. The sequences of *Japanese encephalitis* have been downloaded from the National Center for Biotechnology Information (NCBI) Protein database. Sequence redundancy is another problem of JE virus sequences in public protein databases. Different strains of the same species from samples collected in different location or at different times may possess completely identical sequences. Redundancy and repetition in protein sequences has been carefully removed by using ALIGN software to obtain a unique dataset [13]. Exactly matching sequences taken from multiple sources were eliminated while constructing the dataset. The raw dataset was preprocessed to remove the sequence smaller than 50bp while analyzing with different software.

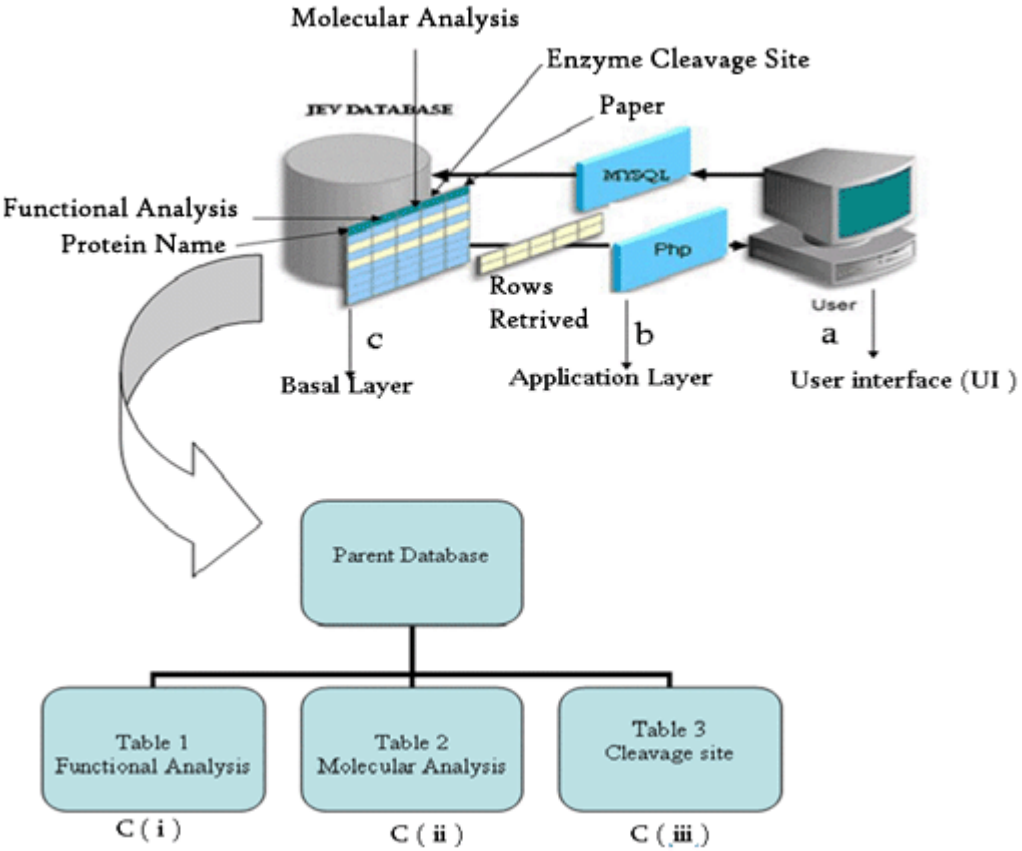

FIGURE 1. System architecture of the JEV database

4. SYSTEM ARCHITECTURE AND DESIGN

A relational database was constructed in MySQL for storage and query of data. It includes three key entities: 'functional analysis', 'molecular analysis' and 'cleavage sites', which simply analyze the protein. The JEVBase consists of three layers: (Fig1), 'The basal layer', 'Application layer' and 'UI'. The user interface (UI) layer has been developed using Php, CSS and JavaScript. Hypertext Preprocessor (php) is a widely used, general-purpose scripting language that was originally designed for web development, to produce dynamic web pages and Cascading Style Sheets (CSS) is a style sheet language used to describe the presentation semantics [14]. The basal layer that is parent database has been divided in to 3 tables. The JEVBase data and

information have been stored in MySQL relational database tables. Meta-information for different types of biological data is placed as individual table in this layer. The application layer between the web interface and the backend relational tables has been implemented using Php. The three layers of JEVBase can be manipulated and developed independently, which provides an optimal environment for maintenance and expansion of the JEVBase. Most of the interface component and application layer were standardized. This was made possible by employing a standardized scheme in building each layer.

5. DATABASE FEATURES

5.1. Data access

JEVBase can be queried to obtain the information about the protein sequences in many ways. Data stored in JEVBase can be accessed in the following ways:

(i) Search by protein name: The user can enter the desired protein name to access the Meta information about the protein sequences.

(ii) Search by protein functional family: The user can select the different protein functional family to find out the protein functional group of different structural and non-structural proteins.

(iii) Search by NCBI locus ID: The user can enter the NCBI locus ID to obtain JEV protein sequence information.

(iv) Search by Instability Index: To find out the stable and unstable protein, user can search by instability index.

JEVBase can be queried to obtain the information about protein-protein comparison. The user can enter the corresponding NCBI locus ID to compare two proteins.

5.2. Visualization

Database visualization helps users process, interpret and act upon large stored data sets. JEVBase provides a number of web-based forms for querying the dataset and selecting one or more protein for either a more detailed view of molecular annotation, Cleavage site and functional family or for viewing the comparison between two selected proteins.

In an effort to improve access to diverse JEV data, The JEVBase has been modified to include an abundance of linkage to other database including pubmed [www.ncbi.nlm.nih.gov/sites/entrez] for related papers and NCBI [www.ncbi.nlm.nih.gov] for corresponding sequences.

After performing a typical search the user is first presented with a summery page detailing the number of proteins matching the search (Fig 2). The following result page then provides the user with a list of proteins and brief descriptions from which individual proteins may be selected for either a detailed view (functional family, molecular annotation & cleavage sites) or a view of the related paper.

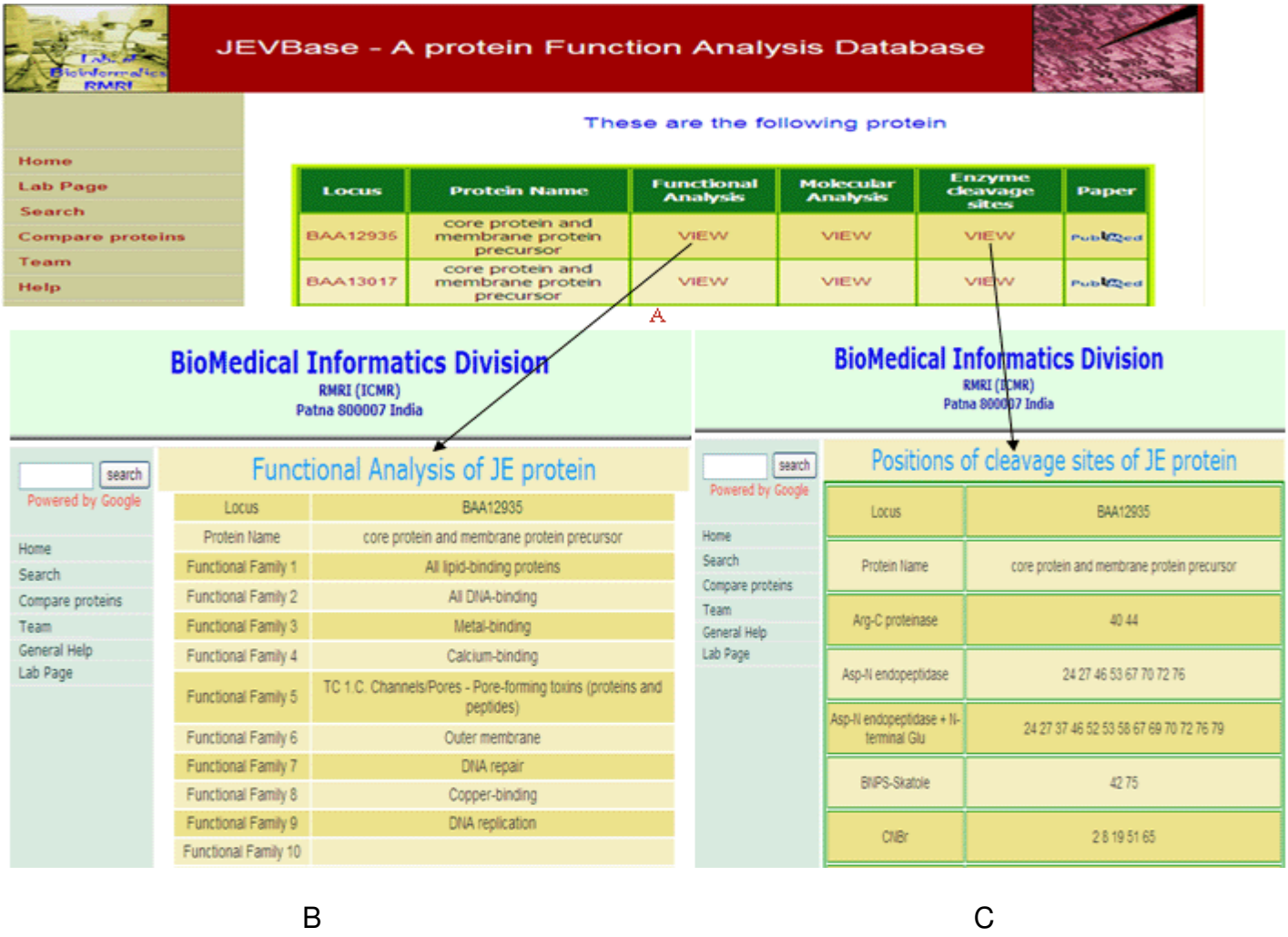

FIGURE 2. Typical screenshots from JEVBase. (A) Search result pages. These pages provide detailed view of each protein identified by a search. (B) Functional family view. This option gives the detailed view of protein functional family. (C) Cleavage sites view. These options give the detailed view of cleavage sites of the protein.

5.3. Data analysis

The protein function family predicted by SVMProt and other research papers are different for each structural and non-structural protein of JE virus strain, some of which may be responsible for virulence or pathogenicity of the virus and others for replication of the virus in the host. Prediction of the functional roles of lipid binding proteins is important for facilitating the study of various biological processes and the search for new therapeutic targets. Comparison of two amino acid sequences of any JE protein will reveal the user the distinguished functional properties of the corresponding protein if there is any amino acid change at any position as SVM works on the basis of physico-chemical properties of the amino acids of the protein e.g. when comparing function assignment of protein of two different NCBI locus numbers ABD84344 and ABD84370 , functions like metal binding, copper binding and DNA repair are common to both the strains whereas lipid degradation, calcium binding , iron binding and DNA binding functions are specific to ABD84344 and outer membrane and magnesium binding functions are specific to ABD84370(fig 3). Like this in molecular analysis of PreM protein, molecular

properties of this protein are also found to be different for each strain. From peptide-cutter and protparam analysis, pattern of restriction sites for all types of restriction enzymes for JE virus protein are visualized from the web server.

Hence the protein function family predicted is different for each structural and non-structural protein of JE virus strain, some of which may be responsible for virulence or pathogenicity of the virus and others for replication of the virus in the host. Prediction of the functional roles of lipid binding proteins is important for facilitating the study of various biological processes and the search for new therapeutic targets.

From this analysis, it is predicted that there is presence of network of functions performed by JEV proteins which brings about severe complicated clinical manifestations e.g. toxin-like pore forming property of core, matrix and NS4B proteins of JEV is responsible for causing acute flaccid paralysis as pore formation in the host causes release of water, micronutrients and macronutrients which can also occur in nerve cells causing severe inflammation of nerve cells and hence the patient suffers from acute meningitis.

This is comparision between two given proteins

FUNCTIONAL ANALYSIS

	PrM/PreM	PrM/PreM
	ABD84344	ABD84370
Functional Family 1		
Functional Family 2		
Functional Family 3	Lipid degradation	
Functional Family 4		
Functional Family 5	Metal-binding	Metal-binding
Functional Family 6	Calcium-binding	
Functional Family 7	Copper-binding	Copper-binding
Functional Family 8	DNA repair	DNA repair
Functional Family 9	Iron-binding	
Functional Family 10		Outer membrane
Functional Family 11		
Functional Family 12		Magnesium-binding
Functional Family 13	All DNA-binding	
Functional Family 14		
Functional Family 15		
Functional Family 16		
Functional Family 17		
Functional Family 18		
Functional Family 19		
Functional Family 20		

MOLECULAR ANALYSYS

	PrM/PreM	PrM/PreM
	ABD84344	ABD84370
Number of amino acids	80	80
Molecular weight	8891.2	8934.2
Negatively charged residues (Asp + Glu)	12	13
Positively charged residues (Arg + Lys)	7	7
Total number of Carbon atom	381	384
Total number of Hydrogen Atom	607	612
Total number of Nitrogen Atom	101	100
Total number of Oxygen Atom	123	124
Total number of Sulfur Atom	10	10
Total number of Atom	1222	1230
Instability index	stable	stable

FIGURE 3 . showing comparative functional analysis of same protein but with different NCBI locus numbers

6. CONCLUSION

JEVBase has been designed to manage and to explore the vast amount of protein data analysis. The current version of JEVBase has provided the basic molecular and functional analysis data of different proteins of *Japanese encephalitis* virus. JEVBase has been developed by keeping pace with the progress of the availability *Japanese encephalitis* proteins. User can search either by protein Functional family or protein name to access the Meta information about the protein sequences. This database facilitates significant relationship between molecular analysis, cleavage sites in the sequence, related paper and possible protein functional family of different proteins. Understanding of the structure-function correlation in viruses is important for finding potential anti-viral inhibitors and vaccine targets.

In near future we aim to include the modeled structures of different JE proteins and analyze quantitative structure–activity relationship of novel ligands targeting different proteins of JE virus. The database will be updated weekly on the basis of availability and analysis of the JE virus information and the amino acid sequences from NCBI and other reliable resources.

7. AVAILABILITY

The JEVBase database is freely available at http://www.jevbase.biomedinformri.org. All questions, comments and requests should be sent by e-mail to ganeshiitkgp@gmail.com .

8. ACKNOWLEDGEMENTS

The authors would like to thank Priya Darsan Sahu, Lingaraj Jena, Mukta Rani and Dr. Sindhuprava Rana for helpful discussion and valuable suggestions. We are thankful to Dr. Meera Singh for helping us during establishment of our division. This work is supported by Indian Council of Medical Research (ICMR), Govement of India.

9. REFERENCE:

[1] T. Solomon, L. T. T. Thao, N. M. Dung, R. Kneen, N. T. Hung, A. Nisalak, D. W. Vaughn, J. Farrar, T. T. Hien, N. J. White and M. J. Cardosa. " Rapid Diagnosis of Japanese Encephalitis by Using an Immunoglobulin M Dot Enzyme Immunoassay". Journal of Clinical Microbiology, 36**,** 2030-2034, 1998

[2] E.L. Buescher and W. F. Scherer. "Ecologic studies of Japanese encephalitis virus in Japan. IX. Epidemiologic correlations and conclusions". Am. J. Trop. Med. Hyg., 8,719-722, 1959

[3] H. Sumiyoshi, C. Mori, I. Fuke, K. Morita, S. Kuhara, J. Kondou, Y. Kikuchi, H. Nagamatu and A. Igarashi. "Complete nucleotide sequence of the Japanese encephalitis virus genome RNA". Virology, 161, 497-510, 1987

[4] F. Zhang, W. Ma, L. Zhang, M. Aasa-Chapman and H. Zhang. "Expression of particulate-form of Japanese encephalitis virus envelope protein in a stably transfected Drosophila cell line". Virology, 4, 17-24, 2007

[5] X. Liu, S. Cao, R. Zhou, G. Xu, S. Xiao, Y. Yang, M. Sun, Y. Li and H. Chen. "Inhibition of Japanese encephalitis virus NS1 protein expression in cell by small interfering RNAs", Virus Genes, 33, 69-75, 2006

[6] D. Eisenberg, C.A. Marcotte, I. Xenarios and T.O. Yeates. "Protein function in the post-genomic era". Nature , 405, 823–826, 2000

[7] S. Henikoff, E. A. Greene, S. Pietrokovski, P. Bork,T.K. Attwood and L. Hood. "Gene families: the taxonomy of protein paralogs and chimeras". Science, 278**,** 609–614, 1997

[8] C.Z. Cai, L.Y. Han, Z.L. Ji, X. Chen and Y.Z. Chen. "SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence". Nucleic Acids Research, 31, 3692–3697, 2003

[9] M. R. Dikhit , S.P. Rana, P. Das, and Ganesh Chandra Sahoo. " CHPVDB - a sequence annotation database for Chandipura Virus". Bioinformation. 3(7): 299–302, 2009

[10] A. Abdo and N. Salim. "Inference Networks for Molecular Database Similarity Searching". International Journal of Biometric and Bioinformatics, 2 (1) 1-16, 2008

[11] G.C. Sahoo, M.R. Dikhit and P. Das. " Functional assignment to JEV proteins using SVM". Bioinformation, 3**,** 1-7, 2008

[12] Y. Qing. "Bioinformatics databases and tools in virology research: An overview". In Silico Biology, 8, 0008 ,2008

[13] http://www.ebi.ac.uk/Tools/emboss/align/index.html .

[14] www.phpro.org