Research Article

Ebolavirus Database: Gene and Protein Information Resource for Ebolaviruses

Rayapadi G. Swetha,[1] Sudha Ramaiah,[1] Anand Anbarasu,[1] and Kanagaraj Sekar[2]

[1]*Medical & Biological Computing Laboratory, School of Biosciences and Technology, VIT University, Vellore 632 014, India*
[2]*Laboratory for Structural Biology and Bio-Computing, Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560 012, India*

Correspondence should be addressed to Anand Anbarasu; aanand@vit.ac.in

Ebola Virus Disease (EVD) is a life-threatening haemorrhagic fever in humans. Even though there are many reports on EVD, the protein precursor functions and virulent factors of ebolaviruses remain poorly understood. Comparative analyses of *Ebolavirus* genomes will help in the identification of these important features. This prompted us to develop the *Ebolavirus* Database (EDB) and we have provided links to various tools that will aid researchers to locate important regions in both the genomes and proteomes of *Ebolavirus*. The genomic analyses of ebolaviruses will provide important clues for locating the essential and core functional genes. The aim of EDB is to act as an integrated resource for ebolaviruses and we strongly believe that the database will be a useful tool for clinicians, microbiologists, health care workers, and bioscience researchers.

1. Introduction

Ebolavirus is responsible for outbreaks of severe haemorrhagic fever in humans, and it is endemic in Equatorial Africa. EVD is reported in many countries, and all of them had the origins in Africa, by a way of travel. *Ebolavirus* infections have high case fatality rates of around 42% [1–4]. Ebolaviruses can spread by direct contact with either infected patients or mortal remains of infected individuals [5]. Currently, there is no effective prophylaxis (including vaccine) for *Ebolavirus* infection [6]. Hence, the Centers for Disease Control and Prevention and the National Institutes of Health have classified *Ebolavirus* as a "category A bioterrorism agent" [7, 8]. The *Ebolavirus* group comprises five viruses, namely, Taï Forest virus, Reston virus, Sudan virus, Ebola virus, and Bundibugyo virus [9]. The first Ebola virus outbreak was reported in 2013, which resulted in death of more than 2,622 people, including many health care workers; and as of today, approximately 11,302 people have succumbed to the disease. The number of deaths in the current outbreak is higher when compared to that of all the previous outbreaks combined. Thus, *Ebolavirus* outbreaks represent a major global public health problem [10–12]. The

comparative analyses of *Ebolavirus* genomes may result in the identification of highly conserved regulatory regions, which may play important roles in *Ebolavirus* biology. We have developed the *Ebolavirus* Database (EDB), a database exclusively for *Ebolavirus* (the EDB is freely available through the following URL: http://bioserver1.physics.iisc.ernet.in/EDB/). The EDB provides a powerful, user-friendly interface to perform various Boolean searches, sequence and literature based searches. The in-built tools in EDB, namely, BLAST and RNA motif search, can be used to compare the genomes of *Ebolavirus*, and they can also be used to identify the RNA motifs within *Ebolavirus* genomes. The BLAST in EDB consists of set of similarity search programs to perform varied homology searches. It can be used to explore all the Ebola viral genomes. It provides a powerful way to compare the novel sequences with previously characterized Ebola viral genes. The BLAST tool highlights the regions of local alignment to detect relationships among sequences that share only isolated regions of similarity [13]. Functional RNA molecules are involved in numerous biological processes, ranging from gene regulation to protein synthesis. The RNA motif search tool helps in the analyses of functional RNA motifs and elements in ebolaviruses genomes and provides useful

information for deciphering RNA regulatory mechanisms in ebolaviruses. The Generic Genome Browser (GBrowse) tool in EDB is a combination of database and interactive web pages. It can be employed for manipulating and displaying the various annotations on Ebola viral genomes [14]. The database is also interfaced with the Jmol plugin to visualize three-dimensional structure of *Ebolavirus* proteins [15]. This feature helps the users to analyze the functionally active proteins associated with *Ebolavirus* pathogenicity. The Ebola virus belongs to *Zaire ebolavirus* species and *Zaire ebolavirus* infections have case fatality rate of 42.2%. In a simple case study using RNA motif search and GBrowse tools in EDB, we find that CpG motif (5′-GTCGTT-3′) is more prominent in *Zaire ebolavirus*, and the GC content is higher in 2,000 to 2,399 positions than any other positions in the genome of *Zaire ebolavirus*. This is one of the most important observations from our case studies. Researchers can make use of EDB to locate interesting and significant regions in both genomes and proteomes of *Ebolavirus*. In addition, EDB provides the detailed information on virulent proteins, reservoirs, epidemiology, pathogenesis, and laboratory diagnosis for *Ebolavirus* infections.

2. Materials and Methods

The complete genomes of all known ebolaviruses were obtained from the National Center for Biotechnology Information [16] in Genome Feature Format and FASTA format. The genomes of ebolaviruses in these two formats were added in GBrowse. The proteomes of ebolaviruses were retrieved from UniProt database [17]. We used a relational database management system, MySQL, to store and manage the complete data of EDB.

The EDB was developed using PERL/CGI and PERL/DBI modules and the user-friendly web forms were coded in HTML, JavaScript, and Ajax. Solaris server which is well known for its adaptability, security, and scalability was used to host EDB. The database has been thoroughly checked and validated on different platforms (Windows, Linux, iOS, and Solaris) and works well with different web browsers (IE, Chrome, Opera, and Firefox).

3. Results and Discussion

3.1. Complex, User-Friendly Search Options. The simple or advanced Boolean-based search tools available in EDB are helpful in exploring the complete annotations of genes/proteins of *Ebolavirus*. In a simple text based search, the users can search for genes and proteins of *Ebolavirus* by entering the gene/protein name in the text box. In gene search, the genes can also be searched by entering their corresponding gene number. The user can also browse for the complete genome annotations of the ebolaviruses. The advanced search option in EDB is used to retrieve the list of proteins localized in a specific cellular compartment. EDB allows the users to obtain the proteins based on clusters of orthologous groups and on specific pattern/profile for the downstream system level analysis. EDB aids the users to obtain the proteins based on their status (reviewed/unreviewed). In

Table 1: The number of occurrences of CpG motif (GTCGT{2}) in ebolaviruses by "RNA motif search tool."

Species name	Number of occurrences
Bundibugyo ebolavirus	2
Reston ebolavirus	2
Sudan ebolavirus	7
Taï Forest ebolavirus	1
Zaire ebolavirus	11

addition, the links to *Ebolavirus* related PUBMED literatures are provided in EDB.

3.2. Facilitating Sequence Based Motif and BLAST Searches. The most important components in the analyses of gene regulation are the sequence motifs with known biological function [18, 19]. The motifs are typically found nonrandomly in the genome [20]. EDB is interfaced with a search tool, "RNA motif search," which is used to locate the user specified motifs within the coding sequences of ebolaviruses genomes. The tool accepts a stretch of RNA sequence of different length in IUPAC format and then the tool converts the input sequence into a regular expression. In addition to "RNA motif search" tool, BLAST tool is provided in EDB through which the user can perform the sequence based similarity searches for either protein or nucleotide sequences against a particular or all known ebolaviruses [13, 21]. A major advantage of using EDB is that it makes the user save the results of multiple searches in the hard disk of a local computer either as a text document or as a portable document format file.

3.2.1. Case Study. The CpG motifs are simple dinucleotide sequence of 5′-cytosine-guanosine-3′. The outcome of several studies on immunotherapy of cancer, vaccination, antisense therapy, and gene therapy highlights the importance of CpG motif [22]. The CpG motif 5′-GTCGTT-3′ is identified to be the best stimulatory motif for human cells [23]. The CpG motif was searched against all the ebolaviruses genomes through RNA motif search tool and the results are given in Table 1. Interestingly, we observed that *Zaire ebolavirus* genome has the highest number of occurrences (11) compared to genome of other ebolaviruses (Figure 1). This is just one example of how integration of this tool can lead to new insights while analyzing ebolaviruses genomes.

3.3. Genome Sequences Utilizing GBrowse. The genome content of ebolaviruses has to be easily accessible to researchers for their quick interpretation. To facilitate this, GBrowse developed by Stein et al. [14] has been incorporated in EDB. The browser has special features like navigate, scroll, and zoom in and zoom out over the random regions of the genome. In GBrowse, a specific region of a genome or a landmark can be searched by entering them in the search box available at the top left corner of the page. Then, the browser redirects to the user specified region and displays five tracks: (i) genes, (ii) proteins, (iii) GC content, (iv) 3-frame translation, and (v) 6-frame translation. Each track carries a link to the corresponding information available in
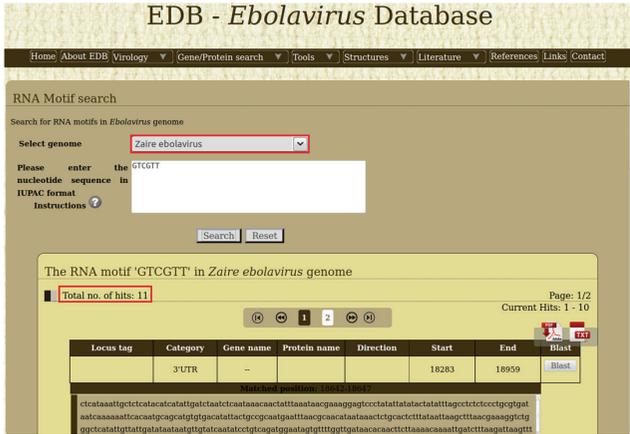
FIGURE 1: The number of occurrences of CpG motif (GTCGT{2}) in *Zaire ebolavirus* by "RNA motif search tool."
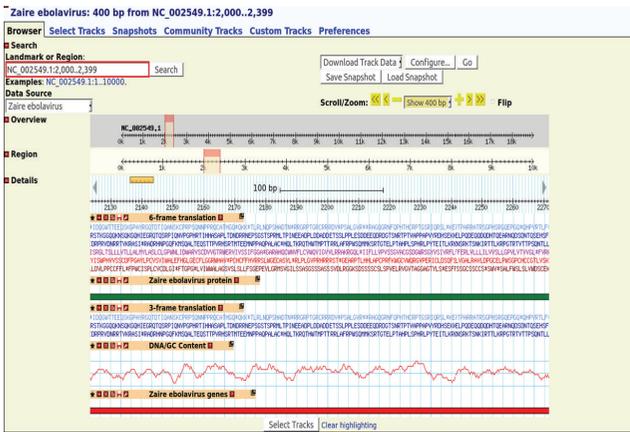


FIGURE 2: The genes, proteins, GC content, 3-frame translation, and 6-frame translation tracks of *Zaire ebolavirus* from 2,000 to 2,399 in GBrowse.


FIGURE 3: GP proteins of *Zaire ebolavirus* visualized in GBrowse.
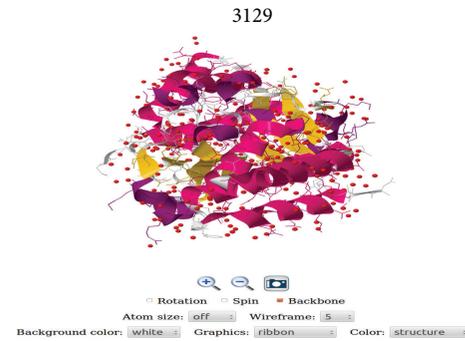
FIGURE 4: The Jmol view of three-dimensional structure of RNA binding protein (PDB ID: 3L29).

EDB or NCBI. Thus, GBrowse enables the user to easily view the genomic content of all ebolaviruses. The genome of *Zaire ebolavirus* (Figure 2) and glycoproteins (GP) (Figure 3) are visualized in GBrowse.

3.3.1. Case Study. The *Zaire ebolavirus* outbreak has resulted in the death of 12,452 persons since its discovery in 1976. The size of *Zaire ebolavirus* genome is 18.96 Kb with 41.1% of GC content [24]. Figure 2 shows various tracks of *Zaire ebolavirus* genome from positions 2,000 to 2,399 where the GC content is found to be notably high.

The only viral protein present on the envelope of *Ebolavirus* is GP. GP is a viral determinant of *Ebolavirus* pathogenicity and probably contributes to haemorrhage during infection [25]. The second secreted GP, spike GP, and small secreted GP are the most important factors responsible for viral entry into the host cell. The coding regions of these GPs are positioned from 5,900 to 8,305 (Figure 3). Interestingly, we found that matrix proteins and minor nucleoprotein are anchored to the sides of these GPs [26].

This viral assembly is unique to *Ebolavirus* and it might help to design effective anti-*Ebolavirus* compounds. Researchers can exploit this option to obtain important features in Ebola viral proteins.

3.4. Other Features in EDB. As of March 25, 2016, 63 three-dimensional structures of *Ebolavirus* proteins were available in Protein Data Bank [27]. It is necessary to include these structures in EDB to examine the functionally active proteins. The interactive graphics Java based plugin, Jmol, is incorporated in EDB to visualize these structures. Figure 4 shows an example of Jmol viewer displaying the three-dimensional structure of a RNA binding protein of *Zaire ebolavirus* (PDB ID: 3L29) [28]. Additionally, the information on virulent proteins, reservoirs, epidemiology, pathogenesis, laboratory diagnosis, and treatment of EVD is provided under "Virology" menu to provide the preliminary knowledge about the ebolaviruses. The links for various resources related to ebolaviruses are provided in the "Links" menu.

4. Conclusion

The comparative study of *Ebolavirus* genomes provides insights into the identification of highly conserved regions that play a significant role in *Ebolavirus* pathogenicity. The

EDB provides a flexible, user-friendly interface and also offers links to tools such as BLAST and RNA motif search to facilitate the comparative study of *Ebolavirus* genomes. These analyses can be exploited for identifying the putative essential and core *Ebolavirus* genes. We believe that EDB will act as universal single point access of educative and information archive for EVD. EDB will provide valuable genomic and proteomic information on *Ebolavirus* for clinicians, microbiologists, health care workers, and bioscience researchers. The database will be updated on a periodic basis.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] H. Feldmann and T. W. Geisbert, "Ebola haemorrhagic fever," *The Lancet*, vol. 377, no. 9768, pp. 849–862, 2011.

[2] H. Feldmann, "Ebola—a growing threat?" *The New England Journal of Medicine*, vol. 371, no. 15, pp. 1375–1378, 2014.

[3] M. J. Murray, "Ebola virus disease: a review of its past and present," *Anesthesia and Analgesia*, vol. 121, no. 3, pp. 798–809, 2015.

[4] T. Hoenen, A. Groseth, D. Falzarano, and H. Feldmann, "Ebola virus: unravelling pathogenesis to combat a deadly disease," *Trends in Molecular Medicine*, vol. 12, no. 5, pp. 206–215, 2006.

[5] A. S. Khan, F. K. Tshioko, D. L. Heymann et al., "The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995," *The Journal of Infectious Diseases*, vol. 179, supplement 1, pp. S76–S86, 1999.

[6] Y. Yazdanpanah, J. R. Arribas, and D. Malvy, "Treatment of Ebola virus disease," *Intensive Care Medicine*, vol. 41, no. 1, pp. 115–117, 2015.

[7] National Institute of Allergy and Infectious Diseases, "NIAID Emerging Infectious Diseases/Pathogens," 2015, http://www.niaid.nih.gov/topics/biodefenserelated/biodefense/pages/cata.aspx.

[8] Centers for Disease Control and Prevention, *Interim Guidance for Environmental Infection Control in Hospitals for Ebola Virus*, Centers for Disease Control and Prevention, 2014, http://www.cdc.gov/vhf/ebola/healthcare-us/cleaning/hospitals.html.

[9] A. A. Bukreyev, K. Chandran, O. Dolnik et al., "Discussions and decisions of the 2012–2014. International Committee on Taxonomy of Viruses (ICTV) Filoviridae Study Group, January 2012–June 2013," *Archives of Virology*, vol. 159, pp. 821–830, 2014.

[10] L. R. Baden, R. Kanapathipillai, E. W. Campion, S. Morrissey, E. J. Rubin, and J. M. Drazen, "Ebola—an ongoing crisis," *The New England Journal of Medicine*, vol. 371, no. 15, pp. 1458–1459, 2014.

[11] A. S. Fauci, "Ebola—underscoring the global disparities in health care resources," *New England Journal of Medicine*, vol. 371, no. 12, pp. 1084–1086, 2014.

[12] T. R. Frieden, I. Damon, B. P. Bell, T. Kenyon, and S. Nichol, "Ebola 2014—new challenges, new global response and responsibility," *The New England Journal of Medicine*, vol. 371, no. 13, pp. 1177–1180, 2014.

[13] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

[14] L. D. Stein, C. Mungall, S. Shu et al., "The generic genome browser: a building block for a model organism system database," *Genome Research*, vol. 12, no. 10, pp. 1599–1610, 2002.

[15] Jmol: an open-source Java viewer for chemical structures in 3D, http://www.jmol.org/.

[16] D. L. Wheeler, T. Barrett, D. A. Benson et al., "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 35, pp. D5–D12, 2007.

[17] A. Bairoch, R. Apweiler, C. H. Wu et al., "The universal protein resource (UniProt)," *Nucleic Acids Research*, vol. 33, pp. D154–D159, 2005.

[18] P. D'haeseleer, "What are DNA sequence motifs?" *Nature Biotechnology*, vol. 24, no. 4, pp. 423–425, 2006.

[19] R. G. Swetha, D. K. Kala Sekar, S. Ramaiah, A. Anbarasu, and K. Sekar, "*Haemophilus influenzae* Genome Database (HIGDB): a single point web resource for *Haemophilus influenzae*," *Computers in Biology and Medicine*, vol. 55, pp. 86–91, 2014.

[20] D. Halpern, H. Chiapello, S. Schbath et al., "Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling," *PLoS Genetics*, vol. 3, no. 9, pp. 1614–1621, 2007.

[21] R. G. Swetha, D. K. K. Sekar, E. D. Devi et al., "*Streptococcus pneumoniae* Genome Database (SPGDB): a database for strain specific comparative analysis of *Streptococcus pneumoniae* genes and proteins," *Genomics*, vol. 104, no. 6, pp. 582–586, 2014.

[22] R. K. Scheule, "The role of CpG motifs in immunostimulation and gene therapy," *Advanced Drug Delivery Reviews*, vol. 44, no. 2-3, pp. 119–134, 2000.

[23] C. A. Janeway Jr., "The immune system evolved to discriminate infectious nonself from noninfectious self," *Immunology Today*, vol. 13, no. 1, pp. 11–16, 1992.

[24] H. Ebihara, A. Takada, D. Kobasa et al., "Molecular determinants of Ebola virus virulence in mice," *PLoS Pathogens*, vol. 2, no. 7, p. e73, 2006.

[25] Z.-Y. Yang, H. J. Duckers, N. J. Sullivan, A. Sanchez, E. G. Nabel, and G. J. Nabel, "Identification of the Ebola virus glycoprotein as the main viral determinant of vascular cell cytotoxicity and injury," *Nature Medicine*, vol. 6, no. 8, pp. 886–889, 2000.

[26] V. E. Volchkov, S. Becker, V. A. Volchkova et al., "GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases," *Virology*, vol. 214, no. 2, pp. 421–430, 1995.

[27] H. M. Berman, T. Battistuz, T. N. Bhat et al., "The protein data bank," *Acta Crystallographica—Section D: Biological Crystallography*, vol. 58, no. 6, pp. 899–907, 2002.

[28] K. C. Prins, S. Delpeut, D. W. Leung et al., "Mutations abrogating VP35 interaction with double-stranded RNA render Ebola virus avirulent in guinea pigs," *Journal of Virology*, vol. 84, no. 6, pp. 3004–3015, 2010.

BioMed
Research International

Stem Cells
International

International Journal of
Peptides

Advances in
Virology

International Journal of
Genomics

International Journal of
Zoology

Journal of
Nucleic Acids

Journal of
Signal Transduction
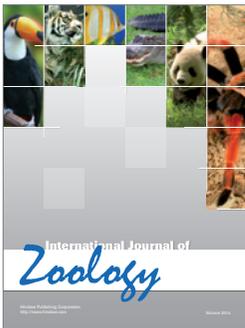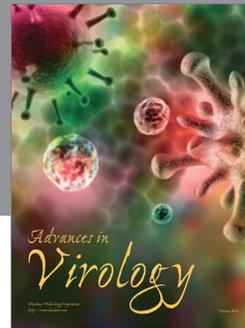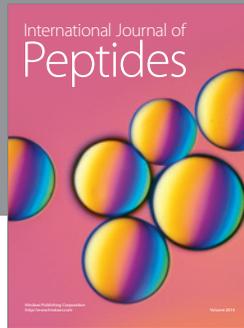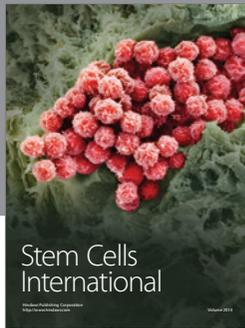
Hindawi

Submit your manuscripts at
http://www.hindawi.com

The Scientific
World Journal

Genetics
Research International

Anatomy
Research International

International Journal of
Microbiology

Biochemistry
Research International

Advances in
Bioinformatics

Archaea

Enzyme
Research

International Journal of
Evolutionary Biology

Molecular Biology
International

Journal of
Marine Biology