



# SeeHaBITaT: A server on bioinformatics applications for Tospoviruses and other species



Seethalakshmi Sakthivel, S.K.M. Habeeb \*

Department of Bioinformatics, School of Bioengineering, SRM University, Kattankulathur, Chennai 603203, India

## ARTICLE INFO

### Article history:

Received 12 February 2016

Accepted 6 March 2016

### Keywords:

Tospovirus  
Tomato spotted wilt virus  
Biological database  
Sequence analysis

## ABSTRACT

Plant viruses are important limiting factors in agricultural productivity. Tospovirus is one of the severe plant pathogens, causing damage to economically important food and ornamental crops worldwide through thrips as vectors. Database application resources exclusively on this virus would help to design better control measures, which aren't available. SeeHaBITaT is a unique and exclusive web based server providing work bench to perform computational research on tospoviruses and its species. SeeHaBITaT hosts Tospoviruses specific database Togribase, MOLBIT, SRMBIT and SS with PDB. These applications would be of immense help to the Tospovirus scientific community. The server could be accessed at <http://bit.srmuniv.ac.in/>.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Viruses result in a major setback to crop production and yield. Control measures largely rely on genetic resistance, hygienic measures or eradication of diseased crops, due to unavailability of antiviral products (Hanssen et al., 2010). *Bunyaviridae*, the largest RNA virus family is a home to 350 different virus types spread across five genera; the vertebrate infecting *Orthobunyavirus*, *Nairovirus*, *Phlebovirus*, and *Hantavirus* and the plant infecting Tospovirus (Elliott, 1997; Kuhn et al., 2014). Tospovirus species are rapidly spreading over large geographic areas (Hanssen et al., 2010). Thrips-transmitted tospoviruses cause significant economic losses to tomato, chili and many other important crops worldwide (Whitfield et al., 2005). Tospoviruses are species specific, having greater host range and are known to be transmitted by at least fourteen thrips species (Riley et al., 2011; Pappu et al., 2009). Tospoviruses are transmitted from plant to plant in a very specific manner by these thrips. Tomato Spotted Wilt Virus [TSWV] is one of the ten deadly plant viruses and it has a wide host range among plant viruses (Rotenberg et al., 2015), infecting more than 1090 plant species; both dicots and monocots, across 80 plant families (Parrella et al., 2003). The tripartite ambisense genome of tospovirus has three segments (Large, Medium and Small) coding for 5 genes involved in basic machinery required for their functioning and existence (Flick and Bouloy, 2005; Walter and Barr, 2011; Beaty and Bishop, 1988).

The geographical distribution of tospoviruses is widespread (Rotenberg et al., 2015) and these viruses are reported from Asia, Australia, Europe, Africa, North and the South America as well (Pappu et al., 2009). Twenty tospovirus species are recorded worldwide (Rotenberg et al., 2015) and the greatest diversity of tospovirus species

has been recorded in the Asian continent. A few have worldwide distribution, e.g., TSWV and INSV, whereas others remain restricted to the Eurasian or American continents. The known natural host range of some tospoviruses is increasing. So far, the largest diversity of tospoviruses is observed in the eastern part of Asia, where many species can be found, i.e., Tomato Spotted Wilt Virus (TSWV), Peanut bud necrosis virus (PBNV), Watermelon silver mottle virus (WSMoV), Watermelon bud necrosis virus (WBNV), Peanut chlorotic fanspot virus (PCFV), Melon yellow spot virus (MYSV), Iris yellow spot virus (IYSV), and a tentative species from *Gloxinia* also found and meanwhile reported from Australia as *Capsicum chlorosis* virus (CaCV) (Pappu et al., 2009).

Symptomatic and morphological identifications could go wrong many times, often leading to incorrect control measures. The combined efforts of plant virologists, plant pathologists, plant breeders, geneticists, weed scientists, entomologists, agronomists, and agricultural economists have been successful in managing few species (Riley et al., 2011). This is highly attributed to the knowledge gained in the fields of molecular biology, virus – plant and host – virus interaction and their epidemiology. The greatest task still remains because of the viral resistance and new strains are being reported from various corners of the world (Pappu et al., 2009).

Emerging biotechnology fields such as bioinformatics and genomics play a vital role in the current research scenario. Increasing data due to large scale genome sequencing projects worldwide and their analysis would pave way for the better understanding of the virus and sort out new control measures. Emerging sets of viral genome data from virus families and sporadic isolates are prompting a focused interest in establishing standards for consistent and comprehensive genome sequence storage and annotation for viral genomes (Seto, 2010). Several virus specific databases have been developed Dpvweb (Adams, 2006), Viralzone, (Hulo et al., 2010) and Virusbanker Database (Fourment and Gibbs, 2008). Therefore, the information gained from diversified

\* Corresponding author.

E-mail address: [habeeb\\_skm@yahoo.co.in](mailto:habeeb_skm@yahoo.co.in) (S.K.M. Habeeb).

fields of science on tospovirus species could be utilized to develop a database and an annotation server. These apart, a comprehensive network of complete cross intra-tospovirus species comparative study coupled with other handy applications under one platform would be of great importance to the tospovirus specific community. Hence, in this study, we have developed an application exclusively for tospoviruses species employing bioinformatics techniques.

## 2. Methodology

The architecture of the SeeHaBITaT follows a three-tier model. A front-end Apache web server (version: 2.2.17) serves content to the client browsers. A middle dynamic content processing and generation layer consists of PERL-CGI (version: 5.10.1), PHP scripts and other programs to connect NCBI Standalone Blast and ClustalW software. Finally a backend MySQL database which holds agricultural and genomics information about Tospovirus.

### 2.1. Data collection & database construction

Togribase was exclusively developed to store agricultural information related to 20 known Tospovirus species. Data collected from literature was reviewed, refined and stacked. Genomic data related to Tospovirus were collected from NCBI includes 4758 nucleotide sequences, 4357 protein sequences and 9 whole genome sequences. Sequences downloaded in FASTA format were formatted using the PERL program for further usage. The whole genome of 9 Tospovirus species was compared with each other and the data necessary to built Genome browser and comparative genomics were compiled using Perl programmes. Further, the data were classified into agricultural and molecular biology data which was stored in two relational databases Togribase and Molbit respectively.

## 3. Result and discussion

The web interface of SeeHaBITaT provides access to four major sections (Fig. 1). (i) Togribase (Tospovirus agricultural Database) (ii) MOLBIT (Molecular Biology of Tospovirus) (iii) SRMBIT (SRM Bioinformatics Tools) and (iv) SS with PDB (Secondary Structure with PDB Validation).

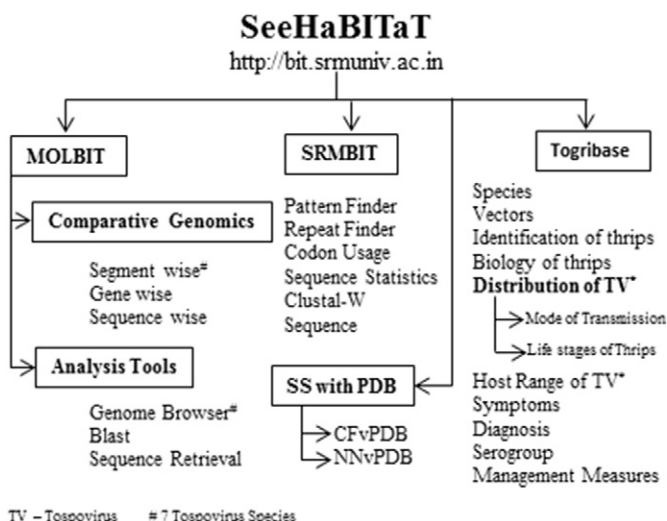


Fig. 1. Hierarchy of web application available under SeeHaBITaT.

### 3.1. Togribase: Tospovirus aGRicultural dataBASE

Togribase was developed in an effort to create a unique platform which provides complete agricultural information on Tospovirus. The information, content ranges from a basic level, such as different species under the genus Tospovirus, various vectors which are known to transmit tospovirus; to advance level such as the diagnosis and management measures. The virus is known to be transmitted mainly by thrips, therefore, their biology and identification hints also stored in this database along with its pictorial representation for the perusal of the scientific community working with this group of virus. The effort was also taken to give a brief outline of the various hosts of all the known Tospovirus species. This content is also complemented by the vectors section of Togribase, which highlights the highly favored thrips for a type virus. This database provides well formatted agricultural information about Tospovirus species than existing resources (Adams, 2006; Hulo et al., 2010; Fourment and Gibbs, 2008).

### 3.2. MOLBIT: MOlecular Biology of Tospovirus

The study of any virus contributes to our understanding about the molecular basis of viral infection in general (Mettenleiter, 2008). As was stated above, the idea of the genetics and molecular biology of the viral genome would help us to gather more information on its cellular functioning. This has the potential to explore new control strategies. MOLBIT forms the platform for interspecies comparative as discussed below.

#### 3.2.1. Comparative genomics

Complete genome sequence of nine available viruses has been explored and was compared with each other. This comparison is based on three levels such as segment wise, gene wise and sequence wise comparison. A graphical output is generated upon a segment comparison between the species; leading to the identification of the specific location of the different segments in the genome. The comparison can be furthered between complete sequences of the species by using the sequence wise comparison tab on a global scale employing Needleman–Wunsch algorithm (Mount, 2001).

#### 3.2.2. Genome browser

Genome browser an important aspect of our server has been developed to give a vivid description of the location of genes and proteins encoded with it. Genome Browser provides rapid and reliable information on any requested portion of the genome. This information is provided at a basic level under the genome summary which further gets expanded with information about the individual segment in a complete manner under the segment summary in a user friendly graphical display.

#### 3.2.3. BLAST

A server is incomplete if it cannot meet the basic requirements of Bioinformatics applications such as sequence analysis. To ease the verification of the origins of new virus species at the sequence level, BLAST was implemented to perform similarity search. Similarity search can be performed for a viral sequence against the exclusive tospovirus database and against all types of viral genomes as well. These results are furnished as a summary table which holds essential information that is extracted from the BLAST result.

#### 3.2.4. Sequence retrieval

Sequence retrieval provides a way to perform three different types of searches, enabling easy retrieval of Tospovirus sequences. (i) Basic Search – retrieve sequences based on keyword or accession number. (ii) Advanced Search – retrieve sequences based on organism name or sequence length. (iii) Property Search – retrieves sequence based on sequence statistical property.

### 3.3. SRMBIT

SRMBIT (SRM Bioinformatics Tools) provides a collection of web based programs written in PERL-CGI to perform protein and nucleotide sequence analysis in an efficient manner. This hierarchy was planned as an extension to the MOLBIT application for further analysis of the sequences present in the Togribase. SRMBIT can also be made use by the users to analyze their sequences from other organisms retrieved from various data repositories. These applications include Pattern finder, Repeat finder, Codon usage, ClustalW (Web interface), Sequence fragmenter and Sequence statistics. Each performs an important task under sequence analysis.

### 3.4. SS with PDB

Though, it is feasible to predict the tertiary structure of protein from amino acid sequences through methods like homology modeling, threading, ab-initio; however intricate in some cases. Under the circumstances, it's a practice to predict the secondary structure of protein, which is an intermediate in tertiary structure which also provides an insight into protein function (Kloczkowski et al., 2002; Qian and Sejnowski, 1988).

Two mostly employed methods of secondary structure prediction Chou Fasman algorithm (CFvPDB) (Ashok, 2013) and Neural Networks (Qian and Sejnowski, 1988) (NNvPDB) (Sakthivel, 2015) based secondary structure prediction can be done at SS with PDB. The most distinguishing part of these two applications is automated validation of the predictions made by these applications with experimentally solved structures in PDB is done. A unique aspect which is not available in any other leading secondary structure prediction applications. The level of accuracy achieved by these applications is 64.46% and 79.56% by CFvPDB and NNvPDB respectively.

## 4. Conclusion

Tospoviruses have significant impact on agricultural crops. Thus far, there is no exclusive web resource for Tospovirus. In this study, we worked with two types of databases which are unique for Tospovirus (i) Togribase: Tospovirus aGRicultural database and (ii) MolBIT: Molecular Biology of Tospovirus. These two databases were provided with two further applications under a separate hierarchy (iii) SRMBIT and (iv) SS with PDB. The objective of this study was to provide the scientific community an environment where complete information related to the dangerous Tospovirus species. SeeHaBITaT provides the user from basic knowledge related to tospoviruses to its diagnosis and control measures; information on its molecular biology to its sequence and structural analysis. We hope our server meets the needs of the end user. Further, extending the server with more applications such tertiary structure

prediction, epitope identification, protein–protein interaction, and evolutionary studies would be a suitable fit to make the SeeHaBITaT the most sought server by the Tospovirus scientific community.

### Conflict of interest

The authors have no conflict of interests.

### Acknowledgment

The authors sincerely acknowledge SRM University, India for continuous support and encouragement toward research.

### References

- Adams, M., 2006. DPVweb: a comprehensive database of plant and fungal virus genes and genomes. *Nucleic Acids Res.* 34 (90001), D382–D385.
- Ashok, Kumar T., 2013. CFSSP: Chou And Fasman secondary structure prediction server. *Wide Spectrum* 1 (9), 15–19.
- Beaty, B., Bishop, D., 1988. Bunyavirus–vector interactions. *Virus Res.* 10 (4), 289–301.
- Elliot, R.M., 1997. Emerging viruses: the Bunyaviridae. *Mol. Med.* 3 (9), 572–577.
- Flick, R., Bouloy, M., 2005. Rift Valley fever virus. *CMM* 5 (8), 827–834.
- Fourment, M., Gibbs, M., 2008. The VirusBanker database uses a java program to allow flexible searching through Bunyaviridae sequences. *BMC Bioinf.* 9 (1), 83.
- Hanssen, I., Lapidot, M., Thomma, B., 2010. Emerging viral diseases of tomato crops. *MPMI* 23 (5), 539–548.
- Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., et al., 2010. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* 39, D576–D582 (Database).
- Kloczkowski, A., Ting, K., Jemigan, R., Garnier, J., 2002. Combining the GOR V algorithm with evolutionary information for protein secondary structure prediction from amino acid sequence. *Proteins Struct. Funct. Genet.* 49 (2), 154–166.
- Kuhn, J., Dürrwald, R., Bào, Y., Briese, T., Carbone, K., Clawson, A., et al., 2014. Taxonomic reorganization of the family Bornaviridae. *Arch. Virol.* 160 (2), 621–632.
- Mettenleiter, Thomas C., Sobrino, Francisco, 2008. *Animal Viruses*. Caister Academic Press, Norfolk, UK.
- Mount, David W., 2001. *Bioinformatics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Pappu, H., Jones, R., Jain, R., 2009. Global status of tospovirus epidemics in diverse cropping systems: successes achieved and challenges ahead. *Virus Res.* 141 (2), 219–236.
- Parrella, G., Gognalons, P., Gebre-Selassie, K., Vovlas, C., Marchoux, G., 2003. An update of the host range of tomato spotted wilt virus. *J. Plant Pathol.* 85, 227–264.
- Qian, N., Sejnowski, T., 1988. Predicting the secondary structure of globular proteins using neural network models. *J. Mol. Biol.* 202 (4), 865–884.
- Riley, D., Joseph, S., Srinivasan, R., Diffie, S., 2011. Thrips vectors of tospoviruses. *J. Integr. Pest Manag.* 1, 1–10.
- Rotenberg, D., Jacobson, A., Schneweis, D., Whitfield, A., 2015. Thrips transmission of tospoviruses. *Curr. Opin. Virol.* 15, 80–89.
- Sakthivel, S., S.K.M. H., 2015. NNvPDB: neural network based protein secondary structure prediction with PDB validation. *Bioinformation* 11 (8), 416–421.
- Seto, D., 2010. Viral genomics and bioinformatics. *Viruses* 2 (12), 2587–2593.
- Walter, C., Barr, J., 2011. Recent advances in the molecular and cellular biology of bunyaviruses. *J. Gen. Virol.* 92 (11), 2467–2484.
- Whitfield, A., Ullman, D., German, T., 2005. Tospovirus–thrips interactions. *Annu. Rev. Phytopathol.* 43 (1), 459–489.