

ISSN 1996-0719

International Journal of  
**Plant**  
Pathology

<http://knowledgiascientific.com>

**Knowledgia**  
SCIENTIFIC  
A Place to Publish Outstanding Research

## Genomics and Proteomics Characterization of Alphasatellite in Weed Associated with Begomovirus

Rajneesh Prajapat, Avinash Marwal, Vivek Bajpai and Rajarshi Kumar Gaur

Department of Science, Faculty of Arts, Science and Commerce, Mody Institute of Technology and Science, Lakshmangarh, Sikar-332311, Rajasthan, India

Corresponding Author: Rajarshi Kumar Gaur, Department of Science, Faculty of Arts, Science and Commerce, Mody Institute of Technology and Science, Lakshmangarh, Sikar-332311, Rajasthan, India Tel: +91-9352911723

### ABSTRACT

The uncharacterized alphasatellite of begomovirus associated with a common weed *Verbesina encelioides* was characterized by using molecular and *in silico* tools and techniques. *Verbesina encelioides* leaf curl alphasatellite (HQ631431) shows 87 % nucleotide sequence identity with *Sida yellow vein disease* associated DNA1 (FN806782). Translated 3-frame of HQ631431 (HQ631431/3-f.pdb) and its homologous 2HWT.pdb had respectively 67 and 77.2% residues in most favorable region of its Ramachandran plot therefore both of these two models cannot be placed in good quality. The 3d2GO server showed the hydrolase activity as possible function for HQ631431/3-f.pdb with 0.58% confidence level. *In silico* prediction of results can be used to confirm the begomovirus not only in host weeds but also in other crops also.

**Key words:** *Verbesina encelioides*, begomovirus, alphasatellite, HQ631431, homology modeling

### INTRODUCTION

Medicinal important weeds and other economical important plants can be affected by different type of diseases that shows a wide range of symptoms and the causative agents of these diseases are of biotic or abiotic nature (Simsek *et al.*, 2002). Among the biotic disease agents, viruses can attack all types of plants. Viruses are infectious pathogens that are too small to be seen with a high power microscope and can replicate only inside the living cells of organisms (Koonin *et al.*, 2006).

Geminiviruses are plant viruses that belong to the family Geminiviridae (Goodman, 1977a, b; Fauquet *et al.*, 2005) and characterized by the unique Gemini shape of a fused icosahedral viral particle. Geminiviruses have circular single stranded DNA genomes encapsulated in twinned quasi isometric particles and are responsible for major crop losses worldwide (Moffat, 1999). The family Geminiviridae is divided into four genera (*Mastrevirus*, *Curtovirus*, *Topocuvirus* and *Begomovirus*), based on genome structure, type of insect vector and host range (Stanley *et al.*, 2005; Medina-Ramos *et al.*, 2008). Begomovirus is the largest genus of family *Geminiviridae* (Dhakar *et al.*, 2010) and obligatory transmitted by the whitefly *Bemisia tabaci* (Funayama *et al.*, 2001; Sidhu *et al.*, 2009; Govindappa *et al.*, 2011) that are common in the tropical and subtropical regions of the world.

The geminivirus genome is organized in one (monopartite) or two (bipartite) covalently closed, circular, ssDNA molecules of about 2.5-2.9 kb. The genes in monopartite and bipartite

geminiviruses are arranged in two divergent clusters 280 to 350 nucleotides, each separated by the Intergenic Region (IR) (Lazarowitz and Shepherd, 1992). The single genomic component of monopartite geminiviruses (Mastreviruses and Curtoviruses) contains all the information necessary for virus replication and infectivity (Hanley-Bowdoin *et al.*, 1996). Bipartite Begomoviruses have seven genes distributed in the two genomic components (Fig. 1) designated A and B (Tiendrebeogo *et al.*, 2008). The A component contains genes involved in virus replication and encapsidation and the B component contains the genes involved in virus movement. The A and B components each have a common region which consists of a block of approximately 200 bp within the IR (Sunter and Bisaro, 1991). These common regions are identical in a given bipartite begomovirus but completely different in among the other geminivirus with the small exception of 30 nucleotide stem loop (identified as origin of replication) region. The common region also contains two divergent promoters which differentially regulate the temporal expression of the viral genes.

The Alphasatellite components are satellite-like, circular ssDNA molecules of approximately 1375 nucleotides in length (Fig. 2). They encode a single gene, a rolling circle replication initiator protein and are capable of autonomous replication in plant cells. Closely related to the replication associated protein encoding components of nanoviruses (a second family of plant infecting ssDNA viruses) (Vetten *et al.*, 2004), from which they are believed to have evolved, they require a helper Begomovirus for movement within and between plants (Mansoor *et al.*, 1999; Saunders and Stanley, 1999). Interestingly, alphasatellite components are phenotypically silent; playing no part in the symptoms of the complex and their precise function remains unclear (Tiendrébéogo *et al.*, 2010).

*Verbesina encelioides* (Astraceae) is an erect annual commonly seen weed up to heights of 1 to 5 feet and distributed not only in India but also several regions of the world (Ball *et al.*, 1951; Robbins *et al.*, 1951; Sindhu *et al.*, 2010). Leaves of *Verbesina encelioides* are toothed or lobed (Everist, 1957) and have two distinct growth patterns: lower leaves are opposite and triangular, while the upper leaves are alternate and lanceshaped (Wagner *et al.*, 1990). Both upper and lower leaves feature fine white hairs on the underside (Everist, 1957). These fine white hairs are also present on the stem of *Verbesina encelioides* which grows from a taproot system (Parker, 1972).

Begomovirus associated symptoms were observed in several *Verbesina encelioides* growing as common weed in Rajasthan (India). The symptoms of the disease consist of mosaic and curling of leaf, leaf yellowing and stunting of plants. Begomovirus was suspected as a causative pathogen because of observed large population of whitefly (*Bemisia tabaci*, the vector of Begomovirus (Geminiviridae)) on the plants. DNA isolated from leaves with symptoms and apparently without symptoms by using CTAB (Cetyl Trimethyl Ammonium Bromide) method and subjected to Rolling

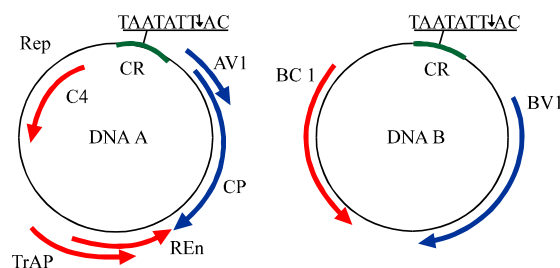


Fig. 1: Genomic Organization of a Begomovirus distributed in the two genomic components designated A and B

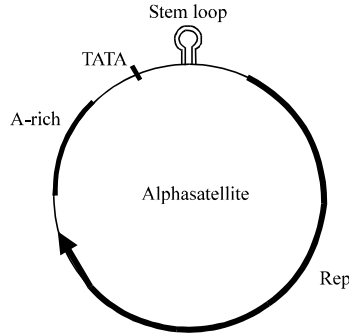


Fig. 2: Structure of alphasatellite component, encodes a single gene in the positive orientation (a replication initiator protein (Rep)) and a region of sequence rich in adenine (a-rich)

Circular Amplification (RCA) which uses a high fidelity  $\phi$  29 DNA polymerase along with random hexamers to detect the genomes of various Begomoviruses. Restriction analysis was performed on positive RCA products in order to detect mixed Begomovirus infections. The digested product was subjected to cloning and partially sequenced by using M13 universal primer and sequence information was submitted in NCBI-bankIt (GenBank accession HQ631431).

At the first stage sequence analysis begins with sequence alignment by using any one of the BLAST programs (BLASTn, BLASTp etc.). After sequence alignment different bioinformatics tools and techniques can be used for homology modeling procedure which can be divided into four sequential steps: template selection, target template alignment, model construction and model assessment (Marti-Renom *et al.*, 2000). Homology modeling is a helpful tool for the production of 3D structure of protein by using *in silico* methods and by this approach we can overcome the limitations of X-Ray crystallography and NMR. The selected model can be validated by using protein structure checking tools such as Procheck for reliability and 3d2GO for function prediction server.

Bioinformatics tools and techniques based sequence analysis has application of research and development in the field of genomics, transcriptomics and proteomics at gene expression, structural and function production level. The *in silico* drug designing principles can be applied for *in silico* antiviral agent designing and that is the new path of research in the field of plant virology. The aim of the study was, characterization of *Verbesina encelioides* leaf curl alphasatellite (HQ631431) by using molecular (e.g., DNA isolation, Rolling Circular Amplification [RCA], cloning, sequencing) and *in silico* (e.g., NCBI-bankIt sequence submission, Homology modeling and function prediction by using 3d2GO) techniques.

## MATERIALS AND METHODS

**Samples collection and DNA extraction:** Due to the current realization that weed-infecting begomoviruses may infect crops and those weeds may serve as a reservoir for crop-infecting geminiviruses (Roye *et al.*, 1997) we made an extensive survey for the epidemiology of begomovirus in 2009-2010. We collected a begomovirus associated with a common weed namely *Verbesina encelioides* across Rajasthan, India. To investigate the possibility the leaf samples from *Verbesina* sp. exhibiting typical begomovirus symptoms, including yellow and golden mosaic, chlorotic mottling and blistering, were taken for the study. Total DNA from plant samples were extracted by using a CTAB (Cetyl Trimethyl Ammonium Bromide) method (Manen *et al.*, 2005).

**Rolling Circular Amplification (RCA):** Isolated DNA was subjected to Rolling Circular Amplification (RCA) which uses a high fidelity  $\phi$ 29 DNA polymerase along with random hexamers to detect the genomes of various Begomoviruses. Genome concatemers generated during amplification were digested with restriction enzymes to release unit-length genomes. A preliminary screening was carried out to identify enzymes capable of linearizing amplification products. After digestion, genomic DNA was gel-purified and cloned into pBS KS (Stratagene).

Restriction analysis was performed on positive RCA products in order to detect mixed Begomovirus infections. Each sample was cleaved with restriction enzymes: *EcoRI* and *SalI* essentially. The digested product was taken for cloning and partial sequencing by using M13 universal primer. After sequencing the nucleotide sequence was submitted in NCBI-bankIt (GenBank accession HQ631431).

**Retrieval of sequences from NCBI translation:** Published information about complete nucleotide sequence of *Verbesina encelioides* leaf curl alphasatellite (>gi|319428633|gb|HQ631431.1) was retrieved from GenBank-NCBI (<http://www.ncbi.nlm.nih.gov/>) in FASTA format and was used for BLASTn alignment.

Nucleotide sequence of *Verbesina encelioides* leaf curl alphasatellite (HQ631431) was translated into protein sequence by using DNA to protein translation server (<http://insilico.ehu.es/translate/>) (Fig. 3).

**Phylogenetic tree construction:** Phylogenetic relationship of HQ631431 was determined based on multiple sequence alignment by using the CLC Main Workbench 5.7. The CLC Main Workbench is developed for Windows, MacOSX and Linux. The software for either platform can be obtained from <http://www.clcbio.com>. The Neighbor Joining algorithms (Saitou and Nei, 1987) were used for construction of phylogenetic trees. Bootstrap values were computed using 100 replicates to evaluate support for the groupings. This analysis clustered each one of the isolates with the other previously sequenced isolate of the respective species.

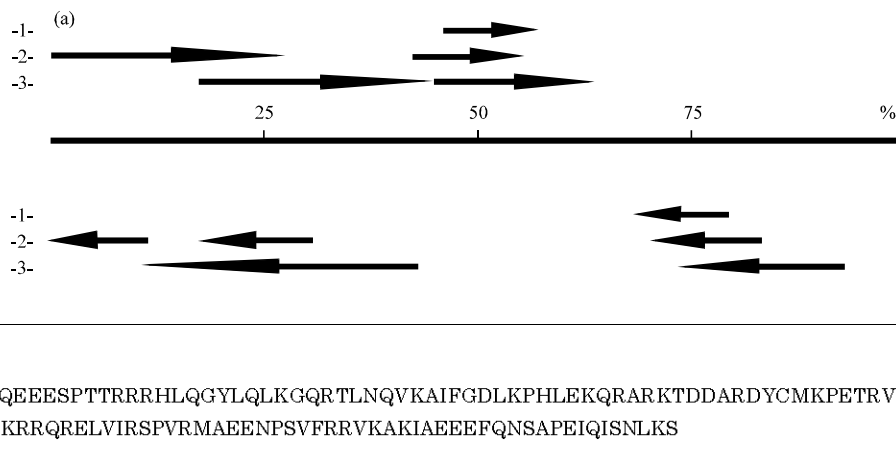


Fig. 3: (a) Different reading frames of complete sequence of *Verbesina encelioides* leaf curl alphasatellite (HQ631431), created by using DNA to protein translation server (<http://insilico.ehu.es/translate/>) that have 132 amino acids residues. (b) Amino acid sequence of 3-frame of HQ631431 (HQ631431/3-f)

**Selection of template:** The homology modeling requires a query sequence with an unknown 3D structure and the target sequence that have known 3D structure with at least 35% similarity. BLAST (Basic Local Alignment Search Tool) (BLASTp 2.2.24+) was used (Altschul *et al.*, 1997, 2005) to search against the PDB (Protein Databank) to find out the related homologues of the query/template (3-frame of HQ631431 or HQ631431/3-f) sequence (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). By the BLASTp search of translated protein HQ631431/3-f against Protein Data Bank (PDB), the Master-Rep protein nuclease domain (2-95) from the *Faba Bean Necrotic yellows virus* (2HWT) showed highest sequence identity of 40%, positives 64% and gaps were 5%.

The PDB file of the target sequence of 2HWT was downloaded from PDB (<http://www.rcsb.org/pdb>) and the FASTA format of target sequence (PDB: 2HWT) was mined from GenBank-NCBI as follows:

```
> gi|151567636|pdb|2HWT|A Chain A, NMR Solution Structure Of The Master-Rep Protein Nuclease Domain (2-95)
From The Faba Bean Necrotic Yellows Virus ARQVICWCFTLNNPLSPLSL HDSMKYLVYQTEQGEAGNIHFQGYI
EMKKRTSLAGMKKLIPG AHFEKRRGTQGEARAYS MKEDTRLEGPWEYGE
```

The 3-frame of HQ631431 (HQ631431/3-f) was find suitable for sufficiently close template in structure database to confidently model of the sequence and rest of translated frames were not find suitable.

```
>PTLLCFGQEEESPTTRRRHLQGYLQLKGQRTLNVKALFVGLKPHLEKQRARKTDDARDYCMKPETRVSGPFGEYCPAG
SHKRRQRELVIKRVFVMAEENPSVFRVVKAKIAEEEFQNSAPEIQISNLKS
```

Those FASTA sequences of template 3-frame of HQ631431 (HQ631431/3-f) and target (2HWT) were uploaded on the 3D-JIGSAWN (Protein Comparative Modeling Server) for the construction of its PDB files. 3D-JIGSAWN (<http://bmm.cancerresearchuk.org/~3djigsaw/>) sends the PDB files on the e-mail address that assigned to the modeling server. The PDB file of query and homologous target sequence were further utilized for 3D model energy validation and docking studies (Heinrichs, 2008). The predicted model was validated with the program Procheck and Ramachandran plot statistics was used to evaluate the stability of the model.

**Model evaluation and validation:** The UCLA-DOE (<http://nihserver.mbi.ucla.edu/Verify-3D/>) Structure Evaluation server provide a visual analysis of the quality of a putative crystal structure for protein. Verify 3D expects this crystal structure to be submitted in PDB format. The validation for structure models obtained from the three software tools was performed by using PROCHECK (Laskowski *et al.*, 1996). In order to study the energy validation of query and homologous target proteins, we uploaded PDB files of both proteins on structure analysis and validation server (SAVA, <http://nihserver.mbi.ucla.edu/SAVES/>). SAVA service provides by NIH-MBI laboratory for structural genomics and proteomics (Bowie *et al.*, 1991; Luthy *et al.*, 1992) analysis. In this study the model was checked with Verified-3D (Fig. 4) and Ramachandran plot through Procheck (Fig. 7) server (Prajapat *et al.*, 2011).

PDB file of query (HQ631431/3-f.pdb) utilized for the ribbon representation model construction using 3dLigandSite server Jmol (Fig. 4a) and homologous target protein (2HWT) NMR solution structure that was uploaded from Protein Data Bank (Fig. 4b).

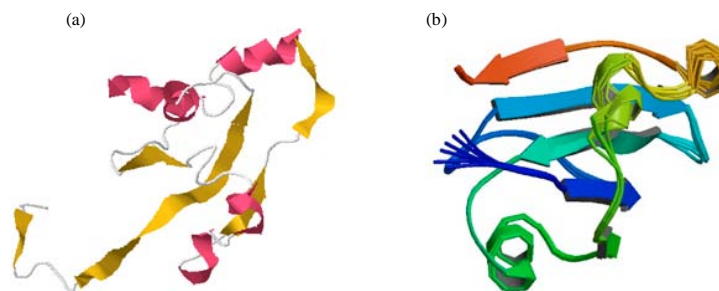


Fig. 4: (a) Ribbon representation of model of 3-frame protein sequence of HQ631431/3-f.pdb created by using 3dLigandSite server Jmol.  $\alpha$ -helices (Pink) and  $\beta$ -sheets (Yellow) are shown as helices and ribbons (b) 2HWTNMR solution structure of the Master-Rep protein nuclease domain (2-95) from the Faba Bean Necrotic Yellows Virus that uploaded from PDB, as best homologous HQ631431/3-f.  $\alpha$ -helices (Green) and  $\beta$ -sheets (Red, Green, Blue and Sky color) are shown as helices and ribbons.

**Function prediction:** The 3d2GO server (<http://www.sbg.bio.ic.ac.uk/phyre/pfd/html/help.html>) was used for the prediction of functions of the predicted model using sequence and structure in the reference of Gene Ontology (GO). 3D2GO predicts the function of protein using some information e.g., overall topological similarity to structures with known function, geometric and residue similarity of predicted functional sites to regions of known structures and sequence homology to functionally annotated sequences. The MAMMOTH structural alignment program was used for full topology search of the model (Ortiz *et al.*, 2002). The MUSCLE program was used for functional site prediction of the predicted model (Edgar, 2004). Functional residue prediction was done using the Jenson-Shannon Divergence (JS Divergence), an information-theory approach to determine relative residue conservation (Capra and Singh, 2007) such conservation is related to the functional importance of residues.

**3D-ligand site prediction:** The Protein ligand binding residues was predicted by using program 3dLigandSite (<http://www.sbg.bio.ic.ac.uk/3dligandsite/>) which uses Critical Assessment of protein Structure Prediction experiment (CASP) (Wass and Sternberg, 2009). This was based on the approach to identify binding sites by combining the use of the predicted structure of the targets with both residue conservation and the location of ligands bound to homologues structures.

## RESULTS AND DISCUSSION

*Veresina encelioides* symptoms are of bright yellow spots finds along the midrib which coalesce to give a mosaic, a reduction in leaf size and stunting of the plant. Field survey showed that *Veresina encelioides* is very common in Rajasthan state (India), it grow in water rich places of the field and show competition with economical important crop plants for nutrition and water and may serve as a reservoir for crop-infecting geminiviruses. Overall results of molecular analysis indicate that there are high infection rate of Begomovirus persisting with yellow mosaic disease of *Verbesina encelioides* in Rajasthan (India).

Positive RCA products were restriction analyzed in order to detect mixed begomovirus infections and each sample was cleaved with restriction enzymes: *EcoRI* and *SalI* essentially. All the samples digested with *EcoRI* and *SalI* in RCA, a full-length fragment at approximately 1.3 kb was found.

Complete sequence of *Verbesina encelioides* leaf curl alphasatellite (HQ631431) was used as query for template search by using BLASTn program. The alignment of HQ631431 by using BLASTn revealed the highest 85% nucleotide sequence identities (Query coverage 100%) with *Cyamopsis tetragonoloba* leaf curl alphasatellite (GU385877), 86% (Query coverage 92%) with *Tobacco leaf curl PUSA alphasatellite* (HQ180392) also 87% (Query coverage 87%) with *Sida yellow vein disease associated DNA 1* (FN806782).

Phylogenetic relationship were determined based on multiple sequence alignment of complete nucleotide sequence of *Verbesina encelioides* leaf curl alphasatellite (HQ631431), performed using the CLC Main Workbench 5.7. Bootstrap values were computed using 100 replicates to evolution support for the groupings. The *Verbesina encelioides* leaf curl alphasatellite (HQ631431) placed in monophyletic clusters with 100 bootstrap value of *Cyamopsis tetragonoloba* leaf curl alphasatellite (GU385877) in Neighbor-Joining tree (Fig. 5).

Complete sequence of *Verbesina encelioides* leaf curl alphasatellite (HQ631431) was converted into 6 different reading frames through DNA to protein translation server (<http://insilico.ehu.es/translate/>). All these frames protein sequences was uploaded on 3D-JIGSAW protein comparative modeling server ([bmm.cancerresearchuk.org/~3djigsaw/](http://bmm.cancerresearchuk.org/~3djigsaw/)). Sufficient close template was generated only for first 3-reading protein sequence frame of HQ631431 (HQ631431/3-f), to confidently model this sequence and find unsuccessful with rest of other reading frames.

During BLASTp analysis of query sequence (3-frame protein sequence of HQ631431 or HQ631431/3-f) against PDB (<http://www.rcsb.org>) showed 40% identity with Chain A of master rep protein nuclease domain of the Faba Bean Necrotic Yellows Virus (2HWT). The PDB file of 2HWT was uploaded from Protein Data Bank (<http://www.rcsb.org>) for further analysis.

Further Verify-3D and Procheck (Laskowski *et al.*, 1993) was used to perform full geometric analysis as well as stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry.

Verify-3D (Eisenberg *et al.*, 1997) works best on proteins with at least 100 residues. Verify3D analyzes the compatibility of an atomic model (3D) with its own amino acid sequence.

The Profile score above zero in the Verify 3D graph corresponds to acceptable environment of the model (Mau Goh *et al.*, 2008). The high score of 0.29 indicate that environment profile of the model is good (Fig. 6).

Homology modeling study is an important method to know the 3D structure of the protein whose structure is not available. After running Procheck, Ramachandran plot of query protein sequence (3-frame of HQ631431 or HQ631431/3-f.pdb) had only 67% residues were in core/favored region, 28.4% in the additional allowed region, 4.5% in the generously allowed region and 0% of the residues in the disallowed region (Fig. 7a) which made this model more acceptable as compared to other predicted models (Table 1). Similar approach was also used in the prediction of homologous 2HWT. Ramachandran plot of 2HWT.pdb had 77.2% in core, 19.0% in allowed, 2.5% in gener and 1.3% disallowed regions (Fig. 7b).

A good quality Ramachandran plot have over 90% in the most favoured or core regions (Prajapat *et al.*, 2010). Ramachandran plot of HQ631431/3-f.pdb had 67% residues in the most



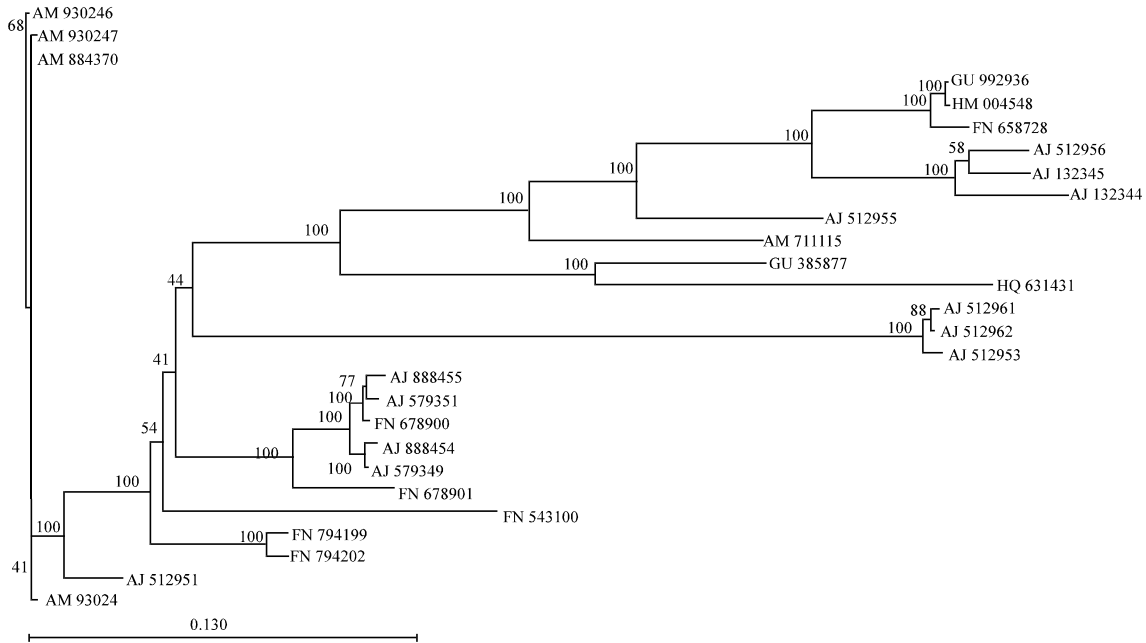


Fig. 5: A Neighbor-Joining tree based on complete nucleotide sequence of *Verbesina encelioides* leaf curl alphasatellite (HQ631431) and other begomovirus sequences available in GenBank. Bootstrap values at major nodes are indicated. Horizontal distances are proportional to the genetic distance between isolates and vertical distances are arbitrary. Scale bar indicates the proportion of sites changing along each branch. The *Verbesina encelioides* leaf curl alphasatellite (HQ631431) used for comparison: Nanovirus-like particle rep gene (AM930246), Nanovirus-like particle rep gene (AM930247), Nanovirus-like particle, rep gene (AM884370), Cotton leaf curl Burewala alphasatellite (GU992936), Cotton leaf curl Burewala alphasatellite (HM004548), Cotton leaf curl Burewala alphasatellite (FN658728), Nanovirus-like particle rep gene (AJ512956), Nanovirus-like particle rep gene (AJ132345), Nanovirus-like particle rep gene (AJ132344), Nanovirus-like particle rep gene (AJ512955), *Cyamopsis tetragonoloba* leaf curl alphasatellite (GU385877), Nanovirus-like particle rep gene (AJ512961), Nanovirus-like particle partial rep gene (AJ512962), Nanovirus-like particle partial rep gene (AJ512953), Tobacco leaf curl Yunnan virus (AJ888455), Tobacco curly shoot virus associated DNA 1 rep gene (AJ579351), Tobacco curly shoot virus associated DNA 1 rep gene (FN678900), Tobacco curly shoot virus associated DNA 1 rep gene (AJ579349), Tobacco curly shoot virus associated DNA 1 rep gene (FN678901), Ageratum enation virus associated DNA 1 (FN543100), Nanovirus-like particle (FN794199), Nanovirus-like particle (FN794202), Nanovirus-like particle rep gene (AJ512951), Nanovirus-like particle rep gene (AM930244)

favoured region and 2HWT.pdb has 77.2% residues in the most favoured (Table 1), therefore, both of these models cannot place in good quality.

Protein 2HWT.pdb is more stable than HQ631431/3-f.pdb due to high percentage in core/favored region and less percentage of residues in allowed region and generously allowed

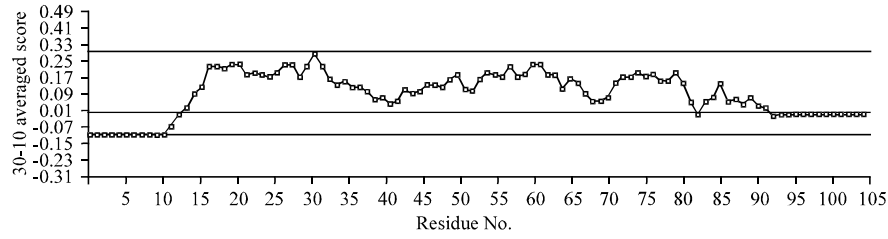


Fig. 6: Verified 3D graph/Profile search plot of HQ631431/3-f.pdb (No. of Residues = 104) with score of 0.29

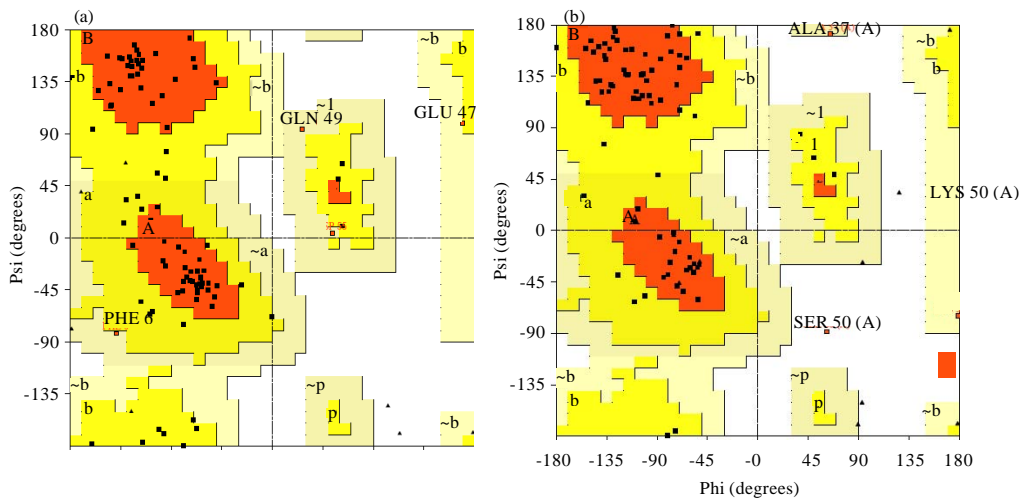


Fig. 7: Ramachandran plot of HQ631431/3-f.pdb and its homologous 2HWT.pdb, model predicted using PROCHECK

Table 1: Comparative analysis of Ramachandran statistics in the HQ631431/3-f.pdb and homologous 2HWT.pdb predicted models

Predicted structure	Ramachandran statistics			
	Most favored region	Additional allowed region	Generously allowed region	Disallowed region
HQ631431/3-f.pdb	67.0	28.4	4.5	0.0
2HWT.pdb	77.2	19.0	2.5	1.3

Table 2: Residues in generously allowed region that affect the stability of protein

Amino acid	Position	PHI	PSI	Omega	Chi1	Chi2
Phe	6	-138.70	-82.60	-179.79	175.56	80.77
Gln	49	26.32	93.86	176.68	-59.67	-85.19
Glu	47	169.52	99.46	-173.25	-171.44	170.72
Asp	55	53.58	4.01	-170.87	-57.23	-53.73

Table 3: Result showing the function prediction of the query modeled protein with 3d2GO (Protein function prediction server)

GO term	Description	Confidence
0016787	Hydrolase activity	0.58
0043169	Cation binding	0.50
0003824	Catalytic activity	0.46
0007411	Axon guidance	0.40
0043167	Ion binding	0.37

Table 4: Different ligand clusters information shows that Cluster 1 has ligands and structures (1 each) with the average Mammoth

Cluster	Ligands	Structures	MAMMOTH scores		
			Av	Min.	Max.
1	1	1	7.1	7.1	7.1

Table 5: List of amino acid residues observed in cluster 1 of predicted protein with number of contacts of ligand, average distance and JS divergence

Residue	Amino acid	Contact	Av distance	JS divergence
74	Phe	1	0.39	0.49
75	Gly	1	0.00	0.58
76	Glu	1	0.00	0.21

region of Ramachandran plot. Phe 6, Gln 49, Glu 47 and Asp 55 present in the generously allowed region, therefore, they may affect the stability of protein (Table 2).

The 3d2GO server uses several methods of function prediction, using sequence and structure, to predict Gene Ontology (GO) terms for the protein (Shankaracharya *et al.*, 2011). Various GO terms, their description and the confidence have been listed in Table 3. Confidence ranges from 0 to 1, with 1 being the most confident prediction. Result show that the predicted query protein HQ631431/3-f.pdb file has functions like hydrolase activity and cation binding with good confidence (Table 3) which confirms its functions.

Hence, the result shows that Tyr60, Gln21, Lys63, Cys61, Tyr23 and Ala57 were more conserved residue in the structure. The predicted site one had Gln21, His45, Tyr23, Ala57, Leu20, Gly22 and in the predicted site two have Tyr60, Gln21, Ala57, Met62, Lys48, Lys53.

Different ligand clusters information shows that Cluster 1 has ligands and structures (1 each) with the average MAMMOTH score (Shankaracharya *et al.*, 2010) of 7.1 (Table 4). Phe74, Gly75 and Glu76 List observed in cluster 1 of predicted protein and there number of contacts, average distance and JS divergence shown in the Table 5.

Prediction of protein 3D structure helps us to find out their active sites, binding site (Amir *et al.*, 2010) etc. The model developed through homology modeling and subsequently the predicted functional characteristics of 3-frame of HQ631431 (HQ631431/3-f) provide a base for structural and functional characteristics of *Verbesina encelioides* leaf curl alphsatellite (HQ631431/3-f) and suitable role in leaf curl mechanism therefore, provide better control on begomovirus infection not only for *Verbesina encelioides* but also for other plants species infected by begomovirus in various tropical and subtropical part of the world.

## CONCLUSION

*Verbesina encelioides* (family Asteraceae) is a common weed plant in the fields of Rajasthan, India. In recent years, large numbers of plants were exhibiting typical Begomovirus infection

symptoms. Isolated DNA from leaf samples was subjected to Rolling Circular Amplification (RCA) to detect the genomes of various Begomoviruses. All samples were digested with *EcoRI* and *SaII* and a full-length fragment of size approximately 1.3 kb was obtained. RCA product was sequenced and submitted to GenBank-NCBI (accession HQ631431). *Verbesina encelioides* leaf curl alphasatellite (HQ631431) sequence alignment was done by using BLASTn that revealed about 87% nucleotide sequence identity with *Sida yellow vein disease associated DNA 1* (FN806782). HQ631431 placed in monophyletic clusters with 100 bootstrap value of *Cyamopsis tetragonoloba* leaf curl alphasatellite (GU385877) in Neighbor-Joining tree.

Homology modeling and function prediction study of translated 3-frame of HQ631431 (HQ631431/3-f) was performed, thus shown highest homology with 2HWT. HQ631431/3-f.pdb had 67% residues and 2HWT.pdb had 77.2% residues in the most favoured region of its Ramachandran plot, therefore, both of these models cannot place in good quality. The hydrolase activity, cation binding and catalytic activity were predicted for HQ631431/3-f as important functional site of the model with high confidence through 3d2GO (Protein function prediction server). Tyr60, Gln21, Lys63, Cys61, Tyr23 and Ala57 were found as more conserved residue in the structure. According to our best knowledge it is the first report regarding functional characteristics of *Verbesina encelioides* leaf curl alphasatellite (HQ631431/3-f), that can be used in future for designing *in silico* antibegomoviral agent by using docking tools and techniques. Results of this study will contribute in control begomoviral epidemic level and reduce yield loss in crops of tropical and subtropical regions of the world.

#### ACKNOWLEDGMENT

The authors are thankful to Prof. Shakti Baijal, Dean, FASC, MITS, Rajasthan, India. The authors are also thankful to Department of Biotechnology (DBT), India and Department of Science and Technology, India for financial support for the present studies.

#### REFERENCES

- Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.*, 25: 3389-3402.
- Altschul, S.F., J.C. Wootton, E.M. Gertz, R. Agarwala, A. Morgulis, A.A. Schaffer and Y.K. Yu, 2005. Protein database searches using compositionally adjusted substitution matrices. *FEBS J.*, 272: 5101-5109.
- Amir, A., M.A. Siddiqui, N. Kapoor, A. Arya and H. Kumar, 2010. *In silico* molecular docking of influenza virus (PB2) protein to check the drug efficacy. *Trends Bioinform.*, (In Press).
- Ball, W.S., A.S. Crafts and R.N. Raynor, W.W. Robbins, 1951. *Principals of Weed Control*. In: *Weeds of California*, Robins, W.W. (Ed.). McGraw-Hill Book Co., New York, pp: 13-15.
- Bowie, J.U., R. Luthy and D. Eisenberg, 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253: 164-170.
- Capra, J. and M. Singh, 2007. Predicting functionally important residues from sequence conservation. *Bioinformatics*, 23: 1875-1882.
- Dhakar, K., V.K. Gupta, M.S. Rathore and R.K. Gaur, 2010. Virus resistance and gene silencing in plants infected with begomovirus. *J. Applied Sci.*, 10: 1787-1791.

- Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.*, 32: 1792-1797.
- Eisenberg, D., R. Luthy and J.U. Bowie, 1997. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.*, 277: 396-404.
- Everist, S.I., 1957. *Common Weeds of Farm and Pasture*. A.H. Tucker, Government Printer, Brisbane.
- Fauquet, C.M., M.A. Mayo, J. Maniloff, U. Desselberger and L.A. Ball, 2005. Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, USA., pp: 1259.
- Funayama, S., I. Terashima and T. Yahara, 2001. Effects of virus infection and light environment on population dynamics of *Eupatorium makinoi* (Asteraceae). *Am. J. Bot.*, 88: 616-622.
- Goh, K.M., N.M. Mahadi, O. Hassan, R.N. Zaliha R.A. Rahman and R.M. Illias, 2008. Molecular modeling of a predominant  $\beta$ -CGTase G1 and analysis of ionic interaction in CGTase. *Biotechnology*, 7: 418-429.
- Goodman, R.M., 1977a. A new kind of virus is discovered. *Illinois Res.*, 19: 5-5.
- Goodman, R.M., 1977b. Single stranded-DNA genome in a whitefly-transmitted plant virus. *Virology*, 83: 171-179.
- Govindappa, M.R., I. Shankergoud, K.S. Shankarappa, W.A.R.T. Wickramaarachchi, B.A. Reddy and K.T. Rangaswamy, 2011. Molecular detection and partial characterization of begomovirus associated with leaf curl disease of sunflower (*Helianthus annuus*) in Southern India. *Plant Pathol. J.*, (In Press)
- Hanley-Bowdoin, L., P.A. Eagle and B.M. Orozco, 1996. Gemini Virus Replication. In: *Biology of Plant-Microbe Interactions*, Stacey, G., B. Mullin and P.M. Gesshoff (Eds.). International Society for Molecular Plant-Microbe Interactions, St. Paul, MN., pp: 287-292.
- Heinrichs, A., 2008. Proteomics: Solving a 3D jigsaw puzzle. *Nat. Rev. Mol. Cell Biol.*, 9: 3-3.
- Koonin, E.V., T.G. Senkevich and V.V. Dolja, 2006. The ancient Virus World and evolution of cells. *Biol. Direct.*, 1: 29-29.
- Laskowski, R.A., M.W. MacArthur, D.S. Moss and J.M. Thornton, 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Applied Cryst.*, 26: 283-291.
- Laskowski, R.A., J.A. Rullmannn, M.W. MacArthur, R. Kaptein and J.M. Thornton, 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR.*, 8: 477-486.
- Lazarowitz, S.G. and R.J. Shepherd, 1992. Geminiviruses: Genome structure and gene function. *Crit. Rev. Plant Sci.*, 11: 327-349.
- Luthy, R., J.U. Bowie and D. Eisenberg, 1992. Assessment of protein models with three-dimensional profiles. *Nature*, 356: 83-85.
- Manen, J.F., O. Sinitsyna, L. Aeschbach, A. Markov and A. Sinitsyn, 2005. A fully automatable enzymatic method for DNA extraction from plant tissues. *BMC Plant Biol.*, 5: 23-23.
- Mansoor, S., S.H. Khan, A. Bashir, M. Saeed and Y. Zafar *et al.*, 1999. Identification of a novel circular single-stranded DNA with cotton leaf curl disease in Pakistan. *Virology*, 259: 190-199.
- Marti-Renom, M.A., A.C. Stuart, A. Fiser, R. Sanchez, F. Melo and A. Sali, 2000. Comparative protein structure modeling of genes and genomes. *Ann. Rev. Biophys. Biomol. Struct.*, 29: 291-325.
- Medina-Ramos, G., R. De La Torre-Almaraz, R. Bujanos-Muniz, R.G. Guevara-Gonzalez and N. Tierranegra-Garcia *et al.*, 2008. Co-transmission of *Pepper huastecoyellow vein virus* and *Pepper golden mosaic virus* in chili pepper by *Bemisia tabaci* (Genn.). *J. Entomol.*, 5: 176-184.

- Moffat, A.S., 1999. Plant pathology: Geminiviruses emerge as serious crop threat. *Science*, 286: 1835-1835.
- Ortiz, A.R., C.E. Strauss and O. Olmea, 2002. MAMMOTH (Matching molecular models obtained from theory): An automated method for model comparison. *Protein Sci.*, 11: 2606-2621.
- Parker, K.F., 1972. An Illustrated Guide to Arizona Weeds. University of Arizona Press, Arizona, pp: 322-323.
- Prajapat, R., R.K. Gaur, R. Raizada and V.K. Gupta, 2010. *In silico* analysis of genetic diversity of begomovirus using homology modelling. *J. Biol. Sci.*, 10: 217-223.
- Prajapat, R., V. Bajpai and R.K. Gaur, 2011. Structural modeling and validation of rep protein of begomovirus strains (TLCBV and TYLCTHV). Nature Publishing Group. <http://precedings.nature.com/documents/5617/version/1>.
- Robbins, W.W., M.K. Bellue and W.S. Ball, 1951. Weeds of California. State Department of Agriculture, Sacramento, CA., pp: 505-506.
- Roye, M.E., W.A. McLaughlin, M.K. Nakhla and D.P. Maxwell, 1997. Genetic diversity among geminiviruses associated with the weed species *Sida* sp., *Macroptilium lathyroides* and *Wissadula amplissima* from Jamaica. *Plant Dis.*, 81: 1251-1258.
- Saitou, N. and M. Nei, 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4: 406-425.
- Saunders, K. and J. Stanley, 1999. A nanovirus-like DNA component associated with yellow vein disease of *Ageratum conyzoides*: Evidence for interfamilial recombination between plant DNA viruses. *Virology*, 264: 142-152.
- Shankaracharya, S. Das, D. Prasad and A.S. Vidyarthi, 2010. Theoretical Investigations on Structure and Function of Human Homologue hABH4 of *E. coli* ALKB4. *IBC*, 2: 1-5.
- Shankaracharya, S. Das and A.S. Vidyarthi, 2011. Structure and function prediction of human homologue hABH5 of *E. coli* ALKB5 using *in silico* approach. *Nat. Proceedings*, 10.1038/npre.2011.5597.2
- Sidhu, J.S., R.S. Mann and N.S. Butter, 2009. Deleterious effects of cotton leaf curl virus on longevity and fecundity of whitefly, *Bemisia tabaci* (Gennadius). *J. Entomol.*, 6: 62-66.
- Simsek, O., M. Arici and C. Demir, 2002. Mycoflora of hazelnut and aflatoxin content in hazelnut kernels artificially infected with *Aspergillus parasiticus*. *Nahrung*, 46: 194-196.
- Sindhu, R.K., N. Vasudeva and S.K. Sharma, 2010. Pharmacognostical and preliminary phytochemical Investigations on *Verbesina encelioides* benth roots. *J. Herbal Med. Toxicol.*, 4: 113-118.
- Stanley, J., D.M. Bisaro, R.W. Briddon, J.K. Brown and C.M. Fauquet *et al.*, 2005. Family geminiviridae. In *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*.
- Sunter, G. and D.M. Bisaro, 1991. Transactivation in a geminivirus: AL2 gene product is needed for coat protein expression. *Virology*, 180: 416-419.
- Tiendrebeogo, F., V.S.E. Traore, N. Barro, A.S. Traore, G. Konate and O. Traore, 2008. Characterization of Pepper yellow vein mali virus in *Capsicum* sp. in Burkina Faso. *Plant Pathol. J.*, 7: 155-161.
- Tiendrebeogo, F., P. Lefeuvre, M. Hoareau, J. Villemot and G. Konate *et al.*, 2010. Molecular diversity of Cotton leaf curl Gezira virus isolates and their satellite DNAs associated with okra leaf curl disease in Burkina Faso. *Virol. J.*, 7: 48-48.

- Vetten, H.J., P.W.G. Chu, J.L. Dale, R. Harding and J. Hu *et al.*, 2004. Nanoviridae. In: Virus Taxonomy, Fauquet, C.M., M.A. Mayo, J. Maniloff, U. Desselberger and L.A. Ball (Eds.). 8th Report of The ICTV, Elsevier/Academic Press, London, pp: 343-352.
- Wagner, W.L., D.R. Herbst and S.H. Sohmer, 1990. Manual of the Flowering Plants of Hawaii. Vol. 1. University of Hawaii Press, Bishop Museum Press, Honolulu.
- Wass, M.N. and M.J. Sternberg, 2009. Prediction of ligand binding sites using homologous structures and conservation at CASP8. *Proteins Struct. Funct. Bioinf.*, 77: 147-151.