Acta virologica **63**: 139 – 148, 2019 doi:10.4149/av_2019_202

Characterization and *in silico* proteomic analysis of C2 and C3 proteins of squash leaf curl China virus associated with pumpkin leaf curl disease in Assam, India

G. M. BALDODIYA[1], K. DEVI[1,2], B. K. BORAH[1*], P. D. NATH[3], M. K. MODI[1,2]

[1]Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat-785013, Assam, India; [2]Distributed Information Centre, Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat-785013, Assam, India; [3]Department of Plant Pathology, Assam Agricultural University, Jorhat-785013, Assam, India

Received June 11, 2018; revised July 26, 2018; accepted March 09, 2019

Summary. – The pumpkin leaf curl disease is an emerging disease of pumpkin in Assam, India. Symptomatic pumpkin leaf samples from different locations were immunologically tested using *Begomovirus* specific antibody. PCR with the ELISA-positive samples, using *Geminivirus* universal primers amplified 1.4 kb virus-specific fragments. Sequence of these amplicons showed around 95% identity with squash leaf curl China virus-[Pumpkin: Varanasi] (SLCCV-Pumpkin: Varanasi EU573715). To investigate the possible functions of the viral proteins present in the fragment, the full-length *C2* and *C3* genes were conceptually translated and were subjected to *in silico* proteomic analyses. The phylogenetic analysis of both the proteins divulged the relationship of our isolate with related viruses and isolates. Multiple sequence alignment (MSA) of the proteins revealed the presence of the known viral conserved motifs, *viz.*, zinc-finger (ZNF) motif [$_{36}$CXCX(7)CX(6)H$_{53}$], the arginine-rich nuclear localization signal (NLS) motif ($_{28}$RRRR$_{31}$) as well as the minimal activation domain in C2 protein. In the C3 protein, the $_{91}$LKYLD$_{95}$ and the replication enhancer motif ($_{30}$YFK$_{32}$) were found to be conserved. Finally, 3-D models of the two proteins were predicted via *ab initio* approach and subsequently, the models were validated. To our knowledge, this study is a pioneering attempt to construct the *ab initio* 3-D models of two begomoviral proteins taking a SLCCV isolate as a model.

Keywords: begomovirus; ELISA; ZNF motif; NLS motif; ab initio modelling

Introduction

Geminiviruses are single-stranded DNA-viruses with geminate (twinned) morphology. The family *Geminiviridae* comprises of nine genera; namely, *Becurtovirus, Begomovirus, Curtovirus, Eragrovirus, Mastrevirus, Topocuvirus, Turncurtovirus, Capulavirus* and *Grablovirus* (10th ICTV report, 2017). While the other genera contain only one genome, begomoviruses contain either one (monopartite, DNA-A) or two (bipartite, DNA-A and DNA-B) genomes.

Both monopartite and bipartite geminiviruses are present in India and they cause significant yield loss in different crops (Borah and Dasgupta, 2012). Pumpkin (*Cucurbita pepo*) is an important vegetable and is a common host of various begomoviruses (Vasudeva and Lal, 1943; Varma, 1955; Capoor and Ahmad, 1975; Jayashree *et al.*, 1999; Maruthi *et al.*, 2003; Muniyappa *et al.*, 2003). In Assam, pumpkin leaf curl disease, caused by a begomovirus, is a rising problem.

The circular 2.8 kb long DNA of the monopartite (and its cognate, DNA-A of the bipartite) begomoviruses codes for six proteins. The genes, V1 and V2, in the viral-sense strand, code for the coat protein and the pre-coat protein,

*Corresponding author. E-mail: basantabora@gmail.com; phone: +91-8011050299.

Abbreviations: BLAST = basic local alignment search tool; GRAVY = grand average of hydropathy; MAD = minimal activation domain, MSA(s) = multiple sequence alignment(s); NLS = nuclear localization signal; SLCCV = squash leaf curl China virus; SLCPV = squash leaf curl Philippines virus; TLCNDV = tomato leaf curl New Delhi virus; TYLCV = tomato yellow leaf curl virus; ZNF = zinc-finger

respectively. The complementary-sense genes, C1, C2, C3 and C4, code for a replication associated protein, a transcriptional activator, a replication enhancer and a suppressor of RNA-silencing, respectively. The similar sized DNA-B of the bipartite begomoviruses contains two genes, BV1, coding for a nuclear shuttle protein, and BC1, coding for a movement protein. Some begomoviruses also harbour one or two types of smaller subviral components called alpha- and beta-satellites.

Geminiviruses use the host's replication-machinery to replicate intermediately producing a double-stranded (ds) genome. To initiate replication, only the C1 protein is required (Elmer *et al*., 1988); it introduces a site-specific nick in to the dsDNA, generating a primer function for DNA synthesis (Laufs *et al*., 1995; Gladfelter *et al*., 1997). The C2 protein contains a NLS, a ZNF-like domain of cysteine and histidine residues, and an acidic activation domain (Hormuzdi *et al*., 1995; Dong *et al*., 2003; Wezel *et al*., 2003; Shen *et al*., 2006). The C2 is a transcriptional activator which also acts as a pathogenicity-factor and suppresses host defence. The protein C3, interacts with C1 and initiates rolling circle replication of the viral genome(s). The C4 has divergent functions, playing a major role in symptom development and in virus movement (Jupin *et al*., 1994; Rigden *et al*., 1994; Teng *et al*., 2010). Interaction of the begomoviral proteins is crucial for successful infection. However, for any kind of protein-interaction study, the foremost requirement is to know the 3-D structures. Until now, no experimental or theoretical models of geminiviral proteins are available. Therefore, as an initial step towards understanding the basis of geminivirus-infection, we undertook this study to get a clear insight of the 3-D structure of two proteins of an identified isolate of squash leaf curl China virus (SLCCV).

Materials and Methods

Viral source and immunological detection. Field survey was conducted in different cultivations of Assam and symptomatic samples were collected. Immunological detection was carried out by indirect ELISA with broad-spectrum tomato yellow leaf curl virus (TYLCV) IgG (Bioreba, Switzerland) as the primary antibody; and anti-rabbit species IgG (Sigma, USA) as the secondary antibody (conjugated with alkaline phosphatase). The developed yellow colour from the para-nitropphenylphosphate substrate was analyzed to detect the antigens using the iMark Microplate Reader S/N 10827 at 405 nm.

DNA isolation, PCR and sequencing. The total DNA was extracted from ELISA-positive samples following the Cetyl trimethyal ammonium bromide (CTAB) method (Doyle and Doyle, 1990). The PCR was performed with universal primer-pair (PALIc1960-PARIv722; Rojas *et al*., 1993) using the total DNA as templates (Briddon and Markham, 1994). The PCR was performed in 15 µl volume, containing *Taq* buffer with 25 mM $MgCl_2$, 4 mM dNTP, 5 pmol of each of the primers, 2.5 unit of *Taq*-polymerase (Himedia, India) and 100 ng of the template. The reactions were performed in Gene Amp® PCR System 9700 at 94°C for 4 min followed by 35 cycles of 94°C for 30 s, 52°C for 45 s and at 72°C for 1 min. The cycling was succeeded by final extension for 7 min at 72°C. The samples were also checked for any associated betasatellite by using the primer-pair beta01-beta02 (Briddon *et al*., 2002). PCR-products were resolved in 1% agarose gel and the products of expected sizes were eluted using GenEluteTM Gen extraction kit (Sigma). Sequencing was commercially done by Bioserve Biotechnologies (I) Pvt. Ltd, India using both the forward and the reverse primers, and contigs were generated in Codon Code Aligner 4.2.7.

Primary sequence analysis. The nucleotide sequences were subjected to BLASTn and BLASTx against the non-redundant-database of NCBI to assess the sequence similarity and the presence of coding regions. Conserved domains were explored by Pfam (http://pfam.sanger.ac.uk/), Simple Modular Architecture Research Tool (http://smart.embl-heidelberg.de) and Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). The NCBI ORF Finder tool was employed to identify the ORFs. The families and super-families of the selected proteins were deduced employing InterProScan. The Protparam tool of ExPaSy server was deployed to analyze physicochemical properties of the proteins, and the CONCORD server (http://helios.princeton.edu/CONCORD/) was used to predict the secondary structures.

Phylogenetic analysis. Amino acid sequence of the selected proteins were subjected to BLASTp against nr-database of NCBI, and the entries producing significant similarities (identity ≥75%, query coverage ≥99%) were aligned with the individual ORFs in CustalW. The multiple sequence alignments (MSAs) were visualized using ESPript. The phylogenetic tree was constructed using the distance matrix by the Neighbor-Joining (NJ) method in MEGA v6.1.0.

Preparation of the 3-D models. To generate the 3-D models, initially, homology-based approaches were tested. Due to much less structural similarity and query coverage in Domain Enhanced Lookup Time Accelerated (DELTA)-BLAST against Protein Data Bank, suitable templates could not be selected. Therefore, *ab initio* approach was opted for their modelling. Among large number of rough models, the models with the lowest C-score were selected and those were subjected to loop and side-chain refinement modloop (https://modbase.compbio.ucsf.edu/modloop/); the models were further assessed with Errat plot (http://nihserver.mbi.ucla.edu/ERRAT/). The topology of the models was analysed using PDBsum (http://www.ebi.ac.uk/pdbsum/) database.

Model validation. The loop-refined models of the proteins were validated using the Procheck module in Profunc (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/profunc/). This module quantifies amino acid residues in the available zones of Ramachandran plot and assesses the stereo-chemical quality of the model.

Results and Discussion

Detection of viral infection

In ELISA, all the 20 pumpkin samples showed positive reaction, indicating the presence of a geminivirus. Using the Rojas universal primer-pair, an approx. 1.4 kb PCR product was obtained from the samples of Jorhat, Assam (data not shown). The sequences of the amplicons were assembled to obtain contigs of approx. 1,200 bp long. The longest contig (1,221 bp, several reads were missed from the ends) was further analyzed (GenBank Acc. No. KX087160) and designated as SLCCV-[Pumpkin: Jorhat]. No amplification was found using the satellite-specific primers.

Primary sequence analysis

The selected sequence (KX087160) showed the highest identity (95%) and the maximum query coverage (98%) with SLCCV-[Pumpkin: Varanasi] segment DNA-A (GenBank Acc. No. EU573715). This result ensured that the query sequence belongs to the DNA-A of a bipartite *Begomovirus* and that the suspected virus is an isolate of SLCCV. The BLASTx analysis against nr-database of NCBI showed the presence of the full-length *C2* and *C3* genes, and parts of the *C4* and the *V1* genes. Moreover, conserved domain search against NCBI-CD database revealed the presence of four (full-length and partial) proteins: Geminivirus C3 protein, Geminivirus C2 protein, Geminivirus C4 protein and Geminivirus coat protein. The two full-length genes (C2 and C3) were considered for further analysis.

Sequence analysis revealed that the C2 (15.3 kDa) and the C3 (16.08 kDa) had the iso-electric points of 9.35 and 8.42, respectively, which indicates them to be highly alkaline in nature (Table 1). The aliphatic index of C2 was found to be 55.22 while for C3, the aliphatic index was as high as 113.82. Aliphatic index indicates the relative volume of a protein occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine); higher aliphatic index is an indicator of higher thermostability (Ikai, 1980). Therefore, our results indicated that C3 is more thermostable than C2. It had been reported that, C3 of *Geminivirus* is highly hydrophobic and contains around 62% of aliphatic amino acids (Settlage *et al*., 2005). This conclusion was found essentially true for our SLCCV-C3 also. The high thermostability of the begomoviral proteins might contribute to their survival and ability to infect crops in diverse geo-climatic conditions. It has to be reminded that *Begomovirus* family is the largest group of plant infecting viruses (Nagata *et al*., 2016; Malathi *et al*., 2017) and are highly prevalent in tropical Africa (Idris and Brown, 2002; 2005; Fauquet and Stanley, 2005; Idris *et al*., 2005; Zhou *et al*., 2008).

Table 1. Results of Protparam tool from ExPaSy proteomic server

Sl. No.	Description	C2	C3
1.	Isoelectric point	9.35	8.42
2.	Molecular mass	15.3 kDa	16.08 kDa
3.	Aliphatic index	55.22	113.82
4.	Instability index	77.00	18.67
5.	Grand average hydropathicity (GRAVY) index	-0.910	0.117

The instability index showed that C3 is more stable than C2. An instability index above 40 indicates less stability of a protein *in vitro*. The grand average of hydropathy (GRAVY) of the C2 and the C3 were -0.910 and 0.117, respectively (sum of hydropathy for all the amino acids divided by the total number of residues gives the GRAVY value of a protein). The polarity of a protein can be determined by a positive GRAVY value (http://bif.uohyd.ac.in/cpc/help_pcp.php). Thus, it was clear that the C2 is a non-polar protein whereas the C3 is a polar protein. There has been no study on the polarity of geminiviral proteins so far. However, the polar nature of the viral proteins indicates their water-solubility. This is expected from both C2 and C3, in general, because both of them are known to be involved in viral replication. However, well-designed wet-lab experiments could be useful to understand the significance of the polarity of the proteins.

Multiple sequence alignment of C2 and C3

To reveal the similarity of the SLCCV-C2 with other C2/AC2 proteins, BLASTp was performed where 18 entries (including our isolate) sharing more than 75% identity were retrieved and MSA was performed. The entries belonged to SLCCV (AGE92360.1), SLCCV-[Pumpkin: Varanasi] (ACE79036.1), SLCCV-[Pumpkin: Lucknow] (AAY63955.1), squash leaf curl Philippines virus (SLCPV)-[Taiwan] (ABM92355.1), tomato leaf curl Karnataka virus (AHA82231.1), tomato leaf curl New Delhi virus (TLCNDV)-cucumber: Thailand (BAF69032.1), TLCNDV (AJW80911.1), chilli leaf curl Bijnour virus (CLCBV; AGJ03639.1), luffa yellow mosaic virus (NP852653.1), luffa begomovirus (AAP74758.1), SLCCV-Thailand (BAF69027.1), SLCCV-Wax Gourd:Nakhon Pathom (ACB30400.1), TLCNDV-[Pakistan:Solanum] (CAF04470.1), TLCNDV-Severe [Jessore] (CAI44690.1), TLCNDV- chili pepper (ABB52026.1), TLCNDV-[Multan:Luffa] (CAL22872.1) and TLCNDV-[Potato] (AAP37409.1). Sequences were aligned by ClustalW along with SLCCV-[Pumpkin: Jorhat]. The sequences were highly conserved, especially in the N-terminal ends. It is well documented that the geminiviral C2/AC2 contains a zinc-finger motif (ZNF), a NLS and a minimal activation domain (MAD; Hartitz *et al*., 1999; Dong *et al*., 2003; Trinks *et al*., 2005) (Fig 1a). Dong *et al*. (2003) reported the pres-

ence of ZNF, $_{36}$CXCX(7)CX(6)H$_{53}$ (X: any amino acid), in the central core of the AC2 of TYLCV-China (D88773; Dong *et al*., 2003). SLCCV-C2 also showed the presence of this ZNF motif ($_{36}$CGCSYYMSINCHDHGFTH$_{53}$). This motif was found highly conserved in all the sequences considered for MSA (Fig 1a). Previously, it had been reported that, the ZNF is essential for the C2 to bind zinc and DNA, furthermore it is crucial for C2-mediated induction of necrosis and suppression of post transcriptional gene silencing (PTGS) in plants infected by TYLCV-China (Wezel *et al*., 2002, 2003). The presence of this motif in our C2 indicates that it could also have a similar function, which is expected owing to its transcriptional activation activity. On the other hand, it was found that our C2 possesses a functional, arginine-rich NLS in the N-terminal basic amino acid-rich region with four conserved arginine residues ($_{28}$RRRR$_{31}$). This motif has been reported in several viruses. In Potato virus X, the mutated $_{28}$RRRR$_{31}$ motif (replaced with DVGG) failed to transport fused GFP into nuclei. The mutation also abolished the necrosis-inducing and suppressor of PTGS function (Dong *et al*., 2003). The third reported domain of C2 is the MAD. This domain is essential for the transcriptional activation activity of C2/AC2 (Hartitz *et al*., 1999) and also involved in the silencing suppression (Trinks *et al*, 2005). This domain is small comprising of only 15 C-terminal amino acids (Hartitz *et al*., 1999). In the MSA, MAD, rich in acidic residues, was also found to be present ($_{118}$PNLDGLTPSDWDFLE$_{132}$) however, it was not equally conserved like the earlier motifs.

The TYLCV-AC2 needs functional NLS and ZNF to suppress silencing (Dong *et al*., 2003; Wezel *et al*., 2002). Trinks *et al*. (2005) proved that while mungbean yellow mosaic virus AC2 does serve as a transactivator of viral promoter and as a suppressor of silencing, these two functions are inseparable by mutations in any of the above three domains. These findings suggested that suppression of silencing by C2/AC2 might involve transcriptional activation of one or more endogenous suppressor(s). The presence of all the three domains in our C2 indicates a probable conserved function of the protein.

Similarly, 14 C3/AC3 BLASTp hits for our C3, with more than 90% identity were aligned. These entries were from TLCNDV-Multan: Luffa (CAL22871.1), TLCNDV-Cucumber: Thailand (BAF69031.1), SLCCV-[Varanasi:Pumpkin:08] (ADF28565.1), TLCNDV (AHM26185.1), SLCCV-[Thailand] (BAF69026.1), SLCCV (ABB77274.1), tomato leaf curl Palampur virus (ACS91333.1), SLCCV (BAA78101.1), SLCCV-Wax Gourd: Nakhon Pathom (ACB30401.1), SLCPV-[Taiwan] (ABM92354.1), SLCCV-*Cucurbita pepo*: Lahore (CAL18281.1), SLCCV-Pumpkin: Varanasi (ACE79037.1), SLCCV-Pumpkin: Coimbatore (AAW55551.1) and SLCPV (ACA61086.1) (Fig. 1b). The results showed that although all the entries were quite similar; the N-terminal region was relatively more conserved. In fact, there is a highly conserved polar region in the N-terminal end ($_{3}$TDSRTGEYIT$_{12}$, Motif-I) (Fig. 1b). Earlier, it was reported that the polar residues in both the N- and C-termini of the TYLCV-C3 are important for its interaction with host-encoded retinoblastoma-related protein (pRBR; Settlage *et al*., 2005). In our study, we observed only one N-terminal conserved polar motif. Meanwhile, though the C-terminal residues also were polar, they were less conserved. Further experimentation might give better insight about the functional significance of these polar regions. Settlage *et al*. (2005) found that the mutation in $_{28}$YFK$_{30}$ to AAA in the TYLCV-C3 abolished its replication enhancement activity suggesting that this motif is vital for its function or its regulation (Settlage *et al*., 2005). In MSA, this motif (Motif-II) was found conserved with a positional shift ($_{30}$YFK$_{32}$; Fig 1b). Settlage *et al*. (2005) also discovered that the $_{91}$LKYLD$_{95}$ motif is indispensable for the C3-C1 interaction in TYLCV, and that mutations in this motif can abolish this interaction. In our study, the same motif was found conserved from residue 93 to 97 (Motif-III). Settlage *et al*. (2005) also showed that while the core hydrophobic region of C3/AC3 was involved in interaction of C3/AC3 with itself, C1/AC1 or proliferating cell nuclear antigen (PCNA), the terminal polar residues at both ends were important for C3/AC3-pRBR interaction.

Phylogenetic analysis of C2 and C3

The phylogeny of C2 revealed a dichotomy with two distinct clusters (Fig. 2a); SLCCV-C2 was grouped under cluster-II. As expected, the most close evolutionarily relatives of SLCCV-C2 were AC2 of SLCCV-[Pumpkin: Varanasi] and SLCCV-[Pumpkin: Lucknow]. The cluster-I composed of C2/AC2 of tomato leaf curl viruses, CLCBV-India and luffa infecting begomoviruses, and the cluster-II composed of C2/AC2 of squash leaf curl viruses.

Phylogeny SLCCV-C3 also showed a tree with two clusters (Fig. 2b); cluster-I included SLCCV-C3 as well as C3/AC3 of squash leaf curl viruses and pumpkin yellow mosaic Malaysia virus. Again, as expected, SLCCV-C3 was found to be evolutionarily most close to SLCCV-[Pumpkin: Varanasi]. The cluster-II consisted of tomato leaf curl viruses.

Building and validation of C2 and C3 protein model

The 3D-models of SLCCV-C2 and SLCCV-C3 were generated in ITESSER server following *ab initio* method. The 3D-model of SLCCV-C2 revealed presence of three domains, NLS, ZNF and C-terminal MAD. In detailed structural study for the arrangement of the secondary structures, it was found that SLCCV-C2 comprised of five α-helices, two β-strands and few other secondary structure elements (turns/coils/hairpins). It also showed presence of the same ZNF motif ($_{36}$**C**G**C**SYYMSIN**C**HD**H**GFT**H**$_{53}$, bold residues are highly

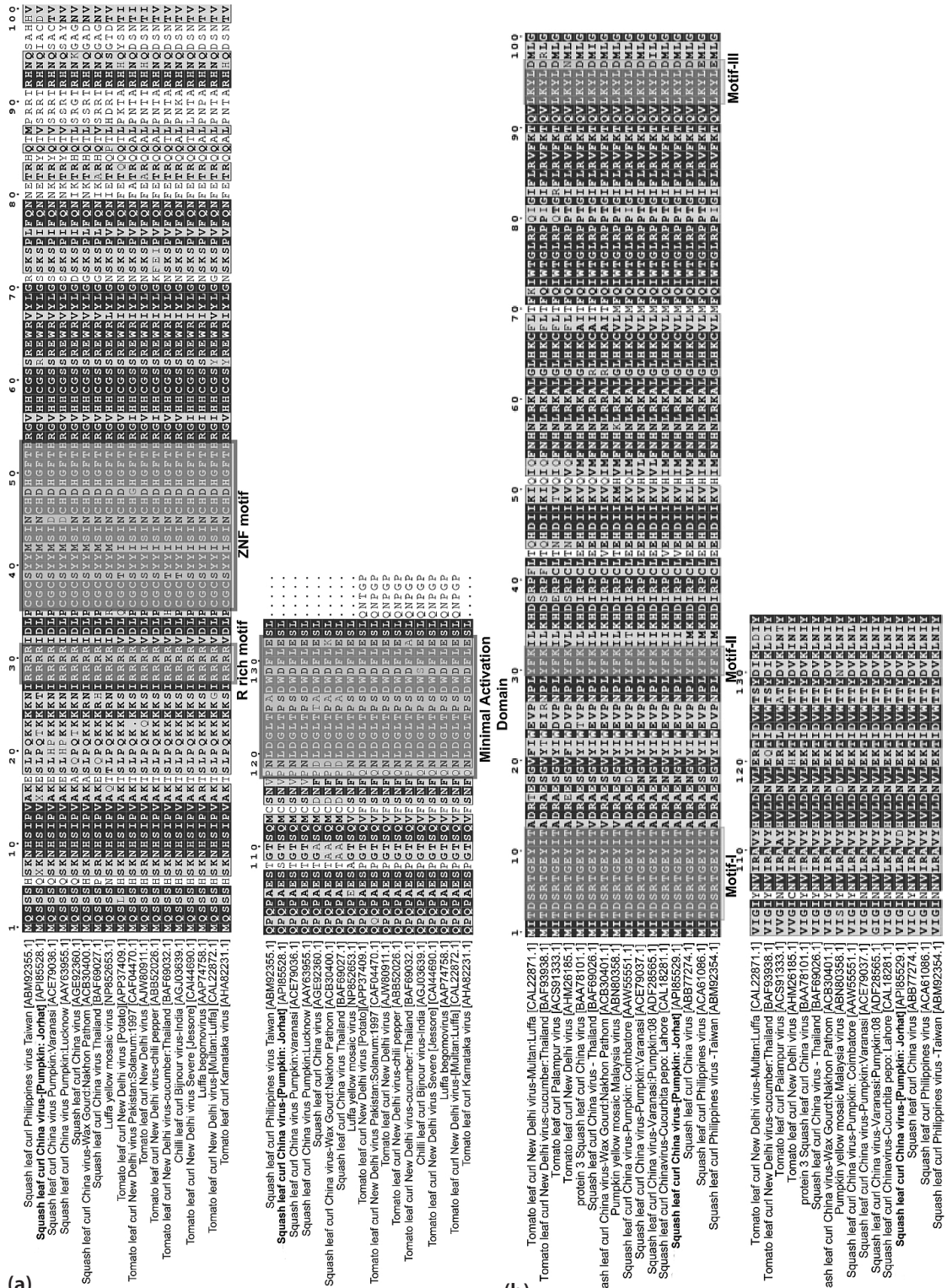

Fig. 1

Multiple sequence alignment

MSA of (**a**) C2 sequences with the seventeen closely related available sequences and (**b**) C3 sequences with fifteen closely related available sequences (MEGA v6.1 and ESPript were used, conserved regions are highlighted in dark shaded boxes labeled in white, whereas partially or conserved residues are marked in light shaded boxes labeled in black).

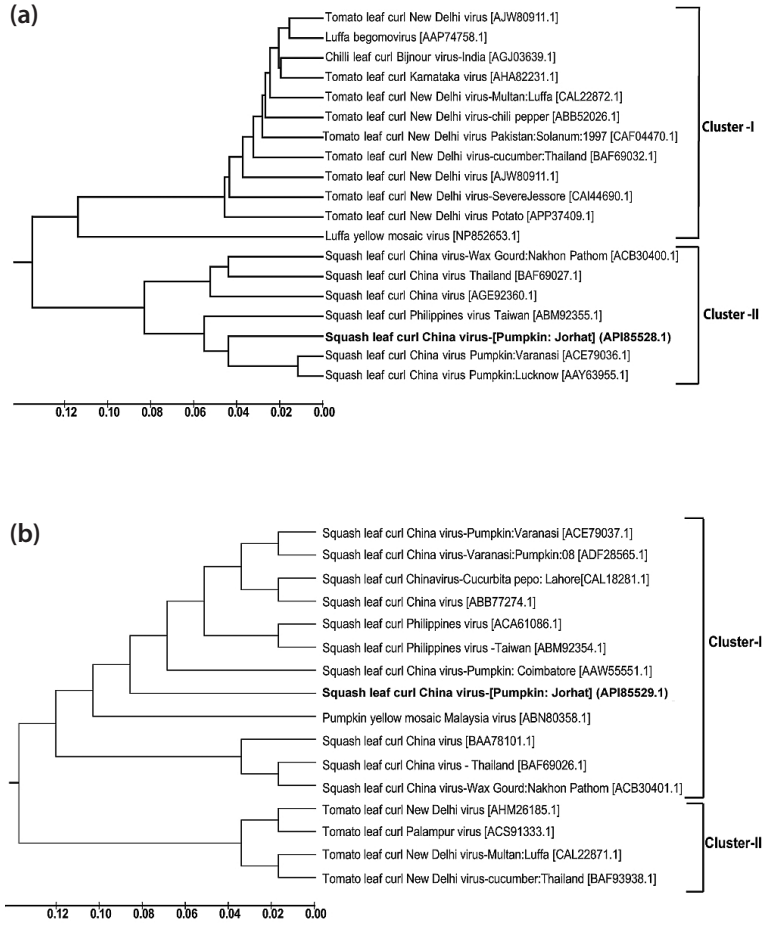

Fig. 2

Phylogenetic analysis of SLCCV C2 and C3 genes

Phylogenetic analysis of **(a)** C2 and **(b)** C3 from SLCCV-[Pumpkin: Jorhat] isolate with its closest homologs using Neighbor-Joining method.

conserved) as in MSA (Fig 3a, b). Zinc fingers generally interact with nucleic acids, and sometimes also with other proteins. Moreover, the conserved cysteine residues in the ZNF could form disulphide bridges within and between proteins (Hartitz *et al*., 1999).

SLCCV-C2 possesses a functional arginine-rich NLS within the N-terminal basic amino acid-rich region with four conserved arginine residues ($_{28}$RRRR$_{31}$). In SLCCV-C2, the sequence of the MAD is $_{118}$PNLDGLTPSDWDFLE$_{132}$. Begomoviral AC2/C2 has a modular structure consisting of an N-terminal basic region and a C-terminal acidic region (Wezel *et al*., 2002, 2003). Several studies have revealed that the basic regions are often involved in DNA-binding and/or nuclear targeting, whereas the acidic regions are characteristic of transcriptional activation domains (Johnson and McKnight, 1989; Struhl, 1989). Overall, the ZNF, the basic N-terminal and the acidic C-terminal regions of SLCCV-C2 could be instrumental in their known transcriptional activation activity.

The 3D-structure of SLCCV-C3 consisted of six α-helices and three β-strands; a conserved motif ($_{30}$YFK$_{32}$) was found to be located in the second β-strand. In TYLCV-C3, this motif plays a pivotal role in replication enhancement (Settlage *et al*., 2005). Another motif, $_{91}$LKYLD$_{95}$, was shown to be crucial for C3-C1 interaction in TYLCV (Settlage *et al*., 2005). This motif was found in the C-terminal part of SLCCV-C3 (Fig. 4a,b) in the second α-helix. The hydrophobic residues in the middle of C3 were implicated in its oligomerization, and its interaction with C1 and PCNA (Settlage *et al*., 2005). These findings indicate the structural-functional conservation of C3/AC3 across the begomoviruses.

Ramachandran plot analysis showed that 62.4% residues of the C2 and 61.8% residues of the C3 reside in the most favoured regions (Φ (phi) and Ψ (psi) angles), and 33.4% residues of the C2 and 36.6% residues of the C3 are located within the additionally and generously allowed regions. This accounted for a total of 95.8% residues of the C2 and 98.4% residues the C3 signifying the accuracy of models (Fig. 5).

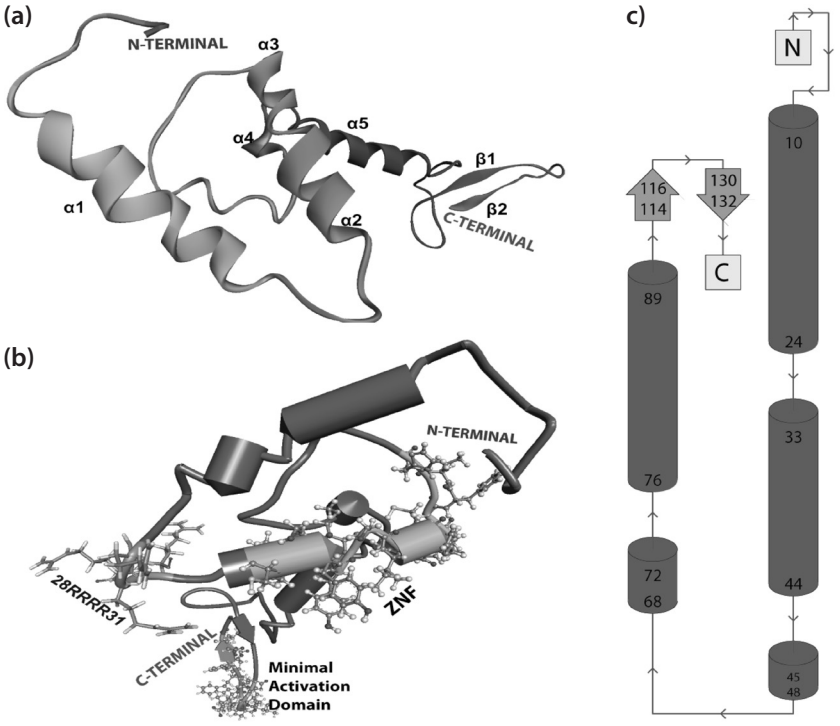

Fig. 3

3D-models of SLCCV C2 protein

(a) 3D-model of SLCCV-C2 and its topology (solid ribbons represent the SLCCV-C2 model colored by its secondary structure elements, secondary structure elements such as α-helices, β-strands and the N- and C-termini are labeled). **(b)** Schematic diagram of SLCCV-C2 showing the conserved motifs and their positions in the 3D structure. **(c)** Topology of SLCCV-C2 (the images were constructed with Discovery Studio and Profunc Web server).

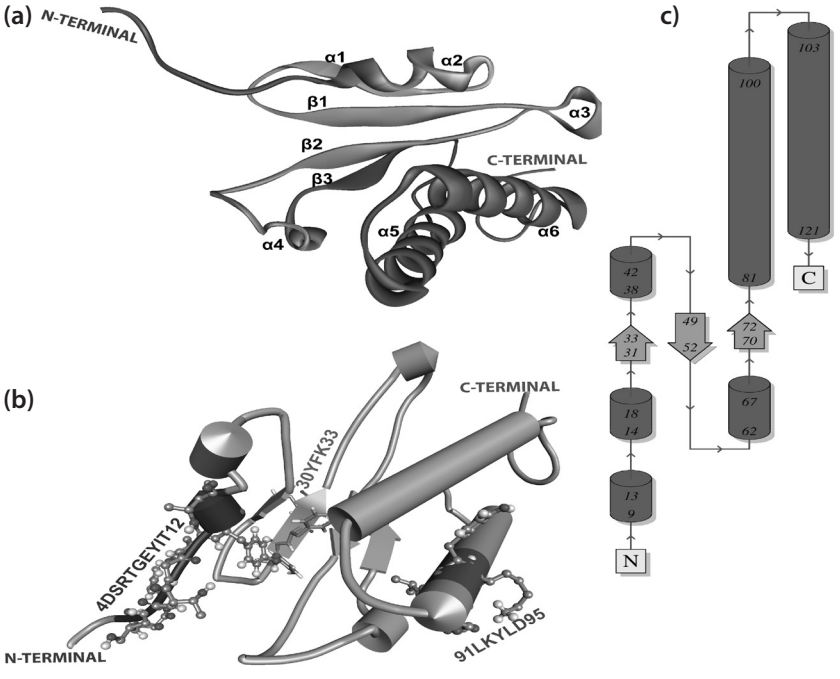

Fig. 4

3D-models of SLCCV C3 protein

(a) 3D-model of SLCCV-C3 and its topology (solid ribbon represents the SLCCV-C3 model colored by its secondary structure elements, secondary structure elements such as α-helices, β-strands and the N- and C-termini are labeled). **(b)** Schematic diagram of SLCCV-C3 showing the conserved motifs and their positions in the 3D structure. **(c)** The topology of SLCCV-C3 (the images were constructed with Discovery Studio and Profunc Web server).

All the bond-distances and the bond-angles lied within the allowed standard ranges further indicating that the models are reasonably good in geometry and stereochemistry. The Φ and Ψ distribution of Ramachandran plot of the non-glycine and non-proline residues, and the templates are summarized in Table 2. The packing quality of each residue of the models was assessed by Verify3D Program, where the compatibility of the residues with their environment is scored; a score above 0.2 is considered reliable. The scores of the models of SLCCV-C2 and -C3 were above 0.2 corresponding to

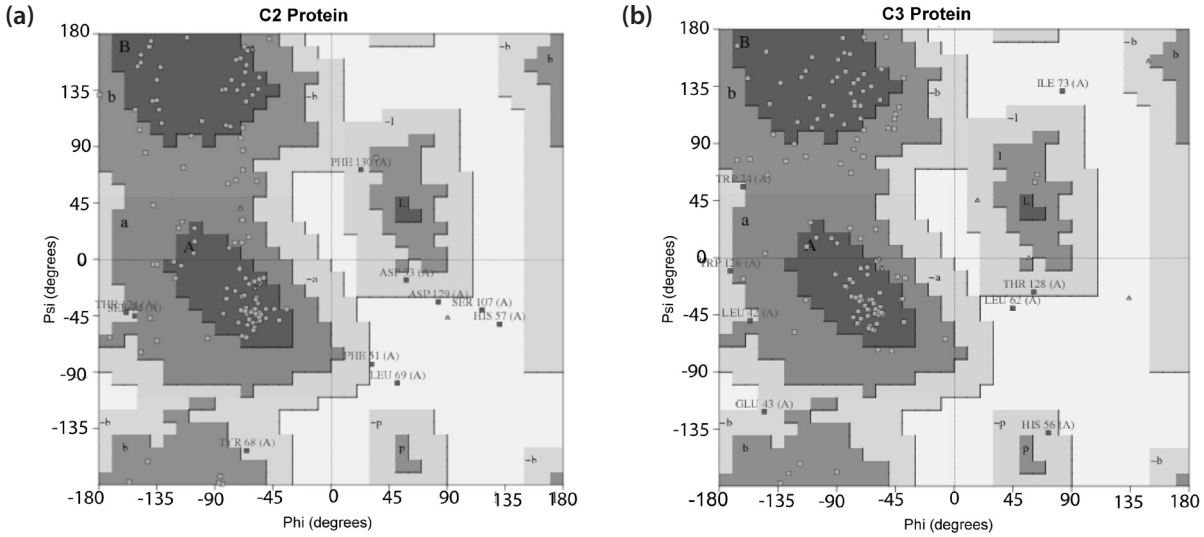

Fig. 5

Ramachandran plot of SLCCV-C2 and C3 protein

Ramachandran plot of **(a)** SLCCV-C2 **(b)** SLCCV-C3. Most favoured regions are shown as [A, B, L], additional allowed regions as [a, b, l, p], generously allowed regions as [~a, ~b, ~l, ~p] and disallowed regions are represented as [XX]. (Procheck embedded in Profunc tool was used for model validation).

Table 2. Ramachandran plot statistics of SLCCV-C2 and -C3 models

Sl. No.	Ramachandran plot statistics	C2	Residues	C3	Residues
1.	Residues in most favoured regions	62.4%	73	61.8%	76
2.	Residues in additionally allowed regions	29.1%	34	31.7%	39
3.	Residues in generously allowed regions	4.3%	5	4.9%	6
4.	Residues in disallowed regions	4.2%	5	1.6%	2
5.	Number of non-glycine and non-proline	100%	117	100%	123
6.	End residues excluding Glycine and non-Proline	-	2	-	2
7.	Number of glycine residues	-	7	-	6
8.	Number of proline residues	-	8	-	5

acceptable side-chain environments. The corroboration of the 3D modelling by the Ramachandran plot puts forward an indication towards the significant correctness of the models. Moreover, the presence of the conserved motifs in the models as well as in the primary structures of C2 and C3 correlated satisfactorily. Most importantly, the structural details elucidated in this study could possibly be applied to other begomoviruses.

The present study on SLCCV-[Pumpkin: Jorhat], in addition to finding its evolutionary relationships, imparts light on its two important proteins and gives a fundamental insight into their functionality. Although, the models were built *ab initio*, the conservation of the functional motifs and domains in the models implies that they could be correct enough to be relied on, at least for the core. The Ramachandran plot analysis further strengthens this conclusion. However, real wet-lab validation would be necessary in future for confident reliance on the models. To our knowledge, this is the first attempt to construct *ab initio* 3D models of two begomoviral proteins. Such studies could possibly be extended to other viral proteins as well. Meanwhile, the accuracy of the constructed models with respect to the role of the functional motifs could be an interesting first step towards understanding the infection process of geminiviruses.

Acknowledgment. We would like to thank the Distributed Information Centre, Dept. of Agricultural Biotechnology (ABT), AAU, Jorhat for providing computational facility. The research in the Dept. of ABT, AAU is supported by DBT, Govt. of India; DST, Govt. of India; ICAR, Govt. of India and DBT-AAU Centre. GMB acknowledges UGC, Govt. of India for providing Rajiv Gandhi National fellowship. The funding agencies are not technically involved with the research.

References

Borah BK, Dasgupta I (2012): Begomovirus research in India: A critical appraisal and the way ahead. J. Biosci. 37, 791–806. https://doi.org/10.1007/s12038-012-9238-y

Briddon RW, Bull SE, Mansoor S, Amin I, Markham PG (2002): Universal primers for the PCR-mediated amplification of DNA-ß: a molecule associated with monopartite begomoviruses. Mol. Biotechnol. 20, 315–318. https://doi.org/10.1385/MB:20:3:315

Briddon RW, Markham PG (1994): Universal primers for the PCR amplification of dicot-infecting geminiviruses. Mol. Biotechnol. 2, 202. https://doi.org/10.1007/BF02921559

Capoor SP, Ahmad RU (1975): Yellow vein mosaic disease of field pumpkin and its relationship with the vector, Bemisia tabaci. Indian Phytopath. 28, 241–246.

Dong X, Wezel RV, Stanley J, Hong Y (2003): Functional characterization of the nuclear localization signal for a suppressor of post transcriptional gene silencing. J. Virol. 77, 7026–7033. https://doi.org/10.1128/JVI.77.12.7026-7033.2003

Doyle JJ, Doyle JL (1990): Isolation of plant DNA from fresh tissue. Focus 12, 13–15. https://doi.org/10.2307/2419362

Elmer JS, Brand L, Sunter G, Gardiner WE, Bisaro DM, Rogers SG (1988): Genetic analysis of the tomato golden mosaic virus. II. The product of the AL1 coding sequence is required for replication. Nucleic Acids Res. 16, 7043–7060. https://doi.org/10.1093/nar/16.14.7043

Fauquet CM, Stanley J (2005): Revising the way we conceive and name viruses below the species level: a review of geminivirus taxonomy calls for new standardized isolate descriptors. Arch. Virol. 150, 2151–2179. https://doi.org/10.1007/s00705-005-0583-0

Gladfelter HJ, Eagle PA, Fontes EP, Batts L, Hanley-Bowdoin L (1997): Two domains of the AL1 protein mediate geminivirus origin recognition. Virology 239, 186–197. https://doi.org/10.1006/viro.1997.8869

Hartitz MD, Sunter G, Bisaro DM (1999): The tomato golden mosaic virus transactivator (TrAP) is a single-stranded DNA and zinc-binding phosphoprotein with an acidic activation domain. Virology 263, 1–14. https://doi.org/10.1006/viro.1999.9925

Hormuzdi SG, Bisaro DM (1995): Genetic analysis of beet curly top virus: examination of the roles of L2 and L3 genes in viral pathogenesis. Virology 206, 1044–1054. https://doi.org/10.1006/viro.1995.1027

ICTV. Virus Taxonomy: 2017 Release. http://ictvonline.org/virusTaxonomy.asp

Idris AM, Briddon RW, Bull SE, Brown JK (2005): Cotton leaf curl Gezira virus-satellite DNAs represent a divergent, geographically isolated Nile Basin lineage: predictive identification of a satDNA REP-binding motif. Virus Res. 109, 19–32. https://doi.org/10.1016/j.virusres.2004.10.002

Idris AM, Brown JK (2002): Molecular analysis of cotton leaf curl virus-Sudan reveals an evolutionary history of recombination. Virus Genes 24, 249–256. https://doi.org/10.1023/A:1015380600089

Idris AM, Brown JK (2005): Evidence for inter specific recombination for three monopartite begomoviral genomes associated with the tomato leaf curl disease from central Sudan. Arch. Virol. 150, 1003–1012. https://doi.org/10.1007/s00705-004-0484-7

Ikai AJ (1980): Thermostability and aliphatic index of globular proteins. J. Biochem. 88, 1895–1898.

Jayashree K, Pun KB, Doraiswamy S (1999): Virus-vector relationship of yellow vein mosaic virus and whitefly (Bemisia tabaci) in pumpkin. Indian Phytopath. 52, 10–13.

Johnson PF, McKnight SL (1989): Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. 58, 799–839. https://doi.org/10.1146/annurev.bi.58.070189.004055

Jupin I, De Kouchkovsky F, Jouanneau F, Gronenborn B (1994): Movement of tomato yellow leaf curl geminivirus (TYLCV): involvement of the protein encoded by ORF C4. Virology 204, 82–90. https://doi.org/10.1006/viro.1994.1512

Laufs J, Traut W, Heyraud F, Matzeit V, Rogers SG, Schell J, Gronenborn, B (1995): In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. Proc. Natl. Acad. Sci. USA 92, 3879–3883. https://doi.org/10.1073/pnas.92.9.3879

Malathi VG (2017): Begomovirus: An introduction. In Saxena S, Tiwari AK (Eds): Begomoviruses: occurrence and management in Asia and Africa. Springer Nature, Singapore, pp. 3–9. https://doi.org/10.1007/978-981-10-5984-1_1

Maruthi MN, Colvin J, Briddon RW, Bull SE, Muniyappa V (2003): Pumpkin yellow vein mosaic virus, a novel begomovirus infecting cucurbits. J. Plant. Pathol. 85, 64–65.

Muniyappa V, Maruthi MN, Babitha CR, Colvin J, Briddon RW, Rangaswamy KT (2003): Characterisation of pumpkin yellow vein mosaic virus from India. Ann. Appl. Biol. 142, 323–331. https://doi.org/10.1111/j.1744-7348.2003.tb00257.x

Nagata AEI, Lima MF, Gilbertson RL (2016): A review of geminivirus diseases in vegetables and other crops in Brazil: current status and approaches for management. Hortic. Bras. 34, 8–18. https://doi.org/10.1590/S0102-053620160000100002

Rigden JE, Krake LR, Rezaian MA, Dry IB (1994): ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. Virology 204, 847–850. https://doi.org/10.1006/viro.1994.1606

Rojas MR, Gilbertson RJ, Russell DR, Maxwell DP (1993): Use of degenerate primers in the polymerase chain-reaction to detect whitefly-transmitted geminiviruses. Plant Dis. 77, 340–347. https://doi.org/10.1094/PD-77-0340

Settlage SB, See RG, Hanley-Bowdoin L (2005): Geminivirus C3 Protein: Replication enhancement and protein interactions. J. Virol. 79, 9885. https://doi.org/10.1128/JVI.79.15.9885-9895.2005

Shen W, Hanley-Bowdoin L (2006): Geminivirus infection upregulates the expression of two Arabidopsis protein kinases related to yeast SNF1- and mammalian AMPK-activating kinases. Plant Physiol. 142, 1642–1655. https://doi.org/10.1104/pp.106.088476

Struhl K (1989): Molecular mechanisms of transcriptional activation in yeast. Annu. Rev. Biochem. 58, 1051–1077. https://doi.org/10.1146/annurev.bi.58.070189.005155

Teng K, Chen H, Lai J, Zhang Z, Fang Y, Xia R, Zhou X, Guo H, Xie Q (2010): Involvement of C4 protein of beet severe curly top virus (family Geminiviridae) in virus movement. PLoS One 5, e11280. https://doi.org/10.1371/journal.pone.0011280

Trinks D, Rajeswaran R, Shivaprasad PV, Akbergenov R, Oakeley OJ, Veluthambi K, Hohn, T, Pooggin MM (2005): Suppression of RNA silencing by a geminivirus nuclear protein AC2 correlates with transactivation of host genes. J. Virol. 79, 2517–2527. https://doi.org/10.1128/JVI.79.4.2517-2527.2005

Varma PM (1955): Ability of the whitefly to carry more than one virus simultaneously. Curr. Sci. 24, 317–318.

Vasudeva RS, Lal TB (1943): A mosaic disease of bottle gourd. Indian J. Agric. Sci. 13, 182–191.

Wezel RV, Dong X, Liu H, Tien P, Stanley J, Hong Y (2002): Mutation of three cysteine residues in Tomato yellow leaf curl virus-China C2 protein causes dysfunction in pathogenesis and posttranscriptional gene silencing suppression. Mol. Plant-Microbe Interact. 15, 203–208. https://doi.org/10.1094/MPMI.2002.15.3.203

Wezel RV, Liu H, Wu Z, Stanley J, Hong Y (2003): Contribution of the zinc finger to zinc and DNA binding by a suppressor of posttranscriptional gene silencing. J. Virol. 77, 696–700. https://doi.org/10.1128/JVI.77.1.696-700.2003

Zhou YC, Noussourou M, Kon T, Rojas MR, Jiang H, Chen LF, Gamby K, Foster R, Gilbertson RL (2008): Evidence of local evolution of tomato-infecting begomovirus species in West Africa: characterization of tomato leaf curl Mali virus and tomato yellow leaf crumple virus from Mali. Arch. Virol. 153, 693–706. https://doi.org/10.1007/s00705-008-0042-9