



Journal of Biological Sciences

ISSN 1727-3048

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

***In silico* Analysis of Genetic Diversity of Begomovirus using Homology Modelling**

R. Prajapat, R.K. Gaur, R. Raizada and V.K. Gupta
Department of Biotechnology, MITS University, Lakshmangarh, Rajasthan, India

Abstract: Structural bioinformatics is concerned with computational approaches to predict and analyse the spatial structure of proteins and nucleic acids. The 3D models can be obtained most easily if the 3D structure of a homologous protein is known (homology modeling, comparative modeling). Homology modeling or comparative modeling is based on the fact that a structure of a protein can be reliably modeled when its sequence is sufficiently similar to a protein sequence with known 3D structure. The Geminivirus taxonomic group of plant viruses is characterized by geminate particles and genomes consisting of single-stranded circular DNA (sscDNA) molecules of about 2-5 to 2.8 kb in size. Here, we analyzed the genetic diversity of begomovirus from different crops and weeds of different region on the basis of homology modeling. This high degree of genetic diversity suggests a recent transfer of indigenous begomovirus from wild hosts into economic important crops.

Key words: Homology modeling, geminivirus, genetic diversity, geminate particles

INTRODUCTION

Today, experimentally we can only predict the structure-based drug design, analysis of protein function, interaction, antigenic behaviour and rational design of protein with increased stability or novel function. Unlike experimental structure, Protein modeling are use to predict a structure from its sequence with an accuracy that allow users to use safely generated *in silico* protein models in all the contexts. In addition, protein modeling is the only way to obtain structural information if experimental techniques fail. Many proteins are too large for NMR analysis and cannot be crystallized for X-ray diffraction.

The UCLA-DOE Structure Evaluation server is a tool designed to help in the refinement of crystallographic structures. It will provide you with a visual analysis of the quality of a putative crystal structure for a protein. Verify 3D expects this crystal structure to be submitted in PDB format. Please note that Verify 3D works best on proteins with at least 100 residues. To submit a crystal structure for analysis, simply select it with the file dialog which is activated by clicking on the Browse button below, then click the Send File button.

Verify3D analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D). Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar, etc). A collection of good structures is used as a reference to obtain a score for each of the 20 amino acids in this structural class. The scores of a sliding 21-residue

window (from -10 to +10) are added and plotted for individual residues.

Genomic organization of begomovirus: The genus *Begomovirus* contains more than 200 species (Fauquet *et al.*, 2008) and belongs to the taxonomic family Geminiviridae. They are plant viruses that as a group have a very wide host range, infecting dicotyledonous plants. Worldwide they are responsible for a large amount of economic damage to many important crops such as tomatoes, beans, squash, cassava and cotton.

Virus particles are non-enveloped. The nucleocapsid is 38 nm long and 15-22 nm in diameter. While particles have basic isocahedral symmetry, they consist of two incomplete isocahedra-missing one vertex-joined together. There are 22 capsomeres per nucleocapsid.

The geminivirus taxonomic group of plant viruses is characterized by geminate particles and genomes consisting of single-stranded circular DNA (sscDNA) molecules of about 2-5 to 2.8 kb in size (Lazarowitz, 1987). There are currently three subgroups recognized by the International Committee on Taxonomy of Viruses (ICTV): subgroup I viruses have single component, ssDNA genomes of about 2-7 kb, mainly infect monocotyledonous plants and are obligate leafhopper-transmitted; Sub group II single genomic component closely resembles DNA A of III, transmitted by leafhoppers or tree hoppers and infect dicotyledonous plant subgroup III viruses have two component genomes, infect dicotyledonous plants and are transmitted

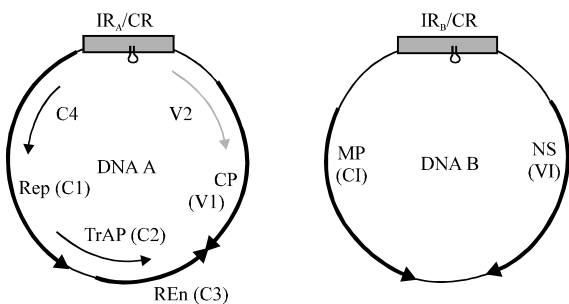


Fig. 1: Genomic organization of a typical begomovirus

mechanically and by the whitefly *Bemisia tabaci* (ICTV Group Descriptions, 1989; G.P. Martelli, personal communication). The two subgroups are distantly evolutionarily related (Howarth and Vandemark, 1989).

Many begomoviruses have a bipartite genome, this means that the genome is segmented into two segments (referred to as DNA A and DNA B) that are packaged into separate particles. Both segments are generally required for successful symptomatic infection in a host cell, but DNA B is dependent for its replication upon DNA A, which can in some begomoviruses apparently cause normal infections on its own. Depending on the particular begomovirus, six different proteins are involved in mediating virus transport through host plants: CP, AV2, AC1, AC2, AC3 and AC4 (Fig. 1) (Rojas *et al.*, 2005). Unlike monopartite begomoviruses, the CP of bipartite begomoviruses is not required for cell-to-cell or long-distance movement. For bipartite begomoviruses, cell-to-cell movement function is carried out by two proteins encoded by the DNA B component, NSP and MP. In the case of monopartite begomoviruses, two proteins, V1 and C4, have been implicated in cell-to-cell movement function (Jupin *et al.*, 1994; Rojas *et al.*, 2001). The opposing transcription units of begomovirus DNA-A and -B molecules are separated by an Intergenic Region (IR) that generally shares a highly conserved region of approximately 200 nt, named the Common Region (CR) (Lazarowitz, 1992). The CR contains an origin of replication (*ori*) that includes a stem-loop structure containing an invariant nonanucleotide (TAATATTAC) sequence whose T7-A8 site is required for cleaving and joining viral DNA during replication and conserved iterated sequences (*iterons*) required for specific recognition and binding by Rep during replication (Arguello-Astorga *et al.*, 1994; Fontes *et al.*, 1994a, b). Based on phylogenetic studies and genome arrangement, begomoviruses have been divided broadly into two groups: the Old World (OW) viruses (eastern hemisphere, Europe, Africa, Asia) and the New World (NW) viruses (western hemisphere, the Americas) (Padidam *et al.*, 1999;

Paximadis *et al.*, 1999; Rybicki, 1994). Begomovirus genomes have a number of characteristics that distinguish Old World and New World viruses. All indigenous New World begomoviruses are bipartite, whereas both bipartite and monopartite begomoviruses are present in the Old World. In addition, DNA-A of bipartite begomoviruses from the New World lacks an AV2 ORF (Rybicki, 1994; Stanley and Latham, 1992).

Besides this few geminiviruses contains a small circular single-stranded DNA termed satellite DNA or DNA- β having single open reading frame (ORF). DNA- β are biologically functional, being required for amplification of DNA A to high levels and for the efficient systemic spread of infection. Since they were first reported in 2000, over one hundred full-length DNA- β s have been cloned and sequenced (Mansoor *et al.*, 2006). They consist of approximately 1350 nt and depend on their helper begomoviruses for replication and encapsidation. The process through which DNA- β regulates pathogenesis is unknown, but it might affect the replication of its helper virus by either facilitating its spread in host plants, or by suppressing host gene silencing (Saunders *et al.*, 2000). The DNA b-encoded protein, bC1, is the determinant of both pathogenicity and suppression of gene silencing (Bridson and Stanley, 2006).

Genetic diversity of ORF region: AV1/AV2: DNA A of the begomovirus contains AV1 ad AV2 at the virus strand. The localization of the product of this ORF in secondary plasmodesmata with the onset of viral lesions is consistent with its role in the movement of monopartite geminivirus (Dickinson *et al.*, 1996). All begomovirus from the new world are bipartite and lack a virion sense Open Reading Frame (ORF) designated AV2. This involved in the movement of monopartite viruses but its function is unknown in bipartite. A recent report has similarly shown that AV2 of Indian Cassava Mosaic Virus (ICMV), another bipartite begomovirus from the Indian subcontinent, is involved in virus spread to neighbouring cells (Shunter and Bisaro, 1992), whereas the V2 (a homolog of AV2) of a monopartite begomovirus is involved in overcoming host defences mediated by post-transcriptional gene silencing as well as in movement (Sunter *et al.*, 2001; Vance and Vaucheret, 2001). The comparison of 153 residues of two begomovirus namely Tomato yellow leaf curl Thailand virus and Tomato leaf curl Karnataka virus shows position diversity of aromatic amino acids as in Fig. 2. The Chi1-Chi2 plot also shows the diversity of favourable and allowed region of the structure (Fig. 3).

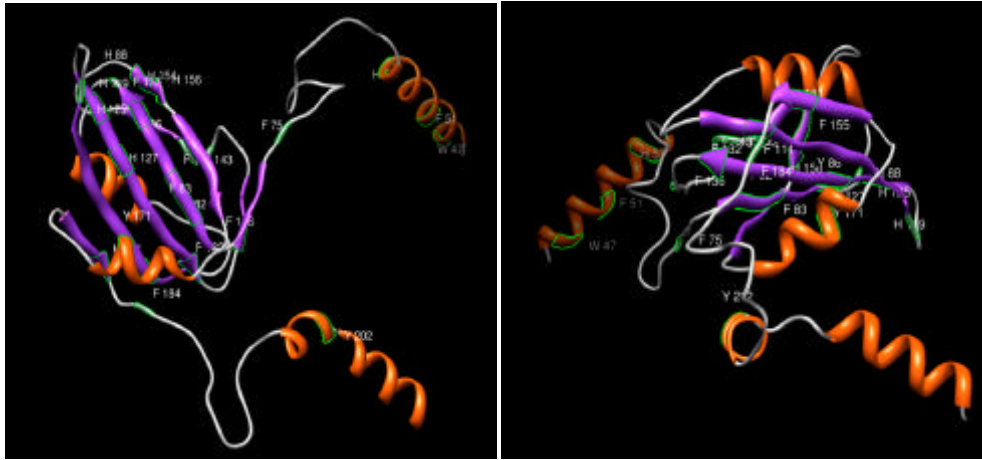


Fig. 2: 3D structure of AV1 coat protein of the two begomovirus showing the diversity of aromatic residue position

AC1: The Rep is present in nuclei of infected plant cells (Nagar *et al.*, 1995), where it plays a key role in geminivirus DNA replication and transcription (Laufs *et al.*, 1995). Rep confers origin of replication and initiates plus-strands DNA replication (Chatterji *et al.*, 2001). Several biochemical activities have been demonstrated for the Rep. it specifically cleaves and ligates viral plus strands DNA within the loop of the hairpin motifs. The protein also possesses DNA nicking activity, site specific topoisomerase and helicase activity (Heyraud-Nitschke *et al.*, 1995). Amino acids 1 to 130 are sufficient for DNA binding, amino acids 134 to 181 are for oligomerization activity. It is suppose that C-terminus is involved in ATP hydrolysis (Desbiez *et al.*, 1995; Fig. 4).

AC2/TrAP: The AC2 gene product or transcriptional activator protein (TrAP) is a trans-acting factor needed for the expression of both AV1 (coat protein) and BV1 genes (Sunter and Bisaro, 1992). The TrAP required for initiating transcription of the virion-sense genes and also involved in suppression of post-transcriptional gene silencing-mediated host defences (Voinnet *et al.*, 1999; Vanitharami *et al.*, 2005). Southern blot analyses of the total DNAs isolated from inoculated protoplasts revealed that EACMCV AC2 expressed in trans under a 35S promoter enhanced ACMV-[CM] DNA accumulation by ~8-fold compared to that of protoplasts transfected with ACMV-[CM] (Vanitharami *et al.*, 2004). ACMV-[CM] AC2 in trans did not alter EACMCV DNA accumulation, but it could act as a mild suppressor of PTGS, as shown for AC2 of ACMV-[KE] (Hamilton *et al.*, 2002; Voinnet *et al.*, 1999). The AC2 genes of ACMV-[CM] and EACMCV show 56% homology and contain conserved zinc finger, transactivator and DNA-binding domains; however, these

genes act differently. In monopartite geminiviruses, it has been demonstrated recently that the C2 protein of TYLCCNV requires the DNA-binding domain, the zinc finger domain and the nuclear localization signal for anti-PTGS activity (Van Wezel *et al.*, 2002, 2003). C2 is a positional homologue of AC2 in bipartite geminiviruses, but it varies functionally in that it does not facilitate the transcriptional activation of coat protein, as in the case of AC2 of bipartite geminiviruses. Therefore, we propose that the different functions of the AC2 proteins that we noticed could be due to differences at the level of point mutations, as indicated by sequence comparisons (data not shown). In addition, AC2 of Tomato golden mosaic virus and C2 of Beet curly top virus have been shown to manipulate host metabolism by interfering with SNF1 kinase and adenosine kinase to provide enhanced susceptibility to virus infection (Wang *et al.*, 2003; Silhavy *et al.*, 2002); however, their roles in relation to PTGS have not been established. All these data support the fact that these tiny proteins are multifunctional in nature.

AC3/Ren: The AC3 is also know as replication enhancer (ren) having the length of 134 nt. Base composition of AC3 shows that it has lower average GC content in comparison to other GC content found among other. It has been observed that ren protein is located in nuclei of infected plant cells at levels similar to Rep (Nager *et al.*, 1995) suggesting that it might act with Rep during initiation of viral DNA replication. Experimental observation suggests that ren protein might increase the affinity of rep for the origin (Mohr *et al.*, 1990). Rep.Ren protein complex formation might also enhance DNA cleavage, ligation or putative helicase (Koonin *et al.*, 1991) activities of Rep.

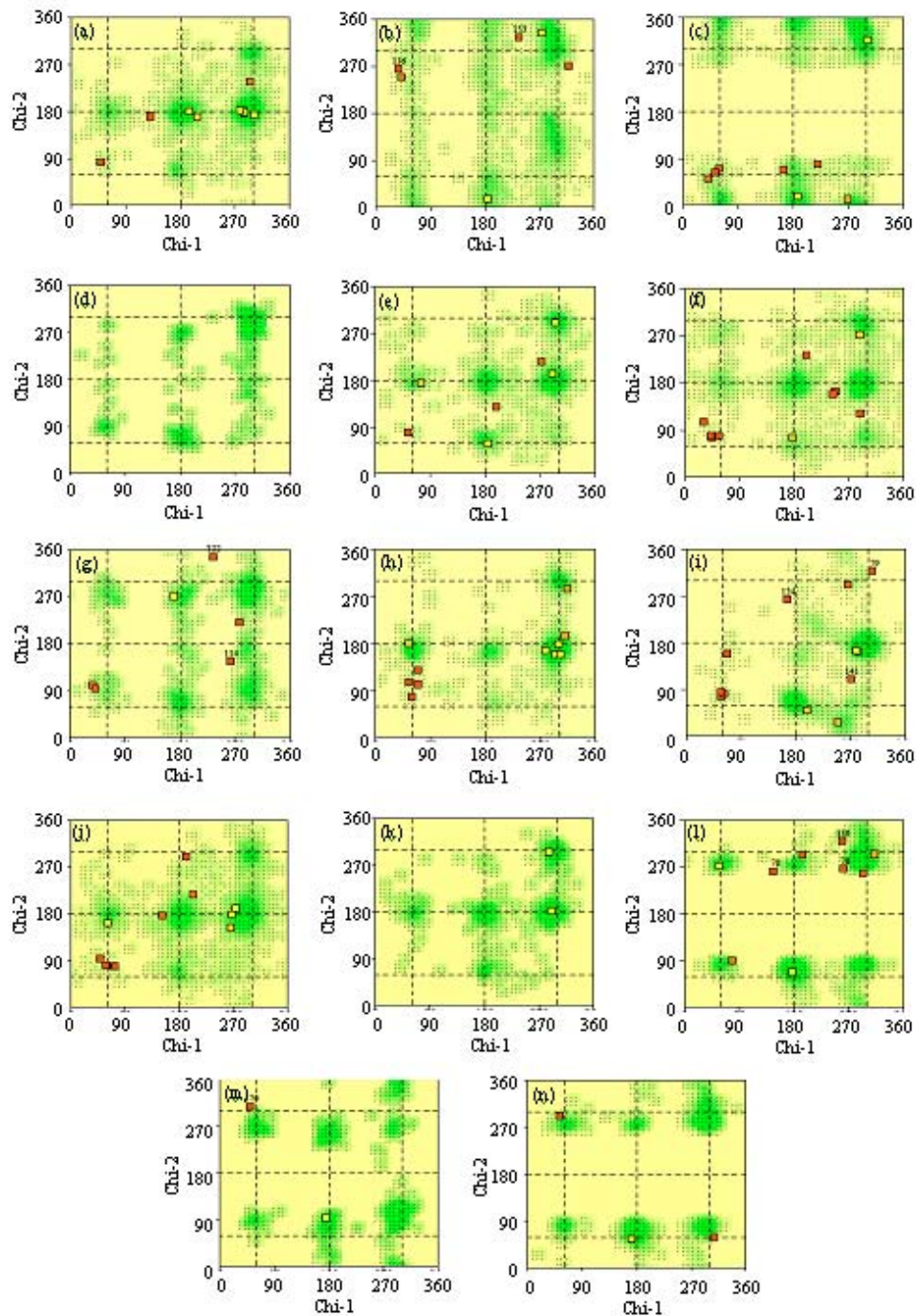


Fig. 3: Chi1-Chi2 plot for AV1 coat protein of a begomovirus. Numbers of residues are shown in brackets. Those in unfavourable conformations (score<-3.00) are labeled. Shading shows favourable conformations as obtained from an analysis of 163 structures are resolution 2.0Å or better. (a) Arg (10), (b) Asn (6), (c) Asp (8), (d) Cys (-), (e) Gln (7), (f) Glu (10), (g) His (6), (h) Ile, (i) Leu (12), (j) Lys (11), (k) Met (2), (l)Phe (10), (m) Trp (2) and (n) Tyr (3)

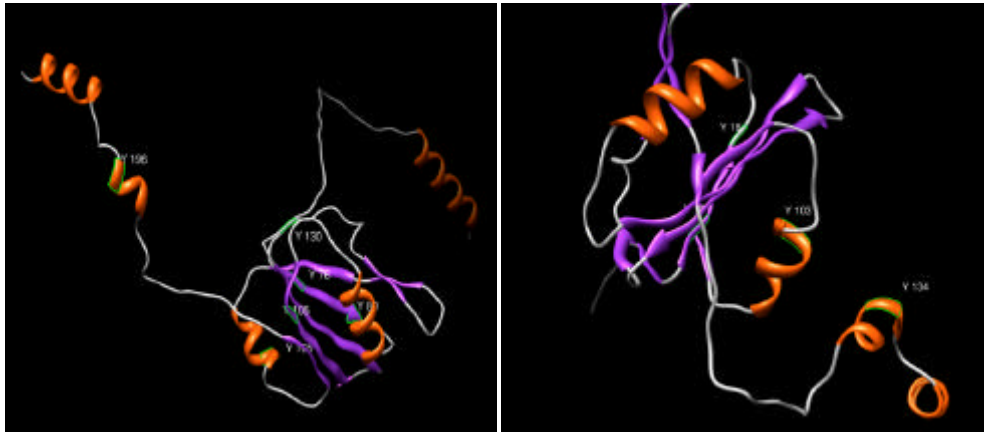


Fig. 4: 3D structure of AC1 proteins of the two different begomovirus showing the diversity of the position of tyrosine

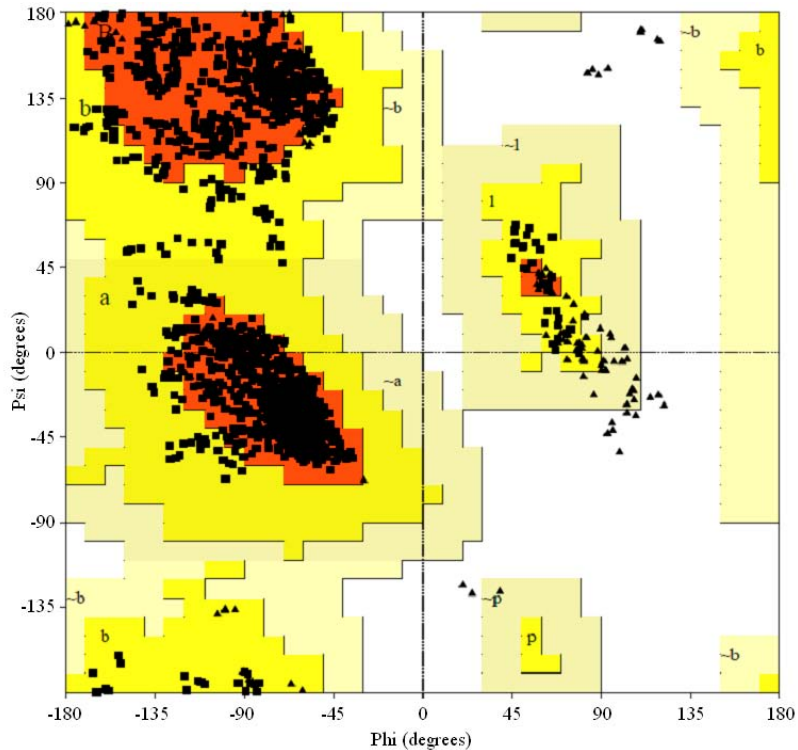


Fig. 5: Ramachandran plot for AC4 protein of a begomovirus based on an analysis of 118 structures of resolution of at least 2.0 angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions

AC4 protein: ACMV-[CM] AC4 expressed in trans from the 35S promoter increased EACMCV DNA accumulation by ~8-fold compared to control transfection with EACMCV alone (Vanitharam *et al.*, 2004). Mutation analysis of AC4 has shown that this protein is involved in symptoms development. Stanley and Latham (1992) introduced a stop codon at two different locations in the C4 gene with out affecting the amino acid sequence of Rep. Agrobacterium inoculation of a mutant of TLCV

(Australia) produced drastically reduced symptoms, although the level of viral DNA was similar to the wild type virus and suggests that C4 is not required by TLCV to replicate or to spread through the host plant but is involved in symptom development (Rigden *et al.*, 1994).

The Ramachandran plot for AC4 of Abutilon mosaic virus (isolate West India) (AbMV) explains that out of the 1988 residues 89.4% residues in most favoured region and 10.6% residue in additional allowed regions (Fig. 5).

CONCLUSION

Homology modeling is an effective way to obtain useful information about the proteins of interest. Structure of unknown proteins can be identified on the basis of amino acid sequence pattern matching of both known and unknown proteins. By selecting five or ten best homologous proteins showing maximum sequence similarity with the unknown protein for homology search we may predict that their structure could be similar.

Breeders working towards the incorporation of begomovirus resistance genes must take into account the high degree of genetic diversity begomoviruses in the world. A given resistance gene might be extremely effective against a particular begomovirus species and totally ineffective against distinct, unrelated species. Also, the rate of evolution of begomoviruses seems to be inordinately fast (Padidam *et al.*, 1999), leading to the quick emergence of new strains or species that might overcome resistance genes. The durability of a resistance based on a single gene is therefore questionable. The genetic diversity of begomoviruses present in additional regions should be determined and the prevalence of distinct species or strains identified, in order to orient breeding programs towards resistance to the species prevalent in each region.

REFERENCES

- Arguello-Astorga, G.R., R.G. Guevara-Gonzalez, L.R. Herrera-Estrella and R.F. Rivera-Bustamante, 1994. Geminivirus replication origins have a group-specific organization of iterative elements: A model for replication. *Virology*, 203: 90-100.
- Briddon, R.W. and J. Stanley, 2006. Subviral agents associated with plant single-stranded DNA viruses. *Virology*, 344: 198-210.
- Chatterji, A., R.N. Beachy and C.M. Fauquet, 2001. Expression of the oligomerization domain of the replication-associated protein (Rep) of tomato leaf curl New Delhi virus interferes with DNA accumulation of heterologous geminiviruses. *J. Biol. Chem.*, 276: 25631-25638.
- Desbiez, C., C. David, A. Mettouchi, J. Laufs and B. Gronenborn, 1995. Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *Proc. Natl. Acad. Sci.*, 92: 5640-5644.
- Dickinson, V.J., J. Halder and C.J. Woolston, 1996. The product of maize streak virus ORF VI is associated with secondary *Plasmodesmata* and is first detected with the onset of viral lesions. *Virology*, 202: 51-59.
- Fauquet, C.M., R.W. Briddon, J.K. Brown, E. Moriones and J. Stanley *et al.*, 2008. Geminivirus strain demarcation and nomenclature. *Arch. Virol.*, 153: 783-821.
- Fontes, E.P., P.A. Eagle, P.S. Sipe, V.A. Luckow and L. Hanley-Bowdoin, 1994a. Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *J. Biol. Chem.*, 269: 8459-8465.
- Fontes, E.P., H.J. Gladfelter, R.L. Schaffer, I.T. Petty and L. Hanley-Bowdoin, 1994b. Geminivirus replication origins have a modular organization. *Plant Cell*, 6: 405-416.
- Hamilton, A., O. Voinnet, L. Chappell and D. Baulcombe, 2002. Two classes of short interfering RNA in RNA silencing. *EMBO J.*, 21: 4671-4679.
- Heyraud-Nitschke, F., S. Schumacher, J. Laufs, S. Schaefer, J. Schell and B. Gronenborn, 1995. Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nucleic Acids Res.*, 23: 910-916.
- Howarth, A.J. and G.J. Vandemark, 1989. Phylogeny of geminiviruses. *J. General Virol.*, 70: 2717-2727.
- Jupin, I., F. De Kouchkovsky, F. Jouanneau and B. Gronenborn, 1994. Movement of tomato yellow leaf curl geminivirus (TYLCV): Involvement of the protein encoded by ORF C4. *Virology*, 204: 82-90.
- Koonin, E.V., A.R. Mushegian, E.V. Ryabov and V.V. Dolja, 1991. Diverse groups of plant RNA and DNA viruses share related movement proteins that may possess chaperone-like activity. *J. Gen. Virol.*, 72: 2895-2903.
- Laufs, J., W. Traut, F. Heyraud, V. Matzeit, S.G. Rogers, J. Schell and B. Gronenborn, 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.*, 92: 3879-3883.
- Lazarowitz, S.G., 1987. The molecular characterization of geminiviruses. *Plant Mol. Biol. Rep.*, 4: 177-192.
- Lazarowitz, S.G., 1992. Geminiviruses: Genome structure and gene function. *Crit. Rev. Plant Sci.*, 11: 327-349.
- Mansoor, S., Y. Zafar and R.W. Briddon, 2006. Geminivirus disease complexes: The threat is spreading. *Trends Plant Sci.*, 11: 209-212.
- Mohr, J.J., R. Clark, S. Sun, E.J. Androphy, P. MacPherson and M.R. Botchan, 1990. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with E2 transactivation. *Science*, 250: 1694-1699.
- Nagar, S., T.J. Pedersen, K.M. Carrick, L. Hanley-Bowdoin, and D. Robertson, 1995. A geminivirus induces expression of a host DNA synthesis protein in terminally differentiated plant cells. *Plant Cell*, 7: 705-719.

- Padidam, M., S. Sawyer and C.M. Fauquet, 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology*, 265: 218-225.
- Paximadis, M., A.M. Idris, I. Torres-Jerez, A. Villarreal, M.E.C. Rey and J.K. Brown, 1999. Characterization of tobacco geminiviruses in the old end new world. *Arch. Virol.*, 144: 703-717.
- Rigden, J.E., L.R. Krake, M.A. Rezaian and I.B. Dry, 1994. ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. *Virology*, 204: 847-850.
- Rojas, M.R., H. Jiang, R. Salati, B. Xoconostle-Cazares and M.R. Sudarshana *et al.*, 2001. Functional analysis of proteins involved in movement of the monopartite begomovirus, Tomato yellow leaf curl virus. *Virology*, 291: 110-125.
- Rojas, M.R., C. Hagen, W.J. Lucas and R.L. Gilbertson, 2005. Exploiting chinks in the plants armor: Evolution and emergence of geminiviruses. *Annu. Rev. Phytopathol.*, 43: 361-394.
- Rybicki, E.P., 1994. A phylogenetic and evolutionary justification for 3 genera of Geminiviridae. *Arch. Virol.*, 139: 49-77.
- Saunders, K., I.D. Bedford, R.W. Briddon, P.G. Markham, S.M. Wong and J. Stanley, 2000. A unique virus complex causes *Ageratum* yellow vein disease. *Proc. Nat. Acad. Sci.*, 97: 6890-6895.
- Silhavy, D., A. Molnar, A. Lucioli, G. Szitty, C. Hornyik, M. Tavazza and J. Burgyan, 2002. A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. *EMBO J.*, 21: 3070-3080.
- Stanley, J. and J.R. Latham, 1992. A symptom variant of beet curly top geminivirus produced by mutation of open reading frame C4. *Virology*, 190: 506-509.
- Sunter, G. and D. M. Bisaro, 1992. Transactivation of geminivirus AR1 and BR1 gene expression by the viral AL2 gene product occurs at the level of transcription. *Plant Cell*, 4: 1321-1331.
- Sunter, G., J.L. Sunter and D.M. Bisaro, 2001. Plants expressing tomato golden mosaic virus AL2 or beet curly top virus L2 transgenes show enhanced susceptibility to infection by DNA and RNA viruses. *Virology*, 285: 59-70.
- Van Wezel, R., X. Dong, H. Liu, P. Tien, J. Stanley and Y. Hong, 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. Interaction*, 15: 203-208.
- Van Wezel, R., H. Liu, Z. Wu, J. Stanley and Y. Hong, 2003. Contribution of the zinc finger to zinc and DNA binding by a suppressor of posttranscriptional gene silencing. *J. Virol.*, 77: 696-700.
- Vance, V. and H. Vaucheret, 2001. RNA silencing in plants-defense and counterdefense. *Science*, 292: 2277-2280.
- Vanitharani, R., P. Chellappan and C. Fauquet, 2005. Geminiviruses and RNA silencing. *Trends Plant Sci.*, 10: 1360-1385.
- Vanitharani, R., P. Chellappan, J.S. Pita and C.M. Fauquet, 2004. Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of *Posttranscriptional* gene silencing. *J. Virol.*, 78: 9487-9498.
- Voinnet, O., Y.M. Pinto and D.C. Baulcombe, 1999. Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. *Proc. Natl. Acad. Sci.*, 96: 14147-14152.
- Wang, H., L. Hao, C.Y. Shung, G. Sunter and D.M. Bisaro, 2003. Adenosine kinase is inactivated by geminivirus AL2 and L2 proteins. *Plant Cell*, 15: 3020-3032.