



In Silico Characterization of the Coat Protein of Whitefly-transmitted Geminiviruses

Akanksha Singh¹, Mohammad Ansar^{2,6}, H. K. Chourasia¹, Auroshikha Mohanty³, Vikash Kumar⁴ and Tushar Ranjan⁵

¹University Dept. of Botany, Tilka Manjhi Bhagalpur University, Bhagalpur, Bihar (812 007), India

²Dept. of Plant Pathology, Bihar Agricultural University, Sabour, Bhagalpur, Bihar (813 210), India

³Division of Plant Physiology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi (110 012), India

⁴Dept. of Soil Science and Agricultural Chemistry, ⁵Dept. of Molecular Biology and Genetic Engineering, Bihar Agricultural University, Sabour, Bhagalpur, Bihar (813 210), India

⁶Division of Plant Pathology, ICAR-Central Institute of Temperate Horticulture, Srinagar, Jammu and Kashmir (191 132), India



Open Access

Corresponding  mail2akanksha25@gmail.com

 0009-0009-2181-6029

ABSTRACT

This analysis was conducted from September, 2025 to June, 2026 at the Applied Microbiology and Plant Pathology Laboratory, University Department of Botany, TMBU, Bhagalpur, Bihar and the Plant Virology Laboratory, Department of Plant Pathology, BAU, Sabour, Bhagalpur, Bihar, India where computational approaches were used to assess the coat proteins of some well-characterized members of the family Geminiviridae. A comprehensive sequence analysis was performed to identify conserved regions within the protein sequences, followed by phylogenetic analysis, secondary structure prediction, homology modelling, and functional annotation to understand their structural and functional features. The tertiary structures of the six selected begomovirus coat proteins revealed a highly conserved eight-stranded antiparallel β -barrel core, indicating a high degree of structural conservation that may be important for capsid assembly and stability. In contrast, variations in surface-exposed loops and terminal regions produced subtle differences in sequence, charge, and flexibility, which may contribute to differences in host adaptation, vector transmission, and intracellular trafficking. The predicted structural models provided valuable insights into the spatial organization of the coat protein and suggested potential mechanisms underlying their multifunctional roles during viral infection. The Ramachandran plot showed good stereochemical quality and reliable backbone geometry, with most residues falling in the favoured regions, suggesting that the coat protein structures are reliable and suitable for further structural analysis. Overall, this study provided computational insights into the structural and functional characteristics of begomovirus coat proteins and highlighted their potential significance as targets for the development of antiviral strategies, diagnostic approaches, and improved crop disease management.

KEYWORDS: Geminiviridae, coat protein, phylogenetic analysis, homology modelling, virion assembly

Citation (VANCOUVER): Singh et al., In Silico Characterization of the Coat Protein of Whitefly-transmitted Geminiviruses. *International Journal of Bio-resource and Stress Management*, 2026; 17(10), 01-10. [HTTPS://DOI.ORG/10.23910/1.2026.7373](https://doi.org/10.23910/1.2026.7373).

Copyright: © 2026 Singh et al. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, that permits unrestricted use, distribution and reproduction in any medium after the author(s) and source are credited.

Data Availability Statement: Legal restrictions are imposed on the public sharing of raw data. However, author have full right to transfer or share the data in raw form upon request subject to either meeting the conditions of the original consents and the original research study. Further, access of data needs to meet whether the user complies with the ethical and legal obligations as data controllers to allow for secondary use of the data outside of the original study.

Conflict of interests: The author has declared that no conflict of interest exists.

1. INTRODUCTION

The family Geminiviridae represents one of the largest and most diverse groups of plant-infecting viruses, comprising numerous species that cause serious diseases in a wide range of economically important crops worldwide (Ansar et al., 2018; Hanley-Bowdoin et al., 2013; Zenda et al., 2024). These viruses are characterized by their unique twinned icosahedral (geminale) particle morphology and circular single-stranded DNA (ssDNA) genomes (Rojas et al., 2005; Singh and Chourasia, 2026a). Members of the Geminiviridae family are primarily transmitted by insect vectors such as whiteflies, leafhoppers, and treehoppers (aphids), which facilitate their widespread distribution and infection across different agroecosystems (Navas-Castillo et al., 2011; Singh et al., 2026a). At present, 549 species of Geminiviridae are divided into 15 genera, including Becurtovirus, Begomovirus, Capulavirus, Citlodavirus, Curtovirus, Eragovirus, Grablovirus, Maldovirus, Mastrevirus, Mulcrilevirus, Opunvirus, Topilevirus, Topocuvirus, Turncurtovirus, and Welwivirus, worldwide. In Begomovirus, there are 464 species, containing the largest of all plant-infecting virus species (Fiallo-Olive et al., 2021; Lefkowitz et al., 2026; Maruthi et al., 2026). These viruses cause devastating diseases such as tomato leaf curl, cassava mosaic disease, cotton leaf curl disease, legume golden mosaic, and okra enation mosaic, leading to significant agricultural and economic losses globally (Varma and Malathi, 2003; Singh et al., 2026b; Singh and Chourasia, 2026b). The genome of geminiviruses encodes 5–7 proteins, among them the coat protein, which plays a crucial role in viral encapsidation, protection of viral DNA, intracellular movement, and vector-mediated transmission (Lazarowitz and Beachy, 1999). Prominently, it mediates interactions with insect vectors such as whiteflies, facilitating efficient transmission between host plants (Navas-Castillo et al., 2011; Whitfield et al., 2015). Earlier studies have shown that the CP is multifunctional and is involved in complex virus–host interactions, influencing viral infectivity, pathogenicity, and host defense responses (Wang et al., 2022; Shakir et al., 2023). Due to its central role in viral assembly, transmission, and host interactions, the coat protein is a key target for understanding viral biology, developing effective antiviral strategies, diagnostic tools, and resistant crop varieties. The conserved regions identified through multiple sequence alignment provide the basis for homology modelling and three-dimensional (3D) protein structure prediction, allowing researchers to investigate the structural organization of the CP and identify residues involved in DNA binding, capsid stability, and vector recognition (Prajapat et al., 2014; Phaneendra et al., 2014). Phylogenetic analyses generated from aligned CP and complete genome sequences have revealed distinct

evolutionary lineages among begomovirus isolates and have provided evidence for geographical clustering, host-specific adaptation, and frequent recombination events that contribute to viral diversification (Leke et al., 2016; Mubin et al., 2010). In addition, comparative genomic investigations have demonstrated that MSA is fundamental for calculating nucleotide sequence identity and establishing taxonomic relationships according to species demarcation criteria, while simultaneously identifying conserved functional domains suitable for molecular diagnostics and resistance breeding (Ramesh et al., 2016; Leke et al., 2015). With advances in computational biology, in silico approaches have become essential for studying viral proteins at both structural and functional levels (Lesk, 2019). Some recent studies also emphasize that topology prediction accuracy has improved significantly with the use of machine learning and deep learning frameworks that integrate sequence, evolutionary, and structural features to refine membrane protein orientation and fold predictions (Wang et al., 2022). These advancements mark a shift from rule-based topology prediction to data-driven approaches, improving reliability and biological interpretation. This study provides a comprehensive baseline in silico characterization of the coat protein, aiming to map the structural dynamics and physicochemical properties of the coat protein across diverse whitefly-mediated geminiviruses, identify conserved evolutionary motifs critical for vector transmission, and regions and alongside establishing a theoretical foundation and generate specific molecular hypotheses that future researchers can target.

2. MATERIALS AND METHODS

2.1. Sequential analysis

The coat protein (CP) sequences of 31 viruses belonging to the Begomovirus genus of the family Geminiviridae were obtained from the NCBI database (<http://www.ncbi.nlm.nih.gov/>). Multiple sequence alignments were done using Clustal Omega (Sievers and Higgins, 2017) and subsequently refined manually to ensure accurate domain superimposition. The viruses include Okra enation leaf curl virus (OELCV) or *Begomovirus abelmoschusenation*, Bhendi yellow vein mosaic virus (BYVMV) or *Begomovirus abelmoschusflavi*, Abutilon golden mosaic virus (ABGMV) or *Begomovirus abutilonis*, Ageratum yellow vein Hualian virus (AYVHV) or *Begomovirus agerahualianense*, Ageratum leaf curl virus (AGLCV) or *Begomovirus agerainvolutionis*, Ageratum leaf curl Sichuan virus (BEGAG) or *Begomovirus agerasichuanense*, Ageratum yellow vein Sri Lanka virus (AYVSV) or *Begomovirus agerasrilankaense*, Hollyhock yellow vein mosaic virus (HYVMV) or *Begomovirus alceamuvisi*, Hollyhock yellow vein virus (HHYVC) or *Begomovirus alceavenae*, Allamanda leaf curl virus

(AMLCV) or *Begomovirus allamandae*, Alternanthera yellow vein virus (ALYVV) or *Begomovirus alternantherae*, Boerhavia yellow spot virus (BYSYV) or *Begomovirus boerhaviae*, Sweet potato leaf curl virus (SPLCV) or *Begomovirus ipomoeae*, Pepper leaf curl virus (PPLCV) or *Begomovirus capsica*, Pepper leaf curl Bangladesh virus (PLCBV) or *Begomovirus capsicumbangladeschense*, Pepper yellow leaf curl virus (PYLCV) or *Begomovirus capsicumchinaense*, Chilli leaf curl Gonda virus (CLCGV) or *Begomovirus capsigondaense*, Chilli leaf curl Kanpur virus (CLCKV) or *Begomovirus capsikanpureense*, Papaya leaf curl China virus (PLCCV) or *Begomovirus caricachinaense*, Chenopodium leaf curl virus (CPLCV) or *Begomovirus chenopodii*, Cotton leaf curl Multan virus (CLCMV) or *Begomovirus gossypimultanense*, Melon chlorotic leaf curl virus (MCLCV) or *Begomovirus chlorocucumis*, Tomato leaf curl Chuxiong virus (TYLCV) or *Begomovirus chuxiongense*, Squash leaf curl Philippines virus (SLCPV) or *Begomovirus cucurbitaphilippinense*, Datura leaf curl Sudan-Datura virus (DLCSV) or *Begomovirus daturae*, Gossypium darwinii symptomless virus (GPDSV) or *Begomovirus gossypimultanense*, Malvastrum leaf curl Philippines virus (MLCPV) or *Begomovirus malvastrumphilippinesense*, Sida yellow leaf curl virus (SYLCV) or *Begomovirus sidacontorsionis*, Tomato leaf curl virus-Bangalore (TLCBV) or *Begomovirus solanumbangaloreense* and Tomato leaf curl Gujarat virus (TLCGV) or *Begomovirus solanumgujaratense*.

2.2. Phylogenetic analysis of coat protein gene of begomoviruses

The phylogenetic trees were constructed using the maximum likelihood (ML) approach (Dhar et al., 2016) using MEGA12 (Tamura et al., 2021). The amino acid sequence data of the CP gene used to construct the phylogenetic tree were previously used in a multiple sequence alignment of amino acids. To perform multiple sequence alignment of the data, the Clustal Omega module was used with default settings. The aligned sequences were checked and manually edited in order to exclude misaligned areas or gaps to avoid any errors. To assess the robustness of the inferred phylogenies, bootstrap analysis was performed with 1,000 replicates (Pattengale et al., 2010). The generated tree visualized and edited within MEGA12, and branch lengths were drawn to scale, representing the number of substitutions per site. All positions containing gaps and missing data were treated according to the default settings of the software.

2.3. Protein topology prediction

Three-dimensional (3D) structures of the target proteins were predicted using the I-TASSER server, which combines multiple threading alignments with iterative structural assembly simulations to generate reliable protein models (Yang et al., 2015). The server identifies structural templates

from the Protein Data Bank (PDB), followed by fragment assembly and atomic-level refinement to produce predicted structures, along with confidence scores such as C-score and TM-score (Yang et al., 2015; Zhang, 2008). The predicted models were downloaded in Protein Data Bank (PDB) format and subjected to structural validation using the SAVES v6.1 platform. This integrated server includes multiple validation tools, among which the PROCHECK module was employed to assess stereochemical quality through Ramachandran plot analysis (Laskowski et al., 1993). Ramachandran plots were produced to evaluate backbone dihedral angles, phi (ϕ) and psi (ψ), for each amino acid residue. These angles define the conformational space of protein backbones and are critical indicators of structural stability. The residues were grouped into the categories of favoured, additionally allowed, generously allowed, and disallowed regions based on established stereochemical criteria (Laskowski et al., 1993). Structures exhibiting a high proportion of residues in favoured and allowed regions, with minimal residues in disallowed regions, were considered suitable for downstream analyses such as functional annotation and interaction studies (Lovell et al., 2003). Thus, the combined use of structure prediction and stereochemical validation ensured the generation of high-quality protein models for subsequent computational and structural investigations.

3. RESULTS AND DISCUSSION

3.1. Multiple sequence alignment of begomoviruses to identify structurally significant domains

Multiple sequence alignment of the amino acids (coat proteins) from diverse begomovirus isolates was performed using Clustal Omega to evaluate evolutionary conservation and identify structurally significant domains (Figure 1). The alignment revealed a remarkably high degree of primary sequence conservation throughout the polypeptide chain, highlighted by three distinct conserved blocks. In the upstream region, a prominent motif (orange-shaded box) exhibits a dense cluster of invariant amino acids, indicating strong evolutionary constraint. Further downstream within the central region of the coat protein, a second highly conserved core motif is characterized by a contiguous, unbroken stretch of five fully invariant residues (*****), with minor variations limited to the immediate flanking regions. Finally, the C-terminal region features a third distinct domain (green-shaded box) that maintains a strict arrangement of conserved hydrophobic and polar residues that act as structural anchors, despite exhibiting a higher baseline level of sequence divergence than the upstream and central motifs. The present analysis revealed that the coat proteins of begomoviruses contain conserved motifs distributed throughout their polypeptide chains; these

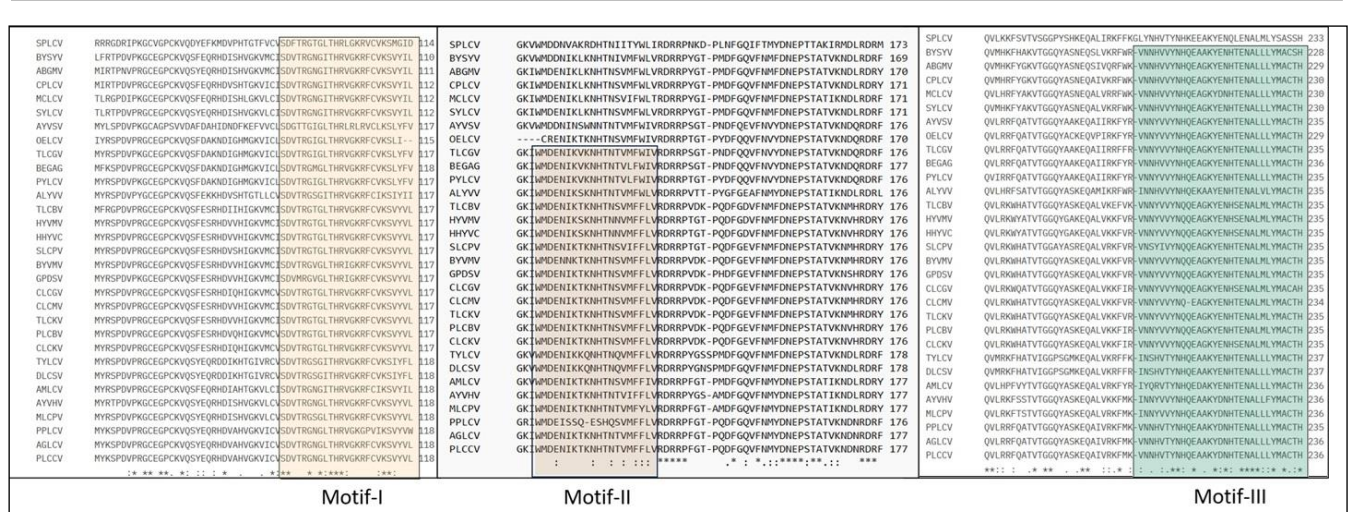


Figure 1: Multiple sequence alignments generated with Clustal Omega were manually refined to ensure accurate domain superimposition. The coloured boxes represent distinct motifs along the polypeptide chain. The highlighted blocks represent distinct regions of high conservation: the upstream motif (orange box), the central core motif featuring a contiguous block of five fully invariant residues, and the C-terminal structural domain (green box). Consensus symbols below the alignment denote the degree of amino acid conservation: asterisks (*) indicate fully conserved residues, colons (:) represent conservation between groups of strongly similar properties, and periods (.) indicate conservation between groups of weakly similar properties. Numbers on the far right indicate relative amino acid positions. Viruses used in analysis are- Okra enation leaf curl virus (OELCV), Bhendi yellow vein mosaic virus (BYVMV), Abutilon golden mosaic virus (ABGMV), Ageratum yellow vein Hualian virus (AYVHV), Ageratum leaf curl virus (AGLCV), Ageratum leaf curl Sichuan virus (BEGAG), Ageratum yellow vein Sri Lanka virus (AYVSV), Hollyhock yellow vein mosaic virus (HYVMV), Hollyhock yellow vein virus (HHYVC), Allamanda leaf curl virus (AMLCV), Alternanthera yellow vein virus (ALYVV), Boerhavia yellow spot virus (BYSV), Sweet potato leaf curl virus (SPLCV), Pepper leaf curl virus (PPLCV), Pepper leaf curl Bangladesh virus (PLCBV), Pepper yellow leaf curl virus (PYLCV), Chilli leaf curl Gonda virus (CLCGV), Chilli leaf curl Kanpur virus (CLCKV), Papaya leaf curl China virus (PLCCV), Chenopodium leaf curl virus (CPLCV), Cotton leaf curl Multan virus (CLCMV), Melon chlorotic leaf curl virus (MCLCV), Tomato leaf curl Chuxiong virus (TYLCV), Squash leaf curl Philippines virus (SLCPV), Datura leaf curl Sudan-Datura virus (DLCSV), Gossypium darwinii symptomless virus (GPDSV), Malvastrum leaf curl Philippines virus (MLCPV), Sida yellow leaf curl virus (SYLCV), Tomato leaf curl virus – Bangalore (TLCBV), Tomato leaf curl Gujarat virus (TCGV)

motifs seem to play a key role in viral assembly. Because these different motifs are perfectly conserved across such a wide range of begomoviruses, they clearly work together to keep the virus functional, likely by building the protective shell, maintaining its stability, and facilitating its interaction with its host (Crespo-Bellido et al., 2024). This tightly maintained structural framework is vital because the coat protein must balance viral DNA binding, nuclear trafficking, and self-binding to initiate shell assembly. Beyond DNA viruses, the presence of highly conserved, functional regions, such as the core jelly-roll domain found in small plant RNA and DNA virus capsid proteins, suggests a deeply rooted evolutionary blueprint optimized for structural stability and host transport machinery recruitment (Koonin et al., 2023). Furthermore, precise crosstalk between replication motifs and encapsidation pathways is essential; specific conserved structural elements within replication machinery often serve a dual purpose by regulating replication cycles while actively coordinating the downstream assembly of infectious virions (Wang et al., 2026).

3.2. Evolutionary relationships among coat proteins (CPs) of various begomoviruses

Phylogenetic trees of coat proteins from viruses in the begomovirus genus were constructed using the Maximum Likelihood method in MEGA 12. The protein sequences were then aligned using Clustal Omega in MEGA 12 with default parameters. Only branches with bootstrap values greater than 50 were retained for final tree reconstruction, using the same parameters as in Kumar et al., 2018, Zhao et al., 2022, and Brown et al., 2023. The unrooted circular phylogenetic tree resolves the analyzed *Begomovirus* isolates into distinct geographical and host-specific evolutionary lineages, radiating from deep central nodes into highly supported clades and nested sub-clades. At the bottom left of the topology, a major malvaceous-infecting clade emerges, which splits into two primary sister groups: sub-clade A1, grouping okra-infecting pathogens such as OELCV and BYVMV, and sub-clade A2, which cleanly clusters cotton-infecting species such as CLCGV. In parallel, the bottom-right of the tree is dominated by

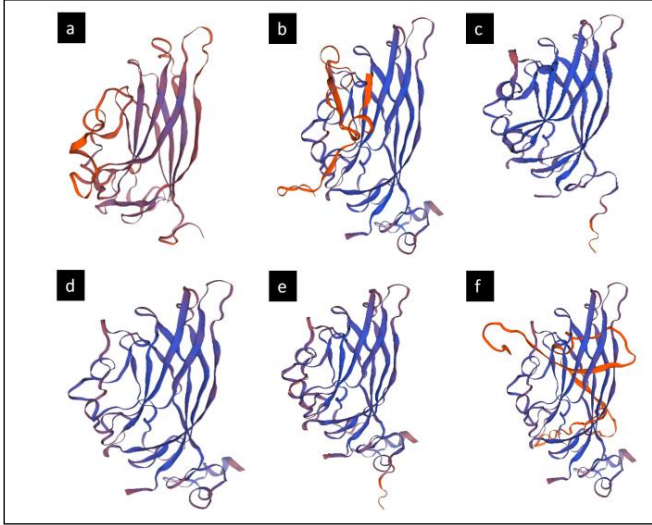


Figure 3: The predicted tertiary structure of coat protein encoded by different viruses from the Geminiviridae family is shown in cartoon view. (a) CP of TYLCV, (b) TLCBV, (c) SYLCV, (d) CPLCV, (e) PLCCV, (f) TLCKV

creating flexible, surface-exposed domains. While the overall folding topology is shared across the six structural panels, minor residue variations within these hypervariable loops govern host-specific adaptations, vector-binding affinities during selective transmission via the whitefly vector *Bemisia tabaci*, and unique interactions with host cellular components during pathogenesis (Navas-Castillo et al., 2011). Ultimately, this unified structural arrangement demonstrates that the evolutionary selection pressures acting on these Begomovirus species preserve the foundational β -barrel capsid architecture to maintain virion stability, while allowing sufficient conformational flexibility in surface domains to facilitate adaptation to diverse plant hosts and distinct ecological niches (Simmonds et al., 2024). In the present analysis, the primary functional regions are structurally the same across all six coat protein models, but they differ in minor key variations in their genetic sequences. Because all six belong to the same genus, *Begomovirus*, they share a strict structural blueprint. However, identical regions differ, creating biological variations such as the Capsid Assembly Core (the β -Barrel), in which the central, rigid blue structure is 100% identical in shape across all six models. Every single one of these viruses requires this exact layout to connect protein monomers into pentons and build a stable, twin (geminate) viral capsid. While the structural boundaries are identical, the amino acid alignment within those regions shows subtle variations.

The highly conserved sequence blocks highlighted in bold colours (e.g., PKGCVGPCKV, CVSDFTRGTGLTHR, and YNHQEAAYKYNHTENAL) form the rigid, inner skeleton of the protein. In the 3D structures (C, D, E), this corresponds to the stable jelly-roll β -barrel fold rendered

in blue/purple. This architecture is strictly preserved across all aligned species (like TYLCV, SLCPV, and CLCMV) to maintain capsid stability. The highlighted orange loops (fig B and F) directly align with the numbers inside the parentheses of the alignment sheet. The Primary Variable Loop corresponds to the sequence insertion marked by (29) or (21) residues between VKSM and RDRRP/YNHQEAAYKYNHTENAL. The Secondary Variable Loop corresponds to the (16) and (14) residue gaps located between the conserved PKGCVGPCKV and CVSDFTRGTGLTHR segments. These loops bulge outward from the surface of the capsid, making them accessible to external molecular interactions. The conserved motifs act as functional structural anchors or molecular recognition signals, such as MTGRMR (Red, N-Terminus), which acts as a nuclear localization signal (NLS) and binds to viral DNA to facilitate genomic packaging into the capsid core. PKGCVGPCKV and CVSDFTRGTGLTHR (Blue/Gold, Core) form the highly stabilized backbone strands of the jelly-roll β -barrel, which keep the structural integrity of the capsid intact under environmental stress. RDRRP (Green, Loop Adjacent) forms a crucial epitope junction right where the internal core transitions into an external loop. The structures of these different viruses appear identical in their blue/purple zones (Figure C, D, E). This perfectly mirrors the unbroken block alignment of the sequence sheets, as all 6 strains possess the identical PKGCVGPCKV (blue) and CVSDFTRGTGLTHR (gold) segments.

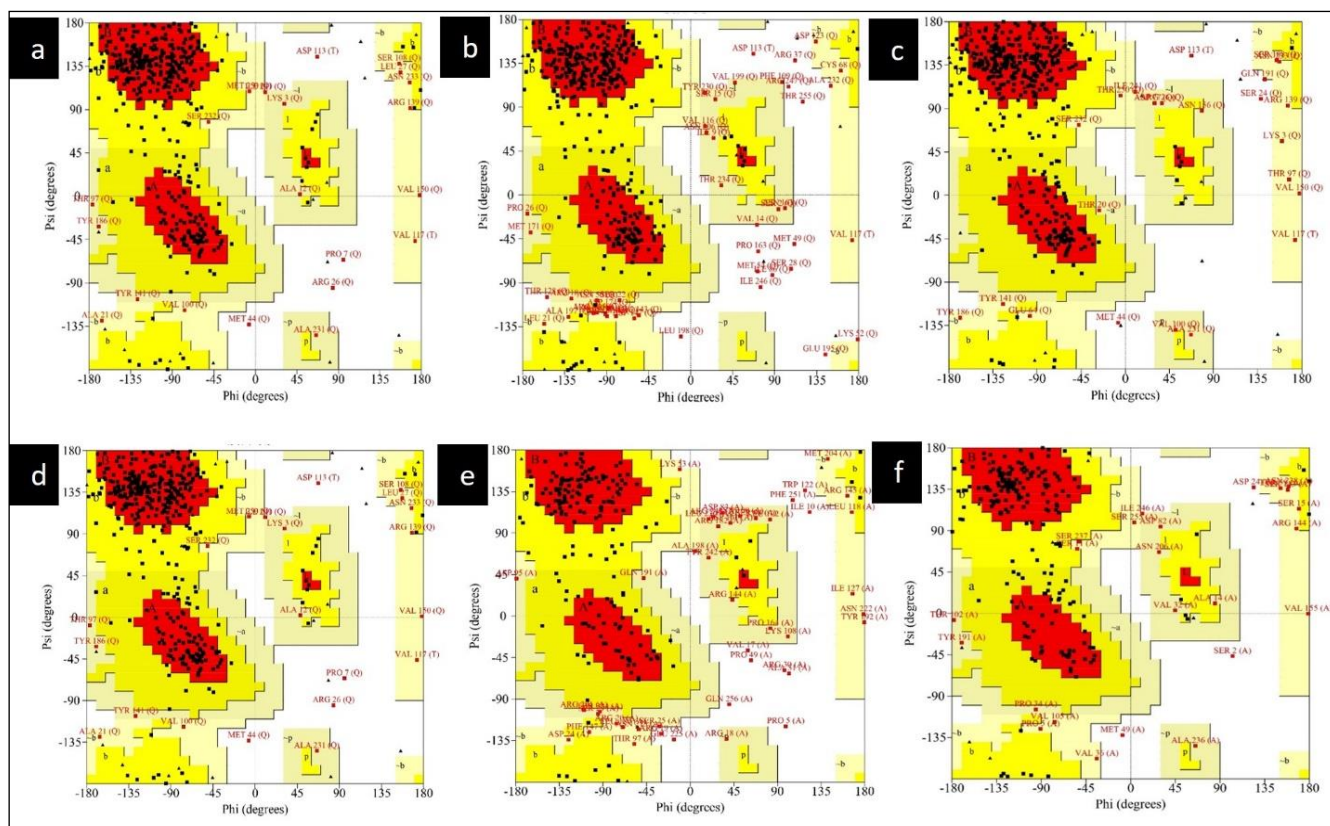
The core differences among these 6 strains reside inside the parentheses (e.g., 55, 60, 16, 29), which represent the exact number of variable amino acids bridging the conserved blocks. These variable gaps build the extended outer surface loops highlighted in orange, such as monomers with shorter loops (SLCPV, BYSV, ABGMV) and monomers with extended loops (TLCGV, AYSV, TYLCV). This creates localized differences across panels A through F, such as differences in vector transmission (the loops). The structural loop spanning residues 129 to 134 is a universal feature. However, in some of the models, like TYLCV or TLCBV, the alignment preserves exact residues (such as Glutamine, Q129/Q134) that fit perfectly with the whitefly vector *Bemisia tabaci*. Specific mutations within the outer loops of the begomovirus coat protein alter the local biochemical charges without disrupting the physical structure of the viral capsid (Wang et al., 2020). These minor chemical transformations directly dictate how efficiently the virus interacts with internal cell receptors across different cryptic species of the *Bemisia tabaci* whitefly complex (Fiallo-Olivé et al., 2020). In other non-tomato isolates, such as SYLCV or CPLCV, minor mutations occur inside this loop. The shape of the loop remains the same, but the

altered surface chemistry can change how effectively that specific virus is acquired or transmitted by different whitefly biotypes. Another difference is in Nuclear-shuttling (The Basic Tails), where, while the N-terminal basic tails of PLCCV and TLCKV display unique charge variations, the remaining four structures reveal distinct host-adapted tail configurations. The core tomato pathogens, TYLCV and TLCBV, exhibit a highly conserved, dense arrangement of Arginine (R) and Lysine (K) residues that optimizes nuclear import kinetics via the importin α -pathway, facilitating rapid systemic infection in plants. In contrast, the weed-associated SYLCV possesses distinct amino acid substitutions within this basic segment, shifting its net charge profile to favour a balanced, low-titer replication cycle that allows persistent survival in wild malvaceous reservoirs. Furthermore, CPLCV presents unique sequence variations in its flexible terminal tail, demonstrating a specialized spatial configuration optimized to interact with the nuclear transport machinery of its specific non-solanaceous host. Collectively, these data confirm that while the unstructured shape of the basic tail is universally maintained across all six

models, minor evolutionary adjustments to sequence length and net charge density allow each *Begomovirus* species to optimize its intracellular trafficking according to its specific ecological niche.

3.3. Ramachandran map analysis of coat protein of various begomoviruses

The stereochemical quality and backbone geometry of the modelled coat protein were validated using the PROCHECK algorithm via the SAVES v6.1 web-based pipeline. This plot was generated to analyze the distribution of phi (ϕ) and psi (ψ) dihedral angles. The model's reliability was assessed by the percentage of residues residing in the most favoured, additionally allowed, and generously allowed regions (Figure 4 a, b, c, d, e, f). Across all six generated plots, the protein backbone dihedral angles phi (ϕ) and psi (ψ) demonstrate an exceptionally robust and statistically sound clustering profile. The vast majority of the amino acid residues occupy the core energetically favored regions (represented by the dark red zones) and the additionally allowed regions (represented by the bright yellow zones).



This concentrated distribution confirms that the primary polypeptide backbones adhere strictly to stereochemically stable geometric parameters and maintain proper spatial conformations across the entire set of models. A minimal fraction of the total residue count occupies the generously allowed regions (pale yellow zones), while only a few highly localized residues migrate into the disallowed white space as statistical outliers. A comparative evaluation of these outlier residues across the panels reveals that they do not indicate structural instability. Instead, these scattered outliers are concentrated almost exclusively within flexible, surface-exposed loop systems and unstructured terminal regions. Because these localized loops are inherently dynamic and deviate naturally from rigid secondary structure constraints, elevated conformational freedom is expected and does not compromise the overall structural integrity of the models. Furthermore, frequent appearances of glycine and proline residues within or near these relaxed boundaries are typical, owing to their unique side-chain properties of extreme flexibility and rigid cyclization, respectively. Consequently, the high proportion of residues aligning within statistically favourable boundaries across the entire multi-panel dataset indicates excellent geometric optimization, confirming that all six models possess the required structural quality and reliable molecular modelling applications. (Similar observations have been reported for structurally validated begomovirus coat protein models, where a high percentage of residues in favoured regions was considered indicative of stable and biologically relevant protein conformations (Kumar et al., 2022; Roy et al., 2010).

4. CONCLUSION

Comparative sequence analysis revealed significant evolutionary, structural, and functional features of six begomovirus coat proteins. All proteins shared a highly conserved eight-stranded antiparallel β -barrel core, whereas variations in surface-exposed and terminal regions showed that these may influence host adaptation, vector transmission, and intracellular trafficking. Structural validation indicated reliable models with good stereochemical quality. These findings shed light on the molecular behaviour and evolutionary dynamics of the coat protein, with potential implications for enhanced virus detection, characterization, and disease management.

5. ACKNOWLEDGEMENT

The authors would like to acknowledge the Plant Virology Laboratory, Department of Plant Pathology, Bihar Agricultural University, Sabour, Bhagalpur, Bihar and the Applied Microbiology and Plant Pathology Laboratory, University Department of Botany, Tilka Manjhi Bhagalpur University, Bhagalpur.

6. REFERENCES

- Ansar, M., Akram, M., Agnihotri, A.K., Srinivasaraghavan, A., Saha, T., Naimuddin, 2018. Characterisation of leaf curl virus in chili and overwintering role of nightshade in linkage between chilli and tomato. *Journal of Plant Pathology* 101, 307–314. DOI: <https://doi.org/10.1007/s42161-018-0182-z>.
- Brown, D., Godfrey, C., Konz, N., Tu, J., Kvinge, H., 2023. Understanding the inner workings of language models through representation dissimilarity. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 6543–6558. DOI: <https://doi.org/10.18653/v1/2023.emnlp-main.403>.
- Crespo-Bellido, A., Hoyer, J.S., Burgos-Amengual, Y., Duffy, S., 2024. Phylogeographic analysis of Begomovirus coat and replication-associated proteins. *The Journal of General Virology* 105(10), 002037. DOI: <https://doi.org/10.1099/jgv.0.002037>.
- Dhar, A., Minin, V., 2016. Calculating higher-order moments of phylogenetic stochastic mapping summaries in linear time. *Journal of Computational Biology* 24(5), 377–399. DOI: <https://doi.org/10.1089/cmb.2016.0172>.
- Felsenstein, J., 1985. Phylogenies and the comparative method. *The American Naturalist* 125(1), 1–15. Available from: <https://www.jstor.org/stable/2461605>.
- Fiallo-Olive, E., Lett, J.M., Martin, D.P., Roumagnac, P., Varsani, A., Zerbini, F.M., Navas-Castillo, J., 2021. ICTV virus taxonomy profile: Geminiviridae 2021. *Journal of General Virology* 102(12), 001696. DOI: <https://doi.org/10.1099/jgv.0.001696>.
- Fiallo-Olive, E., Pan, L.L., Liu, S.S., Navas-Castillo, J., 2020. Transmission of begomoviruses and other whitefly-borne viruses: dependence on the vector species. *Phytopathology* 110(1), 10–17. DOI: <https://doi.org/10.1094/PHYTO-07-19-0273-FI>.
- Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D., Mansoor, S., 2013. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology* 11(11), 777–788. DOI: <https://doi.org/10.1038/nrmicro3117>.
- Koonin, E.V., Dolja, V.V., Krupovic, M., 2023. Plant virus movement proteins originated from jelly-roll capsid proteins of plant viruses, emphasizing the key role of exaptation in virus evolution. *Proceedings of the National Academy of Sciences* 120(27), DOI: <https://doi.org/10.1371/journal.pbio.3002157>.
- Kumar, P., Oraon, P.K., Yadav, P., Roy, A., Goel, S., Reddy, M.K., Mukherjee, S.K., Mandal, B., 2022. Random distribution of nucleotide polymorphism throughout the genome of tomato-infecting begomovirus species occurring in India: implication in PCR-based

- diagnosis. *Virusdisease* 33(3), 270–283. DOI: <https://doi.org/10.1007/s13337-022-00785-9>.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35(6), 1547–1549. DOI: <https://doi.org/10.1093/molbev/msy096>.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26, 283–291. DOI: <https://doi.org/10.1107/S0021889892009944>.
- Lazarowitz, S.G., Beachy, R.N., 1999. Viral movement proteins as probes for intracellular and intercellular trafficking in plants. *The Plant Cell* 11(4), 535–548. DOI: <https://doi.org/10.1105/tpc.11.4.535>.
- Lefkowitz, E., Hendrickson, R., 2026. International Committee on taxonomy of viruses. (2026). ICTV Master Species List (MSL) [Dataset]. International Committee on Taxonomy of Viruses (ICTV). DOI: <https://doi.org/10.5281/zenodo.19154110>.
- Leke, W.N., Mignouna, D.B., Brown, J.K., Kvarnheden, A., 2015. Begomovirus disease complex: Emerging threat to vegetable production systems of West and Central Africa. *Agriculture & Food Security* 4(1), Article 1. DOI: <https://doi.org/10.1186/s40066-014-0020-2>.
- Leke, W.N., Khatabi, B., Mignouna, D.B., Brown, J.K., Fondong, V.N., 2016. Complete genome sequence of a new bipartite begomovirus infecting cotton in the Republic of Benin in West Africa. *Archives of Virology* 161(8), 2329–2333. DOI: <https://doi.org/10.1007/s00705-016-2894-8>.
- Lesk, A.M., 2019. Introduction to bioinformatics (5th ed.). Oxford University Press. Available from: <https://global.oup.com/academic/product/introduction-to-bioinformatics-9780198794141?cc=inandlang=enand>.
- Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Ca geometry: ϕ , ψ and C β deviation. *Proteins: Structure, Function, and Bioinformatics* 50(3), 437–450. DOI: <https://doi.org/10.1002/prot.10286>.
- Maruthi, M.N., Roy, A., Malathi, V.G., Venkataravanappa, V., Prasad, M., Sharma, N., Legg, J.P., Sastry, K.S., Zerbini, F.M., 2026. Taxonomy of family: *Geminiviridae*. In Sastry, K.S., Debat, H., Maruthi, M.N., Dasgupta, I., Rao, G.P., Bhat, A.I. (Eds.), *Taxonomy and classification of plant viruses and viroids*. Springer. DOI: https://doi.org/10.1007/978-981-97-8408-0_2.
- Mubin, M., Amin, I., Amrao, L., Briddon, R.W., Mansoor, S., 2010. The hypersensitive response induced by the V2 protein of a monopartite begomovirus is countered by the C2 protein. *Molecular plant Pathology* 11(2), 245–254. DOI: <https://doi.org/10.1111/j.1364-3703.2009.00601.x>.
- Navas-Castillo, J., Fiallo-Olive, E., Sanchez-Campos, S., 2011. Emerging virus diseases transmitted by whiteflies. *Annual Review of Phytopathology* 49, 219–248. DOI: <http://dx.doi.org/10.1146/annurev-phyto-072910-095235>.
- Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R., Moret, B.M., Stamatakis, A., 2010. How many bootstrap replicates are necessary? *Journal of Computational Biology* 17(3), 337–354. DOI: <https://doi.org/10.1089/cmb.2009.0179>.
- Persson, B., Argos, P., 1996. Topology prediction of membrane proteins. *Protein science: a publication of the Protein Society* 5(2), 363–371. DOI: <https://doi.org/10.1002/pro.5560050221>.
- Phaneendra, C., Sambasiva Rao, K.R.S., Kapoor, R., Jain, R.K., Mandal, B., 2014. Fusion coat protein of *Pumpkin yellow vein mosaic virus* with maltose binding protein: Applications in immunodiagnosis of begomoviruses. *Virus Disease* 25(3), 390–393. DOI: <https://doi.org/10.1007/s13337-013-0189-1>.
- Prajapat, R., Marwal, A., Gaur, R.K., 2014. Recognition of errors in the refinement and validation of three-dimensional structures of AC1 proteins of begomovirus strains by using ProSA-Web. *Journal of Viruses* 2014, 752656, 6. DOI: <https://doi.org/10.1155/2014/752656>.
- Ramesh, S.V., Chouhan, B., Gupta, G., Ramteke, R., Chand, S., Husain, S., 2016. Molecular diversity analysis of coat protein gene encoded by legume begomoviruses and PCR assay to detect yellow mosaic viruses infecting soybean in India. *British Biotechnology Journal* 12(3), 1–10. DOI: <https://doi.org/10.9734/BBJ/2016/24362>.
- Rojas, M.R., Hagen, C., Lucas, W.J., Gilbertson, R.L., 2005. Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. *Annual Review of Phytopathology* 43, 361–394. DOI: <https://doi.org/10.1146/annurev.phyto.43.040204.135939>.
- Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols* 5(4), 725–738. DOI: <https://doi.org/10.1038/nprot.2010.5>.
- Shakir, S., Mubin, M., Nahid, N., Serfraz, S., Qureshi, M.A., Lee, T.K., Liaqat, I., Lee, S., Nawaz-ul-Rehman, M.S., 2023. REPerceptions: How geminiviruses recruit host factors for replication. *Frontiers in Microbiology* 14, 1224221. DOI: <https://doi.org/10.3389/fmicb.2023.1224221>.

- Sievers, F., Higgins, D.G., 2017. Clustal omega for making accurate alignments of many protein sequences. *Protein Science* 27, 135–145. DOI: <https://doi.org/10.1002/pro.3290>.
- Simmonds, P., Adriaenssens, E.M., Lefkowitz, E.J., Oksanen, H.M., Siddell, S.G., Zerbini, F.M., Alfenas-Zerbini, P., Aylward, F.O., Dempsey, D.M., Dutilh, B.E., Freitas-Astua, J., Garcia, M.L., Hendrickson, R.C., Hughes, H.R., Junglen, S., Krupovic, M., Kuhn, J.H., Lambert, A.J., Lobočka, M., Mushegian, A.R., Varsani, A., 2024. Changes to virus taxonomy and the ICTV Statutes ratified by the International Committee on Taxonomy of Viruses (2024). *Archives of Virology* 169(11), 236. DOI: <https://doi.org/10.1007/s00705-024-06143-y>.
- Singh, A., Chourasia, H.K., 2026a. Plant viruses: Diversity, evolution, and their impact on global agriculture-A comprehensive review. *International Journal of Scientific Research and Engineering Development* 9(2), 1280–1290. Available from: <https://ijsred.com/volume9/issue2/IJSRED-V9I2P183.pdf>.
- Singh, A., Chourasia, H.K., 2026b. An overview of begomoviruses: Structure, functions, and interactions. *International Journal of Trend in Scientific Research and Development* 10(2), 463–473. Available from: <https://www.ijtsrd.com/papers/ijtsrd101345.pdf>.
- Singh, A., Ansar, M., Chourasia, H.K., Kumar, V., 2026b. Prevalence of Begomoviruses and their Molecular Detection in Major Solanaceous Crops of Eastern Bihar, India. *Journal of Advances in Microbiology* 26(7), 89–98. DOI: <https://doi.org/10.9734/jamb/2026/v26i71144>.
- Singh, A., Kumar, V., Chourasia, H.K., 2026a. Unravelling the begomovirus-mediated tomato leaf curl disease: An in-depth review of pathogenesis and host interaction. *International Journal of Biochemistry Research and Review* 35(2), 119–137. DOI: <https://doi.org/10.9734/ijbcr/2026/v35i21107>.
- Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution* 38, 3022–3027. DOI: <https://doi.org/10.1093/molbev/msab120>.
- Varma, A., Malathi, V.G., 2003. Emerging geminivirus problems: A serious threat to crop production. *Annals of Applied Biology* 142, 145–164. DOI: <https://doi.org/10.1111/j.1744-7348.2003.tb00240.x>.
- Wang, H.L., Pan, L.L., Liu, S.S., Wang, X.W., 2020. Mutations in the coat protein of a begomovirus result in altered whitefly transmission characteristics. *Scientific Reports* 10(1), Article 4005. DOI: <https://doi.org/10.1093/ve/veaa014>.
- Wang, L., Zhong, H., Xue, Z., Wang, Y., 2022. Improving the topology prediction of α -helical transmembrane proteins with deep transfer learning. *Computational and Structural Biotechnology Journal* 20, 1993–2000. <https://doi.org/10.1016/j.csbj.2022.04.024>.
- Wang, Z., Jamasb, A.R., Hajij, M., Morehead, A., Braithwaite, L., Lio, P., 2025. Topotein: topological deep learning for protein representation learning. DOI: <https://doi.org/10.48550/arXiv.2509.03885>.
- Wang, Y., Zhang, J., Zhou, X., 2026. The geminiviral effector AC4 suppresses nonsense-mediated decay by coordinating target degradation pathways. *Plant, Cell and Environment* 49(2), 452–465. DOI: <https://doi.org/10.1111/pce.70452>.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, A., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* 46, W296–W303. DOI: <https://doi.org/10.1093/nar/gky427>.
- Whitfield, A.E., Falk, B.W., Rotenberg, D., 2015. Insect vector-mediated transmission of plant viruses. *Virology* 479–480, 278–289. DOI: <https://doi.org/10.1016/j.virol.2015.03.026>.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER suite: Protein structure and function prediction. *Nature Methods* 12, 7–8. DOI: <https://doi.org/10.1038/nmeth.3213>.
- Zenda, M.F., Masamba, P., Allie, F., Kappo, A.P., 2024. Geminiviruses and food security: A molecular genetics perspective for sustainable agriculture in Africa. *Plants* 13(19), 2768. DOI: <https://doi.org/10.3390/plants13192768>.
- Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9, 40. DOI: <https://doi.org/10.1186/1471-2105-9-40>.
- Zhao, Q., Yang, M., Cheng, Z., Li, Y., Wang, J., 2022. Biomedical data and deep learning computational models for predicting compound-protein relations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 19(4), 2092–2110. DOI: <https://doi.org/10.1109/tcbb.2021.3069040>.