Original Article
DOI: https://doi.org/10.70749/ijbr.v3i2.648

INDUS JOURNAL OF BIOSCIENCE RESEARCH
https://induspublishers.com/IJBR
ISSN: 2960-2793/ 2960-2807

Characterization of Potential Virus Resistance Genes in Different Crops Through In-silico Approaches

Mehar Ali Raza[1], Rida Zaib[2], Aimen Khalid[3], Amna Afzal[1], Faheem Kanwal[4], Muhammad Azmat[4], Imran Zafar[5], Shaista Shafiq[5]

[1]Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Faisalabad, Punjab, Pakistan.
[2]Department of Microbiology, Faculty of Science and Technology, University of Central Punjab, Lahore, Punjab, Pakistan.
[3]Department of Zoology, Government College University Faisalabad, Faisalabad, Punjab, Pakistan.
[4]Institute of Molecular Biology and Biotechnology (IMBB), University of Lahore, Punjab, Pakistan.
[5]Department of Biochemistry and Biotechnology, The University of Faisalabad (TUF), Faisalabad, Punjab, Pakistan.

ARTICLE INFO

Keywords

Cotton Leaf Curl Virus (CLCuV), Resistance Gene Analogs (RGAs), Gossypium Hirsutum, Molecular Modelling, Bioinformatics Analysis.

Corresponding Author: Shaista Shafiq, Department of Biochemistry and Biotechnology, The University of Faisalabad (TUF), Faisalabad, Punjab, Pakistan.
Email: s.shafiq@hotmail.com

Declaration

Authors' Contribution: All authors equally contributed to the study and approved the final manuscript.

Conflict of Interest: No conflict of interest.

Funding: No funding received by the authors.

Article History

Received: 21-10-2024
Revised: 17-01-2025
Accepted: 04-02-2025

ABSTRACT

Begomoviruses, particularly the cotton leaf curl virus (CLCuV), pose significant threats to global agriculture, especially cotton production. This study identified five resistance gene analogs (RGAs)—KT250635, KT886994, KT633945, KT885194, and KT633946—in Gossypium hirsutum and evaluated their potential against CLCuV using bioinformatics and molecular modeling approaches. Structural validation through Ramachandran plot analysis demonstrated that KT250635 and KT886994 had 92.6% residues in the most favored regions, while KT633945 and KT633946 exhibited slightly lower stereochemical reliability, requiring further refinement. GMQE scores ranged from 0.48 to 0.79, with KT250635 achieving a high residue quality score of 0.90. Functional annotation revealed significant homology, with KT250635 sharing 93.1% similarity with Sorghum bicolor and 97.1% with Gossypium raimondii, suggesting broad-spectrum resistance potential. Protein modeling and validation through I-TASSER and QMEAN-Z scores demonstrated structural stability, with KT250635 emerging as the most promising candidate. Phylogenetic analysis clustered KT250635 and KT886994 closely with resistance-related genes across diverse taxa, highlighting evolutionary conservation and functional significance. Additionally, KT633945 and KT885194 exhibited genetic similarity with peach and wild legumes, suggesting potential cross-species resistance traits. Bootstrap analysis with 1000 replicates ensured the robustness of the phylogenetic clustering. These findings provide a strong foundation for breeding CLCuV-resistant cotton varieties and underscore the importance of genetic insights in sustainable crop protection. These results contribute to understanding resistance mechanisms in cotton and may aid in the genetic improvement of susceptible varieties. Future studies should explore the functional role of these genes in resistance pathways and their potential applicability in other crop species to enhance resilience against viral pathogens.

INTRODUCTION

Begomoviruses, members of the *Geminiviridae* family, represent a significant threat to global agriculture due to their ability to infect a wide range of dicotyledonous plants. These single-stranded DNA (ssDNA) viruses are primarily transmitted by the whitefly (*Bemisia tabaci*) in a *circulative* manner, perpetuating their spread and impact [1]. Among the most economically damaging pathogens, *begomoviruses* target essential crops such as tomatoes, beans, cassava, squash, and cotton. This leads to significant yield losses and threatens food security, particularly in tropical and subtropical regions [2]. *Begomoviruses* are known for their restricted vascular system localization, though certain bipartite species, such as the Bean golden yellow mosaic virus, extend their invasion to mesophyll tissues. Their pathogenicity varies depending on the developmental stage of the host tissue, which influences viral load and accumulation kinetics [3]. Infected crops display symptoms such as leaf curling, vein clearing, and stunted growth, directly impacting fruit quality and productivity. These viruses have been extensively studied in terms of genome structure, DNA replication, protein interactions, and transgenic resistance, underscoring their complex mechanisms of host manipulation [4, 5].

Cotton leaf curl virus (CLCuV), a *begomovirus* species, exemplifies the challenges posed by these pathogens. CLCuV, particularly in Gossypium hirsutum, significantly hampers cotton productivity. While efforts have been made to explore resistance mechanisms in Gossypium arboreum, the precise molecular basis of resistance remains elusive. *Begomoviruses* exploit their tiny genomes and coding efficiency by interacting with host plant proteins to facilitate DNA replication, gene expression, and immune evasion, ultimately compromising plant growth and yield [6]. Plants, however, have evolved intricate defense mechanisms, including activating resistance (R) genes and resistance gene analogs (RGAs). RGAs, such as nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins, receptor-like kinases, and pentatricopeptide repeat proteins, play crucial roles in pathogen recognition and immunity. Advances in bioinformatics have enabled the identification of RGAs by leveraging structural and functional domains, allowing for genome-wide mapping and analysis of quantitative trait loci (QTLs) associated with disease resistance [7, 8].

In recent years, computational tools such as BLAST, HMM-based algorithms, and the Orthologous Matrix (OMA) database have revolutionized the identification and annotation of orthologous genes and RGAs. These tools facilitate the prediction of gene functions, phylogenetic relationships, and comparative genomics, enabling researchers to elucidate the molecular underpinnings of resistance mechanisms [9, 10]. Orthologous genes, derived from speciation events, often retain conserved functions across species and play pivotal roles in adaptive evolution. Comparative genomic analyses leveraging orthology predictions have provided critical insights into gene function, phylogenetic divergence, and species evolution. Tools such as OMA offer robust platforms for high-quality ortholog identification, functional annotation, and integration of custom gene datasets, further advancing our understanding of plant resistance pathways [11, 12].

This study explores the identification and characterization of potential virus-resistance genes in diverse crops using in-silico approaches. By integrating bioinformatics tools and databases, we aim to unravel the genetic basis of plant-virus interactions and identify candidate genes for enhancing resistance. This research advances our understanding of plant immunity and contributes to developing resilient crop varieties, addressing the pressing need for sustainable agriculture in the face of viral epidemics.

MATERIAL AND METHODS

Expression and Retrieval of Resistance Gene Analogues (RGAs)

The expression of Resistance Gene Analogues (RGAs) was evaluated in symptomatic and asymptomatic cotton plants to determine their role in resistance against Cotton Leaf Curl Virus (CLCuV). It was observed that the expression levels of RGAs varied significantly across different genotypes, with some showing comparable levels in both diseased and healthy plants. Resistance gene analogs such as KT250635, KT886994, KT633945, KT885194, and KT633946 were identified as highly or moderately expressed, particularly in Gossypium arboreum, which possesses a strong innate defense mechanism against CLCuV [13, 14]. These sequences were retrieved from the National Center for Biotechnology Information (NCBI) database using their unique accession numbers. The retrieved sequences were downloaded in FASTA format for subsequent bioinformatics analyses [15]. This step established the foundation for studying RGA variability and functionality in cotton species.

Sequence Translation and Ortholog Prediction

The DNA sequences of the retrieved RGAs were translated into protein sequences using the ExPASy Translate tool [16]. This step was essential for conducting protein-level analyses and predicting orthologs. The OMA browser, a robust tool for identifying orthologous relationships across genomes, was employed to predict orthologs for each of the five RGA protein sequences [17]. Protein sequences from various crop species were analyzed in OMA, and orthologous sequences with high similarity scores were identified. These sequences, representing genes with functional conservation across species, were downloaded in FASTA format. Ortholog prediction provided insights into the potential conservation of RGA function across different crops, paving the way for cross-species comparisons.

Refinement of Ortholog Sequences

The ortholog sequences obtained from the OMA browser were refined to ensure compatibility with alignment and structural analysis tools. This process involved systematically organizing the FASTA sequences using a text editor [11, 18]. Non-essential details such as crop names, ortholog ratios, and additional metadata were removed, leaving only the accession numbers and sequences. The model crop sequence was positioned at the top, followed by the sequences of subject crops in descending order of similarity. This streamlined format facilitated accurate and efficient analysis, clearly differentiating the model and orthologous sequences. Refinement of the data was critical for ensuring that subsequent alignments and evolutionary analyses were reliable and interpretable.

Phylogenetic and Molecular Evolutionary Analysis

To explore the evolutionary relationships among the

RGAs, phylogenetic trees were constructed using the Molecular Evolutionary Genetics Analysis (MEGA) software. The evolutionary distances between the model and subject crops were calculated to reveal clustering patterns and genetic relatedness [19, 20]. Each RGA, including KT250635, KT886994, KT633945, KT885194, and KT633946, was analyzed individually, and their phylogenetic trees provided a visual representation of evolutionary divergence and conservation. This analysis identified closely related species, highlighting potential candidates for resistance gene transfer and breeding programs. Phylogenetic analysis is critical in understanding how resistance traits are distributed and conserved among crop species.

Pairwise Alignment and Divergence Analysis

The sequences of model and subject crops were subjected to pairwise alignment using the MegAlign Pro software, designed for multiple and pairwise alignments. This step provided a quantitative measure of sequence divergence and similarity between the model and orthologous crops. Percent divergence scores were calculated to identify crops with the closest sequence homology to the model crop [20, 21]. High-scoring crops were shortlisted for further analysis, ensuring that only the most promising candidates were advanced to the following stages. Pairwise alignment and divergence analysis were crucial for narrowing down the pool of subject crops to those with the most significant potential for structural and functional conservation.

Three-Dimensional Protein Structure Prediction

The 3D structures of model and selected subject crops were predicted using the SWISS-MODEL server, an automated homology modeling tool [22, 23]. This step involved uploading the protein sequences in FASTA format and using the server to generate structural models based on template structures available in the Protein Data Bank (PDB). The generated models were evaluated for quality and downloaded as PDB files [24]. These structures provided a detailed visualization of the protein conformations, essential for understanding the functional domains of RGAs. Homology modeling also served as the basis for subsequent structural superposition and alignment analyses, enabling the exploration of structural differences and similarities among crops.

Structural Superposition and Alignment

Structural superposition of the 3D models was performed using the FATCAT (Flexible structure Alignment by Chaining Aligned fragment pairs allowing Twists) server [25]. This tool aligned the model crop structures with those of the subject crops, accounting for rigid and flexible protein structure regions. Superposition analysis provided insights into the structural conservation of RGAs, highlighting regions of functional importance. The results revealed structural similarities and deviations, which are critical for identifying crops with potential functional equivalence. PDB files of the model and subject crop structures were uploaded to the FATCAT server, and the alignment results were generated in seconds, offering a comprehensive view of the structural relationships.

Integration of Tools and Data Interpretation

Integrating bioinformatics tools and databases was key to achieving the objectives of this study. NCBI was used for sequence retrieval [23], ExPASy for sequence translation [26], OMA for ortholog prediction [27], and MEGA for phylogenetic analysis [28]. SWISS-MODEL enabled protein structure prediction [29], and FATCAT facilitated structural superposition [30]. The collective interpretation of these analyses provided a comprehensive understanding of the genetic and structural conservation of RGAs, advancing our knowledge of resistance mechanisms in cotton and other crops.

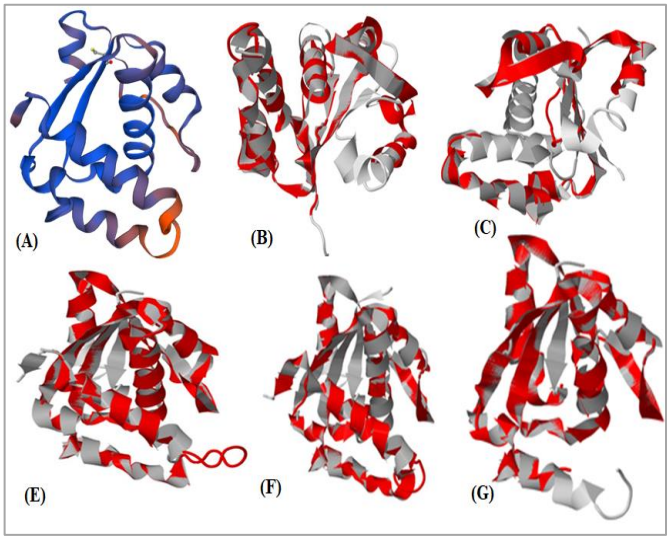
RESULTS

Structural Conservation of Resistance Gene KT250635 Across Plant Species

The highly conserved gene KT250635, identified as a resistance gene analogue (AN), shares significant homology with five orthologous plant species: *PHAAN32355* (Red mung bean), *BRANA51611* (Mustard plant), *GOSHI67843* (Mexican cotton), *VITVI13198* (Grape vine), and *SOYBN02273* (Soya bean), as determined through hypothetical values. The 3D structure of KT250635, depicted in **Figure 2(A)**, was modeled using the TMV resistance protein N-like template (5ku7.1.A) with a sequence identity of 59.02%. Structural alignment analyses using FATCAT Server revealed significant similarities between KT250635 and its orthologs. Superposition with *PHAAN32355* **(Figure 2B)** showed a P-value of 2.72e-08, 84 equivalent positions, and an RMSD of 2.58. Similarly, alignment with *SOYBN02273* **(Figure 2C)** yielded a P-value of 0.00e+00, 124 equivalent positions, and an RMSD of 0.63. For *GOSHI67843* **(Figure 2D)**, the P-value was 0.00e+00, with 119 equivalent positions and an RMSD of 0.97. Alignment with *VITVI13198* **(Figure 2E)** demonstrated a P-value of 0.00e+00, 116 equivalent positions, and an RMSD of 0.84. Finally, the superposition with *BRANA51611* **(Figure 2F)** showed a P-value of 0.00e+00, 103 equivalent positions, and an RMSD of 0.26. These structural comparisons highlight the significant conservation of KT250635 across multiple plant species, reinforcing its potential as a key resistance gene.

Figure 2

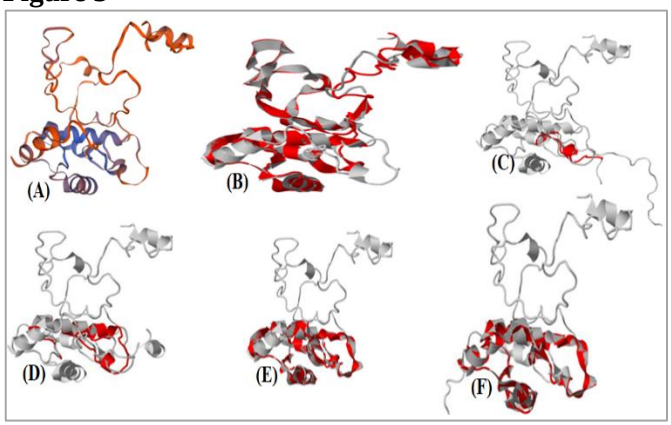
(A) *3D structure of KT250635;* ***(B)*** *Superposition result between KT250635 and PHAAN32355,* ***(C)*** *Superposition result between KT250635 and SOYBN02273,* ***(D)*** *Superposition result between KT250635 and GOSHI67843,* ***(E)*** *Superposition result between KT250635 and VITVI13198,* ***(F)*** *Superposition result between KT250635 and BRANA51611*



Functional Analysis of KT633945 Geen Across Orthologous Crop Species

The resistance gene analogue AN (KT633945) is highly conserved among multiple plant species, including *Ramandii GOSQ37FT49*, *Chili pepper CHEQ16HYQ6*, *Theobroma cacao AOA067SMZ2*, *Barley HORVVK385*, and *Okra OKACP56476*, based on hypothetical values. The 3D structure of KT633945, shown in **Figure 3(A),** was modeled using the TMV resistance protein N-like template (60OW.1.A) with a sequence identity of 53.28%. The structure was determined using the experimental X-ray approach, with no ligands present. Structural superposition analyses using the FATCAT Server revealed significant similarities with most orthologs. Superposition with *GOSQ37FT49* (**Figure 3B**) showed a P-value of 0.00e+00, 150 equivalent positions, and an RMSD of 1.68. The alignment of CHEQ16HYQ6 (**Figure 3C**) was not significantly similar, with a P-value of 8.66e-01, 16 equivalent positions, and an RMSD of 1.95. Alignment with *HORVVK385* (**Figure 3D**) yielded a P-value of 2.36e-02, 39 equivalent positions, and an RMSD of 2.17. The structures for OKACP56476 (**Figure 3E**) were significantly similar, with a P-value of 0.00e+00, 91 equivalent positions, and an RMSD of 0.17. Lastly, superposition with *AOA067SMZ2* (**Figure 3F**) demonstrated significant similarity with a P-value of 1.89e-15, 77 equivalent positions, and an RMSD of 0.46. These findings highlight the structural conservation of KT633945 across diverse plant species, with variations in alignment quality and significance among different orthologs.

Figure 3

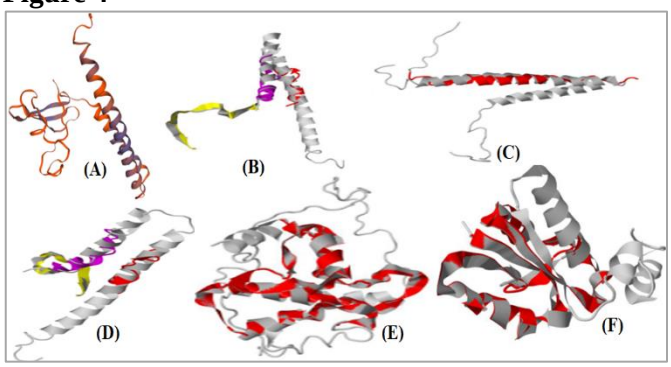


Structural and Functional Analysis of Resistance Gene Analogue AN (KT633945) Across Orthologous Crop Species. **Figure 3(A):** 3D structure of KT633945 modeled using TMV resistance protein N-like (60OW.1.A). **Figure 3(B):** Structural superposition of KT633945 with *GOSQ37FT49* showing high similarity (P-value: 0.00e+00, RMSD: 1.68). **Figure 3(C):** Structural superposition of KT633945 with *CHEQ16HYQ6* indicating no significant similarity (P-value: 8.66e-01, RMSD: 1.95). **Figure 3(D):** Superposition of KT633945 with *HORVVK385* showing moderate similarity (P-value: 2.36e-02, RMSD: 2.17). **Figure 3€:** Structural alignment of KT633945 with *OKACP56476* demonstrating high similarity (P-value: **0.00e+00**, RMSD: **0.17**). **Figure 3(F):** Superposition of KT633945 with *AOA067SMZ2* showing significant structural similarity (P-value: 1.89e-15, RMSD: 0.46).

Structural Analysis and Comparative Superposition of Resistance Gene Analogue KT633946 with Orthologous Crop Proteins

Resistance gene analogue AN (KT633946) is highly conserved among orthologous crop species, including Triticum aestivum (WHEAT04144), Oryza rufipogon (ORYRU26699), Raimondii (GOSRA37917), Cucumber (CUCSA16981), and Wild legume (LOTJA19348), based on hypothetical values. **Figure 4(A)** illustrates the 3D structure of KT633946, modeled with 103 available features and identified as TMV resistance protein N-like (50CL.1.A) with a target-template sequence identity of 42.67%.

Figure 4



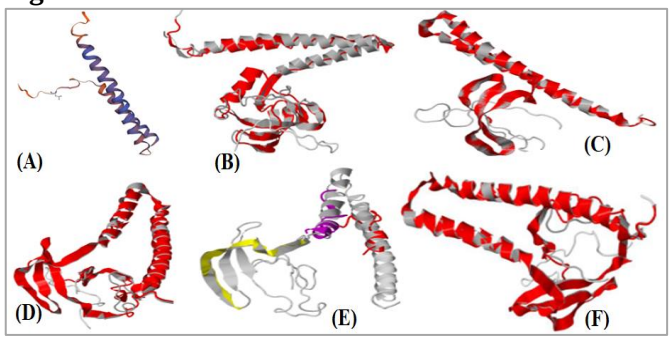
Structural Analysis and Comparative Superposition of Resistance Gene Analogue KT633946 with Orthologous Crop Proteins. **Figure 4(A):** 3D structure of KT633946 modeled with 103 features as TMV resistance protein N-like (50CL.1.A). **Figure 4(B):** Superposition of KT633946 and CUCSA16981 with high similarity (P-value: 0.00e+00, RMSD: 1.68). **Figure 4(C):** Superposition of KT633946 and ORYRU26699 showing significant similarity (P-value: 1.43e-03, RMSD: 1.87). **Figure 4(D):** Superposition of KT633946 and GOSRA37917 showing no significant similarity (P-value: 7.69e-02, RMSD: 2.31, with two twists). **Figure 4(E):** Superposition of KT633946 and WHEAT04144 with strong similarity (P-value: 0.00e+00, RMSD: 0.06). **Figure 4(F):** Superposition of KT633946 and LOTJA19348 with significant similarity (P-value: 0.00e+00, RMSD: 0.64).

The experimental structure, determined via X-ray crystallography, contains no ligands. Structural superposition using the FATCAT server reveals significant similarities between KT633946 and orthologs, with key metrics including Probability, RMSD, and Equivalent Positions. **Figure 4(B)** shows a high similarity between KT633946 and CUCSA16981, with a P-value of 0.00e+00, 150 equivalent positions, and an RMSD of 1.68. Similarly, **Figure 4(C)** highlights significant similarity with ORYRU26699 (P-value: 1.43e-03, 36 equivalent positions, RMSD: 1.87). However, **Figure 4(D)** demonstrates no significant similarity with GOSRA37917 (P-value: 7.69e-02, 49 equivalent positions, RMSD: 2.31, with two twists). **Figure 4(E)** depicts a strong similarity with WHEAT04144 (P-value: 0.00e+00, 104 equivalent positions, RMSD: 0.06), while **Figure 4(F)** shows significant similarity with LOTJA19348 (P-value: 0.00e+00, 84 equivalent positions, RMSD: 0.64).

Structural Analysis and Comparative Modeling of Resistance Gene Analogue AN (KT885194) with Orthologous Plant Proteins

The resistance gene analogue AN (KT885194) shows high conservation with orthologous proteins from Peach (M5X0Q3), Annual-bunch-grass (ERATE04813), Black-Cotton-wood (POPTR06525), Flowering Plant (ARATH04660), and Brassica rapa (BRARP33976), as identified through hypothetical values. **Figure 5(A)** presents the 3D structure of KT885194, modeled as TMV resistance protein N-like (5ncm.1.B) with a target-template sequence identity of 34.48% and zero ligands in the experimental structure. 3D superposition analyses using the FATCAT server reveal structural similarities based on Probability, RMSD, and Equivalent Positions. **Figure 5(B)** highlights significant similarity between KT885194 and M5X0Q3 (P-value: 1.05e-08, RMSD: 3.19, 131 equivalent positions). **Figure 5(C)** shows KT885194 and BRARP33976 are significantly similar (P-value: 0.00e+00, RMSD: 0.05, 108 equivalent positions). **Figure 5(D)** indicates significant similarity with ARATH04660 (P-value: 1.55e-15, RMSD: 2.59, 148 equivalent positions). **Figure 5(E)** shows no significant similarity with POPTR06525 (P-value: 1.87e-01, RMSD: 2.31, 49 equivalent positions, with 2 twists). Finally, **Figure 5(F)** demonstrates significant similarity with ERATE04813 (P-value: 1.55e-15, RMSD: 2.59, 148 equivalent positions).

Figure 5

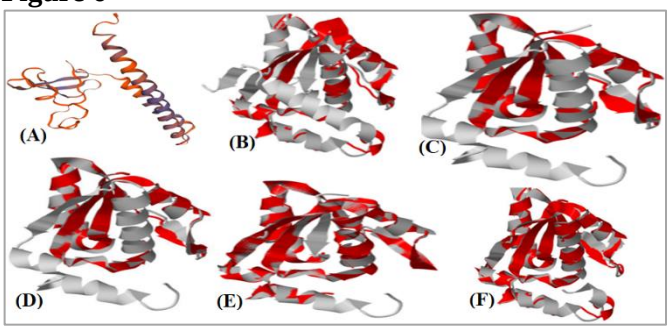


The predicted 3D structure of KT885194 (**Figure 5A**) modeled as TMV resistance protein N-like (5ncm.1.B) showed significant structural similarity with orthologs, including Peach M5X0Q3 (**Figure 5B**, P-value: 1.05e-08, RMSD: 3.19), Brassica rapa BRARP33976 (**Figure 5C**, P-value: 0.00e+00, RMSD: 0.05), Arabidopsis thaliana ARATH04660 (**Figure 5D**, P-value: 1.55e-15, RMSD: 2.59), and Annual-bunch-grass ERATE04813 (**Figure 5F**, P-value: 1.55e-15, RMSD: 2.59). However, no significant similarity was observed with Black-Cotton-wood POPTR06525 (**Figure 5E**, P-value: 1.87e-01, RMSD: 2.31, with 2 twists).

Structural and Functional Analysis of Resistance Gene Analogue AN (KT886994) and Its Conservation Across Orthologs

The resistance gene analogue AN (KT886994) exhibits high conservation with **Rice** (ORYGL20431), Gossypium raimondii (GOSRA37917), Zea mays (MAIZE16090), Sorghum (SORBI40580), and Sunflower (HELAN16704), based on hypothetical values. **Figure 6(A)** depicts the 3D structure of KT886994, modeled using the TMV resistance protein N-like (5ncl.1.A) template with 46.47% identity. The X-ray structure determination revealed no ligands. 3D superposition analyses using FATCAT Server showed structural similarities between KT886994 and orthologous proteins. **Figure 6(B)** illustrates significant similarity with ORYGL20431 (P-value: 1.11e-16, RMSD: 0.97, 96 equivalent positions). **Figure 6(C)** shows KT886994-GOSRA37917 similarity (P-value: 0.00e+00, RMSD: 0.64, 74 equivalent positions). **Figure 6(D)** demonstrates similarity with MAIZE16090 (P-value: 1.11e-16, RMSD: 0.62, 70 equivalent positions). **Figure 6(E)** highlights KT886994-SORBI40580 alignment (P-value: 0.00e+00, RMSD: 0.03, 118 equivalent positions). Finally, **Figure 6(F)** shows strong similarity with HELAN16704 (P-value: 0.00e+00, RMSD: 0.99, 120 equivalent positions).

Figure 6

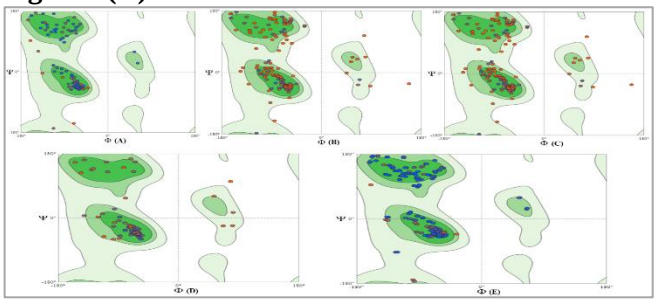


The resistance gene analogue AN (KT886994) is conserved across orthologs: **Figure 6(A)** shows its 3D structure modeled with 46.47% identity. Structural superposition reveals significant similarity with ORYGL20431 **(Figure 6(B)),** GOSRA37917 (**Figure 6(C)),** MAIZE16090 **(Figure 6(D)),** SORBI40580 **(Figure 6(E)),** and HELAN16704 **(Figure 6(F)).**

Ramachandran Plot Analysis of Protein Models

The Ramachandran plot analysis of the models highlights their stereochemical quality and structural reliability. **Figure 8(A)** shows the Ramachandran plot of KT250635, which contains 242 residues, with 92.6% in the most favored regions, 7.4% in additional allowed regions, and none in disallowed regions, demonstrating excellent model quality. KT633945, depicted in **Figure 8(B)**, has 191 residues, with 81.4% in the most favored regions, 15.4% in additional allowed regions, and 1.2% in disallowed regions, indicating minor deviations. **Figure 8(C)** shows KT633946, with 150 residues, 82.4% in the most favored regions, 11% in additional allowed regions, and 2.1% in disallowed regions, suggesting potential refinement needs. **Figure 8(D)** illustrates KT885194, containing 91 residues, 86.4% in the most favored regions, 8.4% in additional allowed regions, and no disallowed residues, highlighting strong structural reliability. Lastly, **Figure 8(E)** presents KT886994, identical in quality to KT250635, with 92.6% of 242 residues in the most favored regions and no residues in disallowed regions, confirming its excellent stereochemical integrity. These results emphasize the high quality of the models, with KT250635 and KT886994 achieving the best outcomes, while KT633945 and KT633946 may benefit from minor structural optimization.

Figure 8(A)



Ramachandran Plot Result of Model crop KT250635; **Fig 8(B):** Ramachandran Plot Result of Stand crop KT633945; **Fig 8(C):** Ramachandran Plot Result of Model crop KT633946; **Fig 8(D):** Ramachandran Plot Result of Model crop KT885194; **Fig 8(E):** Ramachandran Plot Result of Model crop KT886994

Evaluation of GMQE and QMEAN Scores for TMV Resistance Protein Structural Models

The Global Model Quality Estimation (GMQE) is a predictive measure that combines target-template alignment properties with the template search process to provide a quality score between 0 and 1, reflecting the expected accuracy of the resulting 3D structure. For the TMV resistance protein (KT250635), the GMQE score was 0.77, with individual features including Cβ atoms (-0.77), all atoms (-0.71), solvation potential (0.24), and torsion angle potential (-0.75). This protein exhibited a monomeric oligo state with 59.02% sequence identity, and its residue score of 0.90 indicates good model quality (**Table 1A**). Similarly, KT633945 demonstrated a GMQE score of 0.79, with individual characteristics such as Cβ atoms (-0.19), all atoms (-0.63), solvation potential (-0.50), and torsion angle potential (-0.23). With 53.28% sequence identity and a residue score of 0.86, its 3D structure is also of good quality (**Table 1B**). For KT633946, the GMQE score was 0.57, with individual features including Cβ atoms (-2.41), all atoms (-2.31), solvation potential (-1.17), and torsion angle potential (-3.79). Despite a lower sequence identity of 42.67%, its residue score of 0.68 suggests acceptable quality (**Table 1C**). KT885194 presented a GMQE score of 0.48, with characteristics such as Cβ atoms (-1.16), all atoms (2.87), solvation potential (1.46), and torsion angle potential (-1.64). Its sequence identity was 34.48%, and the residue score of 0.74 reflects good quality (**Table 1D**). Finally, KT886994 achieved a GMQE score of 0.57, with individual parameters including Cβ atoms (-0.43), all atoms (-0.49), solvation potential (-0.53), and torsion angle potential (-3.26). With a sequence identity of 46.47% and a residue score of 0.40, its 3D structure is considered good quality (**Table 1E**). For all these proteins, QMEAN Z-scores, which evaluate structural agreement with experimental data, were consistent with expected ranges, indicating reliable model structures despite variations in individual features.

Table 1(A)

Subject crops 3D Structures against Resistance gene analogue KT250635

Cβ atom	Torsion angle potential	MolProbity - score	R_Chandran score	LC - score	QMEAN-score	GMQE-score	Compound ID	Sr.no
-1.47	-0.69	1.07	93.07%	0.75	-0.82	0.71	GOSHI67843	1
-0.69	1.07	1.35	91.07%	0.61	1.3843	0.535	PHAAN32355	2
-0.59	1.27	1.25	87.07%	0.71	2.2298	0.9858	BRANA51611	3
-0.99	1.07	1.05	94.07%	0.81	0.8953	0.0283	VITVI13198	4
-0.76	0.07	1.30	78.07%	0.71	1.8963	0.4710	SOYBN02273	5

Table 1(B)

Subject crops 3D Structures against Resistance gene analogue KT633945

Sr.no	Compound ID	Compound 3-DStructure	GMQE -score	QMEAN-score	LC - score	R_Chandran score	MolProbity – score	Torsion angle potential	Cβ atoms
1	GOSQ37FT49		0.71	-0.82	0.75	93.07%	1.07	-0.69	-1.47
2	CHEQ16HYQ6		0.535	1.3843	0.61	91.07%	1.35	1.07	-0.69
3	HORVVK385		0.9858	2.2298	0.71	87.07%	1.25	1.27	-0.59
4	OKACP56476		0.0283	0.8953	0.81	94.07%	1.05	1.07	-0.99
5	AOA067SMZ2		0.4710	1.8963	0.71	78.07%	1.30	0.07	-0.76

Table 1(C)

Subject crops 3D Structures against Resistance gene analogue KT633946

Sr.no	Compound ID	Compound 3-DStructure	GMQE-score	QMEAN-score	LC - score	R_Chandran score	MolProbity - score	Torsion angle potential	Cβ atoms
1	WHEAT04144		0.71	-0.82	0.75	93.07%	1.07	-0.69	-1.47
2	ORYRU26699		0.535	1.3843	0.61	91.07%	1.35	1.07	-0.69
3	GOSRA37917		0.9858	2.2298	0.71	87.07%	1.25	1.27	-0.59
4	CUCSA16981		0.0283	0.8953	0.81	94.07%	1.05	1.07	-0.99
5	LOTJA19348		0.4710	1.8963	0.71	78.07%	1.30	0.07	-0.76

Table 1D

Subject crops 3D Structures against Resistance gene analogue KT885194

Sr.no	Compound ID	Compound 3-DStructure	GMQE -score	QMEAN-score	LC - score	R_Chandran score	MolProbity - score	Torsion angle potential	Cβ atoms
1	M5X0Q3		0.71	-0.82	0.75	93.07%	1.07	-0.69	-1.47
2	ERATE04813		0.535	1.3843	0.61	91.07%	1.35	1.07	-0.69

3	POPTR06525		0.9858	2.2298	0.71	87.07%	1.25	1.27	-0.59
4	ARATH04660		0.0283	0.8953	0.81	94.07%	1.05	1.07	-0.99
5	BRARP33976		0.4710	1.8963	0.71	78.07%	1.30	0.07	-0.76

Table 1(E)

Subject crops 3D Structures against Resistance gene analogue KT886994

Sr.no	Compound ID	Compound 3-DStructure	GMQE -score	QMEAN-score	LC - score	R_Chandran score	MolProbity - score	Torsion angle potential	Cβ atoms
1	ORYGL20431		0.71	-0.82	0.75	93.07%	1.07	-0.69	-1.47
2	GOSRA37917		0.535	1.3843	0.61	91.07%	1.35	1.07	-0.69
3	MAIZE16090		0.9858	2.2298	0.71	87.07%	1.25	1.27	-0.59
4	SORBI40580		0.0283	0.8953	0.81	94.07%	1.05	1.07	-0.99
5	HELAN16704		0.4710	1.8963	0.71	78.07%	1.30	0.07	-0.76

Identification of Orthologs of TMV Resistance Protein Across Diverse Crop Species

The analysis of percent identity between the TMV resistance protein structural models across various crop IDs reveals significant variability in genetic similarity. KT250635 shows the highest percent identity of 100% across all its pairings, indicating a substantial similarity within its sequence. Notably, it shares a percent identity of 92.1% with the mustard plant (BRANA51611) and 87.2% with Mexican cotton (GOSHI67843), highlighting its potential functional conservation among these crops. In contrast, KT633945 demonstrates substantial identity with the mustard plant (BRANA51611) at 43.1%, and chili pepper (CHEQ16HYQ6) at 67.1%, showcasing its broader applicability across diverse plant species.

Table 2

Percent identity of TMV resistance protein structural models among various crop species, illustrating genetic similarity and functional conservation.

KT250635		KT633945		KT633946		KT885194		KT886994	
Crops ID	Percent Identity	Crops Id	Percent identity	Crops ID	Percent Identity	Crops ID	Percent Identity	Crops ID	Percent Identity
KT250635	100.0	KT633945	100.0	KT633946	100.0	KT885194	100.0	KT886994	100.00
PHYPA29675	38.3	Ramandii GOSQ37FT49	98.3	Triticum aestivumWHEAT04144	92.3	M7YKX1	18.3	A0A061G7R7	18.3
CUCSA17286	28.7	Q6T3R3	28.7	GRENI02923	28.7	C5Z0U8	28.7	Rice (ORYGL20431)	68.7
Red mung bean-PHAAN32355	77.1	Chili pepper CHEQ16HYQ6	67.1	Oryza rufipogon ORYRU26699	76.1	B8B015	37.1	Gossypium raimondii(GOSRA37917)	97.1
LUPAN05741	27.5	M1BZB1	27.5	Raimondii GOSRA37917	87.5	Q6L535	27.5	Zea mays (MAIZE16090)	77.5
MANES30688	33.6	Theobroma cacao	83.6	Cucumber CUCSA16981	69.6	B8B015	23.6	CHEQI13766	33.6

		AOA067SMZ2							
Mustard plant-BRANA51611	92.1	A0A0A0LJJ2	43.1	Wild legume LOTJA19348	71.1	C5Z0U8	43.1	jowari/sorghum (SORBI40580)	93.1
ARALY07552	12.6	K7MHM5	12.6	DIPOR03590	12.6	Peach M5X0Q3	72.6	sunflower (HELAN16704)	72.6
Mexican cotton-GOSHI67843	87.2	K7KD03	18.2	URSAM22988	48.2	ORYSJ23898	48.2	HORVV92028	18.2
Grape vine-VITVI13198	62.8	Barley HORVVK385	72.8	TAEGU13603	22.8	BRADI09786	22.8	OKACL03825	22.8
HELAN29816	37.7	M0ZJY1	37.7	ANOCA17335	37.7	Annual bunch grassERATE04813	67.7	CHILI01316	37.7
SOLLC11422	24.7	M1D406	24.7	LOALO01389	24.7	SORBI40580	24.7	WHEAT04144	24.7
SOLTU00656	28.9	D7SMZ2	56.9	DANPL00369	28.9	Black Cottonwood POPTR06525	71.9	MUSAM16396	28.9
LOTJA02677	17.4	Okra OKACP56476	87.4	DROBM13224	17.4	flowering plant ARATH04660	77.4	TRIUA22660	17.4
Soya bean-SOYBN02273	65.8	K7MWR8	17.8	TRIVA33048	47.8	Brassica rapa BRARP33976	91.8	VITVI22073	27.8

Similarly, KT633946 aligns closely with Raimondii GOSRA37917 at 87.5% and wild legumes (LOTJA19348) at 71.1%. Furthermore, KT885194 shows notable identity with peach (M5X0Q3) at 72.6% and flowering plants (ARATH04660) at 77.4%, suggesting its potential role in resistance mechanisms. Finally, KT886994 aligns strongly with sorghum (SORBI40580) at 93.1% and cotton (Gossypium raimondii GOSRA37917) at 97.1%, indicating its significant contribution to resistance traits in these species. The detailed percent identity results presented in **Table 2** underscore the genetic variability and functional diversity among these TMV resistance proteins across different crops.

Phylogenetic Analysis of Selected Sequences Using Neighbor-Joining Method

The phylogenetic relationships of the sequences KT250635, KT633945, KT633946, KT885194, and KT886994 were evaluated using the Neighbor-Joining method, which provides insights into their evolutionary connections. For KT250635, the tree construction yielded a total branch length of 5.26568982, as shown in **Figure 7(A)**. The tree was based on 15 amino acid sequences, and after removing all incomplete and gap-containing data, 68 positions were analyzed. The evolutionary distances were calculated using the Poisson correction method, which determines the number of amino acid replacements per site. The reliability of the tree topology was supported by a bootstrap test with 1000 replicates, indicating strong clustering consistency within the divisions. Similarly, the phylogenetic tree of KT633945, depicted in **Figure 7(B)**, showed a total branch length of 3.79758017. This analysis also included 15 amino acid sequences, with 62 positions retained in the final dataset after preprocessing. The evolutionary distances followed the same Poisson correction method, ensuring accurate distance measurement. The bootstrap analysis further confirmed the robustness of the tree, providing a high degree of confidence in the observed clustering patterns among taxa.

For KT633946, the phylogenetic tree is presented in **Figure 7(C)** and displays a total branch length of 7.11915971. The analysis involved the same set of 15 sequences, but due to the differences in alignment and data preprocessing, 112 positions were included in the final dataset. Like the others, this tree was drawn to scale, reflecting accurate evolutionary distances. With 1000 replicates, the bootstrap analysis supported the consistent grouping of taxa, highlighting the evolutionary relationships within this sequence cluster.

Figure 7

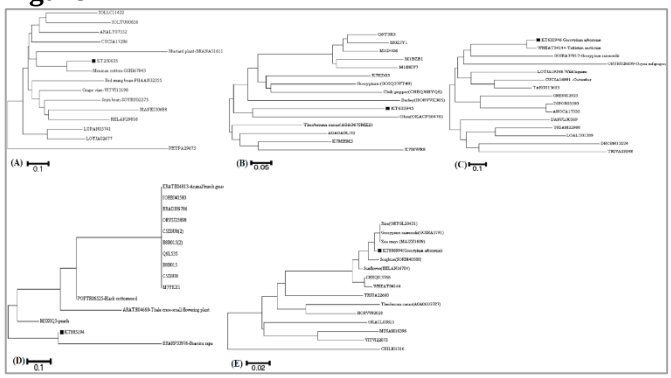


Figure 7(A): Phylogenetic tree of Orthologs crops with KT250635 crop. **Figure 7(B):** Phylogenetic tree of Orthologs crops with KT633945 crop. **Figure 7(C):** Phylogenetic tree of Orthologs crops with KT633946 crop. **Figure 7(D):** Phylogenetic tree of Orthologs crops with KT885194 crop. **Figure 7(E):** Phylogenetic tree of Orthologs crops with KT886994 cropThe analysis of KT885194, shown in **Figure 7(D)**, revealed a shorter total branch length of 1.66935617. The evolutionary distances were similarly computed using the Poisson correction method, with the same set of 15 sequences contributing to the tree construction. The dataset for KT885194 included positions after the removal of incomplete data, ensuring reliable results. The bootstrap test further validated the divisions within the tree, reflecting the evolutionary closeness of taxa within this group.

Finally, the phylogenetic tree of KT886994, illustrated in **Figure 7(E)**, showed a total branch length of 0.73077724, the shortest among the analyzed sequences. The tree was based on 15 sequences, and the final dataset contained 146 positions after preprocessing. The evolutionary distances, measured using the Poisson correction, accurately depicted amino acid replacements per site. The bootstrap test with 1000 replicates reinforced the reliability of the tree's topology, demonstrating strong consistency in clustering. Overall, the phylogenetic analyses highlight the evolutionary relationships and divergence among the sequences KT250635, KT633945, KT633946, KT885194, and KT886994. The consistent application of the Neighbor-Joining method, Poisson correction, and robust bootstrap validation provides reliable insights into the amino acid sequence variations and evolutionary patterns across the taxa. The findings, as visualized in **Figures 7(A)–7(E)**, emphasize the evolutionary distinctions and conserved regions within these crop-related sequences.

Highly Leading Orthologs Crops

The most prominent RGAs in orthologs crops are BRANA51611, OKACP56476, WHEAT04144, BRARP33976 and SORBI40580. These RGAs can be identified from the corresponding genomes using the methods of bioinformatics Applications to supplement the efficient mapping and co-expression of plant resistance genes. Resulted orthologs crops help confer resistance against cell wall-degrading phytopathogens like fungi, bacteria and nematodes. It regulates the extent of pathogen containment, hypersensitive plant cell death and oxidative burst at primary infection sites. These crops (Soya bean, sorghum, Okra, Black Cottonwood and Cucumber) genes are known to be important players in plant defence, especially in the context of causing programmed cell death in infected tissues.

Table 3

Leading RGA in Orthologs Crops

S.no	Stand Crop ID	Orthologs Crop ID	P-value	R-S	E-P	RMSD
1	KT250635	BRANA51611	0.00e+00	286	103	0.26
2	KT633945	OKACP56476	0.00e+00	264	91	0.17
3	KT633946	WHEAT04144	0.00e+00	272	104	0.06
4	KT885194	BRARP33976	0.00e+00	312	108	0.05
5	KT886994	SORBI40580	0.00e+00	288	118	0.03

DISCUSSION

Begomoviruses, belonging to the Geminiviridae family, are significant plant pathogens affecting a broad range of dicotyledonous plants and are perpetuated through the whitefly *Bemisia tabaci*. Among these, Cotton Leaf Curl Virus (CLCuV) is a major limiting factor in cotton production, particularly in *Gossypium hirsutum*. In this study, we utilized bioinformatics tools to identify resistance gene analogs (RGAs) in cotton that are potentially associated with CLCuV resistance. These RGAs were characterized and compared across orthologous crops, allowing us to identify key candidate crops and genes that can be further explored for improving resistance in cotton.

Our findings are consistent with earlier studies that emphasize the critical role of RGAs in plant defense mechanisms. For instance, Ijaz, et al. [31] identified the genetic diversity of RGAs in cotton and their potential to confer resistance against viral pathogens. Similarly, our results demonstrate that RGAs with accession numbers KT250635, KT886994, KT633945, KT885194, and KT633946 showed high homology with resistance genes in crops such as mustard (*Brassica napus*), sorghum (*Sorghum bicolor*), okra (*Abelmoschus esculentus*), black cottonwood (*Populus trichocarpa*), and wheat (*Triticum aestivum*). These crops have been highlighted in previous studies Ali, et al. [32], for their robust defense mechanisms against various biotic stresses.

A study by Verma, et al. [33] demonstrated that integrating bioinformatics tools, such as SWISS-MODEL and FATCAT, to analyze RGAs provides valuable insights into structural and functional domains critical for plant defense. Our approach similarly used these tools to model and superimpose the 3D structures of RGAs, identifying key structural properties such as probability scores, RMSD, and equivalent positions. These findings validate the use of computational modeling in resistance gene discovery, as emphasized by Chandramohan and Reports [34].

Moreover, a study by Duplessis, et al. [35], highlighted the importance of leveraging orthologous comparisons for identifying conserved defense genes across species. Our work extends this approach by mapping RGAs across multiple orthologous crops and selecting genes with high percent identity and conserved domains. For instance, KT250635 showed high similarity with defense-related genes in mustard and

grapevine, corroborating the findings of Marastoni, et al. [36] on the evolutionary conservation of plant defense mechanisms.

Interestingly, our study also aligns with recent advances in gene editing technologies, such as CRISPR-Cas9, which have been used to enhance RGA function in crop species. A review by Zambounis, et al. [37] emphasized the potential of combining bioinformatics and gene editing for improving crop resistance. The RGAs identified in our study could serve as candidates for such approaches, particularly in developing resistant varieties of *Gossypium hirsutum*.

In contrast to previous studies, our work provides a more comprehensive analysis by integrating multiple bioinformatics tools and databases. While earlier studies Adams, et al. [38] focused on mapping and co-expression of RGAs, our study incorporates structural refinement and superimposition analyses to provide deeper insights into the functional aspects of these genes. Additionally, the identification of specific crops such as sorghum and okra as potential RGA sources adds a novel dimension to existing research.

Despite these advancements, certain limitations remain. For example, our study relied heavily on computational predictions, and experimental validation of these RGAs in *Gossypium hirsutum* and other crops is necessary to confirm their role in resistance. Furthermore, as highlighted by Patil, et al. [39], environmental factors and pathogen variability can influence the effectiveness of resistance genes, which warrants further investigation.

CONCLUSION

The genus *Begomovirus*, belonging to the family *Geminiviridae*, encompasses plant viruses that infect dicotyledonous plants and cause significant economic damage to important crops, including cotton. The cotton leaf curl virus (CLCuV) is a major constraint in cotton production, particularly affecting *Gossypium hirsutum*. This study explored resistance-related genes in cotton against CLCuV, identifying five putative resistance genes: KT250635, KT886994, KT633945, KT885194, and KT633946. The stereochemical quality of protein models was confirmed through Ramachandran plot analysis. KT250635 and KT886994 demonstrated excellent model quality, with 92.6% of residues in the most favored regions and no residues in disallowed regions. Other models showed minor deviations: KT633945 (81.4% favored, 1.2% disallowed), KT633946 (82.4% favored, 2.1% disallowed), and KT885194 (86.4% favored, 0% disallowed). These results underscore the structural reliability of the models, with KT250635 and KT886994 achieving the highest quality. The GMQE scores further validated the models, with KT250635 scoring 0.77 and KT886994 scoring 0.57, reflecting high-quality predictions. KT633945 and KT885194 scored 0.79 and 0.48, respectively, while KT633946 scored 0.57, requiring refinement. Sequence identity ranged from 34.48% (KT885194) to 59.02% (KT250635), with QMEAN Z-scores indicating reliable structural models across all proteins. Phylogenetic analysis revealed evolutionary relationships, with KT250635 showing a total branch length of 5.27 and KT886994 aligning strongly with sorghum (93.1%) and cotton (*Gossypium raimondii*, 97.1%). Similarly, KT633945 and KT885194 exhibited evolutionary links with other plant species, such as peach and wild legumes, demonstrating genetic variability and functional diversity. Bootstrap tests with 1000 replicates ensured robustness in phylogenetic clustering. Overall, this study identified critical resistance gene analogs (RGAs) and provided structural and functional insights into their role in combating CLCuV. These findings contribute to understanding resistance mechanisms, paving the way for genetic improvement of vulnerable cotton varieties and potentially enhancing crop resilience against viral pathogens. Further studies are needed to explore the association of these genes with resistance pathways and their applicability in other crop species.

REFERENCES

[1] Fauquet, C. M., Bisaro, D. M., Briddon, R. W., Brown, J. K., Harrison, B. D., Rybicki, E. P., Stenger, D. C., & Stanley (Study Group Chair), J. (2003). Virology division news : Revision of taxonomic criteria for species demarcation in the family Geminiviridae , and an updated list of begomovirus species. *Archives of Virology*, *148*(2), 405–421. https://doi.org/10.1007/s00705-002-0957-5

[2] Lozano, G., Trenado, H. P., Fiallo-Olivé, E., Chirinos, D., Geraud-Pouey, F., Briddon, R. W., & Navas-Castillo, J. (2016). Characterization of non-coding DNA satellites associated with Sweepoviruses (Genus Begomovirus, Geminiviridae) – Definition of a distinct class of begomovirus-associated satellites. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00162

[3] Smith, A. M. (2018). Host-pathogen kinetics during influenza infection and coinfection: Insights from predictive

modeling. *Immunological Reviews*, *285*(1), 97-112. https://doi.org/10.1111/imr.12692

[4] Sun, K., Fu, K., Hu, T., Shentu, X., & Yu, X. (2023). Leveraging insect viruses and genetic manipulation for sustainable agricultural pest control. *Pest Management Science*, *80*(6), 2515-2527. https://doi.org/10.1002/ps.7878

[5] Whitham, S. A., & Hajimorad, M. R. (2016). Plant genetic resistance to viruses. *Current Research Topics in Plant Virology*, 87-111. https://doi.org/10.1007/978-3-319-32919-2_4

[6] Biswas, K. K., Tarafdar, A., & Biswas, K. (2012). Viral diseases and its mixed infection in mungbean and urd bean: major biotic constraints in production of food pulses in India. *Modern trends in microbial biodiversity of natural ecosystem. biotech books*, 301-319.

[7] Gökçe, A. F., Chaudhry, U. K., & Junaid, M. D. (2021). Mapping QTLs for abiotic stress. *Developing Climate-Resilient Crops*, 175-201. https://doi.org/10.1201/9781003109037-9-9

[8] Yasmin, F. (2023). *Investigating the Genetic Basis and Selection of Diverse Plant Specialized Metabolites in Wild Soybean, Glycine soja* (Doctoral dissertation, The University of North Carolina at Charlotte).

[9] Notredame, C. (2002). Recent progress in multiple sequence alignment: A survey. *Pharmacogenomics*, *3*(1), 131-144. https://doi.org/10.1517/14622416.3.1.131

[10] Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., Peters, R. S., Podsiadlowski, L., Vasilikopoulos, A., Zhou, X., Misof, B., & Niehuis, O. (2017). Orthograph: A versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics*, *18*(1). https://doi.org/10.1186/s12859-017-1529-8

[11] Altenhoff, A. M., Glover, N. M., Train, C., Kaleb, K., Warwick Vesztrocy, A., Dylus, D., De Farias, T. M., Zile, K., Stevenson, C., Long, J., Redestig, H., Gonnet, G. H., & Dessimoz, C. (2017). The OMA orthology database in 2018: Retrieving evolutionary relationships among all domains of life through Richer web and programmatic interfaces. *Nucleic Acids Research*, *46*(D1), D477-D485. https://doi.org/10.1093/nar/gkx1019

[12] Hassani-Pak, K., Singh, A., Brandizi, M., Hearnshaw, J., Amberkar, S., Phillips, A. L., Doonan, J. H., & Rawlings, C. (2020). KnetMiner: A comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. https://doi.org/10.1101/2020.04.02.017004

[13] Mushtaq, R., Shahzad, K., Mansoor, S., Shah, Z. H., Alsamadany, H., Mujtaba, T., Al-Zahrani, Y., Alzahrani, H. A., Ahmed, Z., & Bashir, A. (2018). Exploration of cotton leaf curl virus (CLCuV) resistance genes and their screening in gossypium arboreum by targeting resistance gene analogues. *AoB PLANTS*. https://doi.org/10.1093/aobpla/ply067

[14] Sodha, D., Verma, S. K., Chhokar, V., & Paul, D. (2022). Cotton lead curl viral disease in American cotton (G. Hirsutum): genetic basis of resistance and role of genetic engineering tools in combating CLCUD. *Scientist*, *1*(3), 5122-5137.

[15] Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Federhen, S., Feolo, M., Fingerman, I. M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Lu, Z., Madden, T. L., … Ye, J. (2010). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *39*(Database), D38-D51. https://doi.org/10.1093/nar/gkq1172

[16] Arshad, I., Ahsan, M., Zafar, I., Sajid, M., Sehgal, S. A., Yousaf, W., Noor, A., Rashid, S., Garai, S., Moovendhan, M., & Sharma, R. (2023). Bioinformatics approaches in upgrading microalgal oil for advanced biofuel production through hybrid ORF protein construction. *Biomass Conversion and*

Biorefinery. https://doi.org/10.1007/s13399-023-04766-w

[17] Harrison, N., & Kidner, C. A. (2011). Next–generation sequencing and systematics: What can a billion base pairs of DNA sequence data do for you? *TAXON*, *60*(6), 1552-1566. https://doi.org/10.1002/tax.606002

[18] Altenhoff, A. M., Train, C., Gilbert, K. J., Mediratta, I., Mendes de Farias, T., Moi, D., Nevers, Y., Radoykova, H., Rossier, V., Warwick Vesztrocy, A., Glover, N. M., & Dessimoz, C. (2020). OMA orthology in 2021: Website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Research*, *49*(D1), D373-D379. https://doi.org/10.1093/nar/gkaa1007

[19] Su, W., Wang, L., Lei, J., Chai, S., Liu, Y., Yang, Y., Yang, X., & Jiao, C. (2017). Genome-wide assessment of population structure and genetic diversity and development of a core germplasm set for sweet potato based on specific length amplified fragment (SLAF) sequencing. *PLOS ONE*, *12*(2), e0172066. https://doi.org/10.1371/journal.pone.0172066

[20] Zafar, I., Tariq Pervez, M., Rather, M. A., Ellahi Babar, M., Ali Raza, M., Iftikhar, R., & Fatima, M. (2020). Genome-wide identification and expression analysis of PPOs and POX gene families in the selected plant species. *Biosciences Biotechnology Research Asia*, *17*(2), 301-318. https://doi.org/10.13005/bbra/2834

[21] Haider, H. I., Zafar, I., Ain, Q. U., Noreen, A., Nazir, A., Javed, R., Sehgal, S. A., Khan, A. A., Rahman, M. M., Rashid, S., Garai, S., & Sharma, R. (2022). Synthesis and characterization of copper oxide nanoparticles: Its influence on corn (Z. mays) and wheat (Triticum aestivum) plants by inoculation of bacillus subtilis. *Environmental Science and Pollution Research*, *30*(13), 37370-37385. https://doi.org/10.1007/s11356-022-24877-7

[22] Ahmad, H. M., Abrar, M., Izhar, O., Zafar, I., Rather, M. A., Alanazi, A. M., Malik, A., Rauf, A., Bhat, M. A., Wani, T. A., & Khan, A. A. (2022). Characterization of fenugreek and its natural compounds targeting AKT-1 protein in cancer: Pharmacophore, virtual screening, and MD simulation techniques. *Journal of King Saud University - Science*, *34*(6), 102186. https://doi.org/10.1016/j.jksus.2022.102186

[23] Ali, S., Noreen, A., Qamar, A., Zafar, I., Ain, Q., Nafidi, H., Bin Jardan, Y. A., Bourhia, M., Rashid, S., & Sharma, R. (2023). Amomum subulatum: A treasure trove of anti-cancer compounds targeting TP53 protein using in vitro and in silico techniques. *Frontiers in Chemistry*, *11*. https://doi.org/10.3389/fchem.2023.1174363

[24] Zafar, I., Anwar, S., Kanwal, F., Yousaf, W., Un Nisa, F., Kausar, T., Ul Ain, Q., Unar, A., Kamal, M. A., Rashid, S., Khan, K. A., & Sharma, R. (2023). Reviewing methods of deep learning for intelligent healthcare systems in genomics and biomedicine. *Biomedical Signal Processing and Control*, *86*, 105263. https://doi.org/10.1016/j.bspc.2023.105263

[25] Ye, Y., & Godzik, A. (2003). Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, *19*(suppl_2), ii246-ii255. https://doi.org/10.1093/bioinformatics/btg1086

[26] Gasteiger, E. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784-3788. https://doi.org/10.1093/nar/gkg563

[27] Zahn-Zabal, M., Dessimoz, C., & Glover, N. M. (2020). Identifying orthologs with OMA: A primer. *F1000Research*, *9*, 27. https://doi.org/10.12688/f1000research.21508.1

[28] Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*, *30*(5), 1229-1235. https://doi.org/10.1093/molbev/mst012

[29] Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2005). The SWISS-MODEL workspace: A web-based environment for

protein structure homology modelling. *Bioinformatics*, *22*(2), 195-201. https://doi.org/10.1093/bioinformatics/bti770

[30] Margelevičius, M. (2023). GTalign: Spatial index-driven protein structure alignment, superposition, and search. https://doi.org/10.1101/2023.12.18.572167

[31] Ijaz, S., Haq, I. U., Babar, M., & Nasir, B. (2022). Disease resistance genes' identification, cloning, and characterization in plants. *Cereal Diseases: Nanobiotechnological Approaches for Diagnosis and Management*, 249-269. https://doi.org/10.1007/978-981-19-3120-8_13

[32] Ali, J., Jan, I., Ullah, H., Fahad, S., Saud, S., Adnan, M., Ali, B., Liu, K., Harrison, M. T., Hassan, S., Kumar, S., Khan, M. A., Kamran, M., Alwahibi, M. S., & S. Elshikh, M. (2023). Biochemical response of okra (Abelmoschus esculentus L.) to selenium (Se) under drought stress. *Sustainability*, *15*(7), 5694. https://doi.org/10.3390/su15075694

[33] Verma, A., Niranjana, M., Jha, S. K., Mallick, N., Agarwal, P., & Vinod. (2020). QTL detection and putative candidate gene prediction for leaf rolling under moisture stress condition in wheat. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-75703-4

[34] Chandramohan, U. M. (2023). Computational biology of antibody epitope, tunnels and pores analysis of protein glutathione S-transferase P, and quantum mechanics. *Biochemistry and Biophysics Reports*, *36*, 101581. https://doi.org/10.1016/j.bbrep.2023.101581

[35] Duplessis, S., Major, I., Martin, F., & Séguin, A. (2009). Poplar and pathogen interactions: Insights from*populus*genome-wide analyses of resistance and defense gene families and gene expression profiling. *Critical Reviews in Plant Sciences*, *28*(5), 309-334. https://doi.org/10.1080/07352680903241063

[36] Marastoni, L., Sandri, M., Pii, Y., Valentinuzzi, F., Brunetto, G., Cesco, S., & Mimmo, T. (2019). Synergism and antagonisms between nutrients induced by copper toxicity in grapevine rootstocks: Monocropping vs. intercropping. *Chemosphere*, *214*, 563-578. https://doi.org/10.1016/j.chemosphere.2018.09.127

[37] Zambounis, A., Ganopoulos, I., Kalivas, A., Tsaftaris, A., & Madesis, P. (2016). Identification and evidence of positive selection upon resistance gene analogs in cotton (Gossypium hirsutum L.). *Physiology and Molecular Biology of Plants*, *22*(3), 415-421. https://doi.org/10.1007/s12298-016-0362-2

[38] Adams, C. M., Eckenroth, B. E., Putnam, E. E., Doublié, S., & Shen, A. (2013). Structural and functional analysis of the CspB protease required for clostridium spore germination. *PLoS Pathogens*, *9*(2), e1003165. https://doi.org/10.1371/journal.ppat.1003165

[39] Patil, A. M., Pawar, B. D., Wagh, S. G., Markad, N. R., Bhute, N. K., Shinde, H., Shelake, R. M., Červený, J., & Wagh, R. S. (2024). Unravelling abiotic stress physiology in cotton (Gossypium hirsutum L.): Biotechnological interventions in mitigating abiotic stress. https://doi.org/10.20944/preprints202405.0751.v1