Mysore J. Agric. Sci., 55 (4) : 355-363 (2021)

In-Silico Investigation of *Cucumber Mosaic Virus* (CMV) Resistance Genes and Transcription Factors Involved in CMV Hosts Generated from Chilli-CMV Transcriptome Data

SHRIDHAR HIREMATH[1], H. D. VINAYKUMAR[2], PRASANNA S. KOTI[3], M. NANDAN[4], K. S. SHANKRAPPA[5] AND C. N. LAKSHMINARAYANA REDDY[6]

[1,2,4&6]Department of Plant Pathology, College of Agriculture, UAS, GKVK, Bangalore - 560 065

[3]The University of Trans-Disciplinary Health Sciences and Technology, Bengaluru

[5]College of Horticulture, University of Horticultural Sciences, Bagalkot

e-Mail - shridharhi80@gmail.com

ABSTRACT

Cucumber mosaic virus (CMV) severely reduces crop productivity and is considered as a major global plant virus affecting important agricultural and ornamental crops. On infected crops, it causes severe disease symptoms such as mosaic, mosaic mottling, stunting and stem necrosis. Controlling disease propagation has become a serious challenge due to the advent of new strains of CMV, vector biotypes and an expanding host range. The use of resistant cultivars, on the other hand, is the greatest way to tackle this difficulty. In this regard, an attempt was undertaken to explore the previously available chilli-CMV transcriptome data for the presence of resistance (R) genes. Seven R genes were chosen and their physico-chemical characteristics were evaluated. Further, R genes were also extracted from CMV infected hosts using BLAST sequence similarity search tool. The diversity study of R gene proteins revealed that *Capsicum annuum* R gene proteins clustered closely with *Solanum lycopersicum* and *Nicotiana tabacum* R gene proteins and had the highest per cent of amino acid similarity with good query coverage. A venn graph was drawn based on similarity to the proteins database of *Oryza sativa*, *Arabidopsis thaliana*, *Zea mays* and *Vitis vinifera* to determine the distribution of R gene proteins across diverse crops. Four R gene proteins shared homology with proteins from all of the crops mentioned above. One among the three R gene proteins was found in all crops except *Z. mays*. The other two R gene proteins were identified in *Z. mays*, *O. sativa* and *A. thaliana* and *Vitis vinifera* species, respectively. These mined R genes can be used to produce CMV resistance varities. Furthermore, the R proteins can be utilized to study protein-protein interactions.

Keywords : *Cucumber mosaic virus*, R genes, *Capsicum annuum, In-silico* analysis

CUCUMBER MOSAIC VIRUS (CMV), a type species of the genus *Cucumovirus* in the family *Bromoviridae* is a plant virus found worldwide in temperate and tropical climates (Kumari *et al*., 2013). CMV is a tripartite virus with three plus sense, single stranded RNA molecules contained in a distinct particle with an isometric shape. RNAs 1 and 2 are involved in viral genome replication (Ding *et al*., 1994), whereas RNA 3 encodes movement and coat proteins (Lefkowitz *et al*., 2018). CMV has a wide host range and known to infect over 1200 species, inflicting significant harm to the majority of commercially important crops in the *Solanaceae* and *Cucurbitaceae* families (Roossinck, 2002). More than 80 species of aphids readily transmit the virus in non-persistent manner. In addition, this CMV is transmitted through sap and externally on seed, resulting in rapid disease spread and devastation of many crop plants (Palukaitis *et al*., 1992; Palukaitis & Garca Arenal, 2003 and Gildow *et al*., 2008). CMV causes a variety of symptoms on its host plants, including mild mosaic, mosaic mottling, stunting and stem necrosis (Ashwathappa *et al*., 2021). In some geographical areas, the devastating effect of CMV has caused farmers to for sake the growing crop altogether.

The Mysore Journal of Agricultural Sciences

Unlike other pathogens, the management of viral diseases is becoming increasingly difficult due to the emergence of new species / strains or the introduction of new viruses to geographical areas where they were not previously present, also the evolution of more efficient vector biotypes and the expanding host range of viruses (Vinaykumar *et al*., 2018, Venkataravanappa *et al*., 2021). CMV is difficult to handle since it has been successful in fast adapting to different hosts and surroundings due to the high rate of mutation and evolution leading to new strains (Kavyashree *et al*., 2018). It also has an impact on host gene expression by affecting cellular processes and signal transduction pathways *via* interactions with host proteins (Agudelo-Romero *et al*., 2008 and Babu *et al*., 2008). Plant resistance (R) genes have the ability to identify and counteract pathogens. As a result, structural and functional investigation of R genes and their analogues (RGA) is critical for developing disease-resistant crops.

Huge economic impact caused by CMV has sparked intense scientific interest in elucidating its pathogenesis mechanisms and underlying genetic variables. The use of bioinformatics in plant pathology has shown tremendous growth in recent years. It has significantly aided in taking Cucumovirus research to the next level of sophistication, where resistance genes are mined from diverse host species, their properties and structures are predicted and resistance breeding programs are implemented (Salánki *et al*., 2018).

Screening for the presence of R genes against CMV in diverse host plants is a time-consuming approach. However, advances in computational biology can be used to mine the R genes. The mined genes and their functional analyses can aid in understanding the interaction of viral and host R proteins. The existing literature provides limited details on R genes that protect against CMV. Keeping these research limitations in mind, the current study used a part of already available transcriptome data obtained from the chilli samples challenge inoculated with CMV to mine R gene proteins against CMV in diverse hosts.

Material and Methods

Selection of R Genes and Retrieval of Amino Acid Sequences

The transcriptome for CMV-chilli interaction was previously generated and differential gene expression was evaluated at Department of Plant Pathology, CoA, UAS, GKVK, Bengaluru (Vinaykumar, 2020). The transcriptome data generated was at various time intervals (5, 24, 72 and 120 hours) after inoculation of CMV to a resistant (IIHR-3476) and susceptible *Capsicum annuum* lines (IIHR-2541). The R gene proteins and transcription factors (TFs) with a log 2 fold change value greater than 2 in all time points in the resistant variety were chosen for the current investigation based on data (Table 1). The selected five R genes and two TFs amino acid (aa) sequences were obtained from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/).

Computation of Physicochemical Properties of selected CMV R Gene Proteins

The aa sequences of selected R genes and TFs were uploaded in FASTA format to the Expasy Prot Param program (https://web.expasy.org/protparam/) (Gasteiger *et al*., 2003). The amount of amino acids, molecular weight, theoretical isoelectric point (pI), total number of positively charged residues, instability index, aliphatic index (AI) and grand average of hydropathicity (GRAVY) of a protein were estimated using the Prot Param program.

In silico Mining of Resistance Genes in different Host Plants of CMV and other Crops

CMV has a wide host range and for this investigation, homologous sequences of selected R gene proteins from *C. annumm.* Search was made for the R gene and TFs homologues of chilli in *Cucumis sativus, C. melo, Cucurbita maxima, C. pepo, Citrullus lanatus, Benincasa hispida, Lagenaria siceraria, Luffa acutangula, L. aegyptiaca, Momordica charantia, Nicotiana tabacum* and *Solanum lycopersicum,* which are known to susceptible for CMV.

TABLE 1

List of selected R gene protein accession IDs, name of the protein and their putative function. The R gene proteins were selected based on the previously available transcriptome data from *Cucumber mosaic virus* and *Capsicum annuum* interaction

NCBI accession ID	Putative protein name	Function	Reference
NP_001312033	WRKY transcription factor	Involved in disease resistance in Arabidopsis thaliana	Duan *et al*., 2015
XP_016547009	bHLH transcription factor	It facilitates resistance to *Phytophthora sojae* in *Glycine max*	Cheng *et al*., 2018
XP_016581449	LRR-III kinase	LRR-RLK mediated resistance in N. benthamiana infected with Tobacco curly shoot virus	Li *et al*., 2018
XP_016560054	G-Lectin RLK	Plays important role in signal perception durig biotic stresses	Vaid *et al*., 2012
XP_016570057	CDPK Kinase	Universal secondary messanger and enhances activates defense genes	Bundo *et al*., 2016
XP_016556906	Glutathione S-transferase	Multifunctional enzyme and has role in hypersensitive response	Gullner *et al*., 2018
XP_016570736	Phytoharmone brassinosteroid	Modulates peroxidase-mediated oxidative burst and plant defense in rice against *Rice black-streaked dwarf virus*	Zhang *et al*., 2019

The aa sequences of selected R genes and TFs in FASTA format were used to search for homologous sequences in all of the above-mentioned hosts using the NCBI Basic Local Alignment Search Engine (BLAST) search tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). In order to determine the evolutionary relationship of these R gene proteins, phylogenetic trees were generated for each of the seven proteins separately using the neighbour-joining method available in MEGA X software with 1000 bootstrap replications (Kumar *et al*., 2018).

The R gene and TFs protein sequences obtained from *C. annuum* were uploaded to the Ortho Venn 2 webserver (https://orthovenn2.bioinfotoolkits.net/home) (Xu *et al*., 2019) to identify the presence of homologous gene sequences in a variety of agricultural plants. sOrthoVenn2 is a web-based tool for whole-genome comparison and annotation of orthologous clusters from various species. All seven *C. annuum* R gene proteins were compared to *Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinifera*, and *Zea mays* proteins present in the database.

RESULTS AND DISCUSSION

Physico-Chemical Characterization of Retrieved R Gene Protein Sequences

Physicochemical properties are thought to be critical in determining the function and structure of protein sequences. The physicochemical characteristics of the deduced aa sequences of *C. annuum* R genes and TFs against CMV were computed using the Expasy Prot Param tool and the results are shown in Table 2. The calculated molecular weights of R genes and TFs ranged from 11943.65 Da for phytoharmone

TABLE 2

Physico-chemical properties of R gene proteins. The physico-chemical properties were computed using the Expasy Protparam tool

Name of the protein	Number of amino acids	Molecular weight (Daltons)	Theoretical pI	Total number of negatively charged residues	Total number of positively charged residues	The instability index	Aliphatic index	Grand average of hydropathicity (GRAVY)
WRKY transcription factor	636.00	69308.58	6.32	61.00	54.00	53.45	54.25	-0.80
bHLH transcription factor	329.00	36518.38	5.16	44.00	35.00	54.94	62.58	-0.88
LRR-III kinase	831.00	89379.90	8.14	79.00	82.00	33.65	106.67	0.00
G-Lectin RLK	824.00	92380.60	6.77	92.00	90.00	36.68	79.11	-0.31
CDPK Kinase	502.00	56693.64	5.40	75.00	60.00	34.77	84.10	-0.35
Glutathione S-transferase	225.00	26000.78	5.53	33.00	26.00	35.23	90.53	-0.29
Phytoharmone brassinosteroid	105.00	11943.65	6.39	14.00	13.00	46.13	62.95	-0.27

brassinosteroid (XP 016570736) to 92380.60 Da for G-Lectin RLK (XP 016560054). Four of the five R genes and two TFs produced positively charged proteins, with the exception of LRR-III kinase, which was shown to be negatively charged. The computed isoelectric point (pI) also followed a similar pattern, with six proteins having pI<7, suggesting an acidic nature, except for LRR-III kinase, which had a pI of 8.14, indicating a basic nature. The isoelectric point is the pH at which the protein has no electric charge, and this value is important in protein purification because solubility is low, protein is stable and compact (Sahay and Shakya, 2010). The computed instability index (II) for LRR-III kinase ranged from 33.65 to 54.94 for bHLH transcription factor.

The metabolic stability of a protein in a test tube can be characterized based on the value of II, with a value less than 40 indicating a stable protein and a value greater than 40 indicating an unstable nature of the protein in the test tube (Gamage *et al*., 2019). Four R gene proteins, G-Lectin RLK, LRR-III kinase, CDPK Kinase and Glutathione S-transferase, with II values less than 40, were found to be stable *in vitro*, whereas the other three proteins, WRKY transcription factor, bHLH transcription factor and Phytoharmone brassinosteroid, were classified as unstable because their II values were greater than 40. The aliphatic side chains (alanine, valine, isoleucine and leucine) are responsible for the thermal stability of globular proteins and the aliphatic index (AI) of a protein is defined as the relative volume filled by aliphatic side chains (Panda and Chandra, 2012). The AI for resistance gene proteins was shown to have a wide range of values. Among the retrieved proteins, LRR-III kinase with an AI of 106.67 can be considered to have excellent thermostability. The grand average hydropathicity (GRAVY) value is calculated by dividing the sum of a protein's hydropathy values by the number of residues in the sequence and it measures the hydrophobicity or hydrophilicity of a protein (Babnigg and Joachimiak, 2010). Most proteins have a GRAVY value ranging from -2 to + 2, with positively graded proteins being more hydrophobic. All seven proteins had a negative GRAVY rating, indicating their hydrophilic nature. The computed physico-chemical properties assist in the study of a protein's function and nature and this knowledge is critical during numerous *in-vitro* and *in-vivo* protein interaction research (Yu *et al*., 2017).

Mining of Resistance Genes in different Host Plants of CMV and Diversity Analysis

The amino acid sequences of each R gene protein and TFs were examined for the presence of homologues sequences in CMV-infected host plants.

TABLE 3

List of homologous genes to *C. annuum* resistance genes against CMV in its host plants. The information was deduced through BLAST analysis

	NP_001312033 (WRKY trancription factor)			XP_016547009 (bHLH transcription factor)			XP_016581449 (LRR-III kinase)			XP_016560054 (G-Lectin RLK)		
Host plants	ID	%QC	%NS	ID	%QC	%NS	ID	%QC	%NS	ID	%QC	%NS
C. sativus	XP_011658698	83.00	56.75	XP_011657414	99.00	46.49	XP_004137694	95.00	68.92	XP_031738218	100.00	51.25
C. melo	TYK07115	72.00	62.29	XP_008445413	99.00	46.78	KAA0044200	95.00	68.55	TYK26346	96.00	52.17
C. maxima	XP_022969209	72.00	63.71	XP_022997211	99.00	46.36	XP_022983330	98.00	67.68	XP_023001226	97.00	48.58
C. pepo	XP_023531598	82.00	56.35	XP_023546510	99.00	46.06	XP_023536495	93.00	67.52	XP_023519597	99.00	51.05
C. lanatus	NA	NA	NA	NA	NA	NA	BAD26585	41.00	55.90	NA	NA	NA
B. hispida	XP_038889024	72.00	62.50	XP_038886108	99.00	46.94	XP_038892260	93.00	67.69	XP_038895379	96.00	53.97
L. siceraria	NA	NA	NA	NA	NA	NA				NA	NA	NA
L. acutangula	NA	NA	NA	NA	NA	NA	AKH61014	39.00	32.30	NA	NA	NA
L. aegyptiaca	AQX45439	67.00	63.43	NA	NA	NA	NA	NA	NA	NA	NA	NA
M. charantia	XP_022135705	83.00	58.26	XP_022131969	89.00	48.16	XP_022146757	97.00	65.23	XP_022139494	98.00	48.05
N. tabacum	XP_016445607	94.00	82.17	XP_016461736	99.00	61.93	XP_016510803	100.00	86.43	XP_016497189	100.00	83.29
S.lycopersicum	XP_004243486	93.00	81.02	XP_010320830	100.00	75.15	XP_004246299	98.00	89.13	XP_010316635	100.00	87.05

	XP_016570057 (CDPK Kinase)			XP_016556906 (Glutathione-S-transferase)			XP_016570736 (Phytoharmone brassinosteroid)			
Host plants	ID	%QC	%NS	ID	%QC	%NS	ID	%QC	%NS	
C. sativus	XP_004149412	97.00	78.48	XP_004147764	97.00	63.47	XP_004148349	52.00		58.18
C. melo	XP_008461962	97.00	78.69	KAA0062975	96.00	62.67	XP_008443179	78.00		40.00
C. maxima	XP_022981457	97.00	77.51	XP_022984351	96.00	60.55	XP_022971619	78.00		40.00
C. pepo	XP_023521317	97.00	77.30	XP_023523969	96.00	60.55	XP_023534575	52.00		63.64
C. lanatus	BAD26573	31.00	67.50	NA	NA	NA	NA	NA	NA	
B. hispida	XP_038899801	97.00	79.10	XP_038896278	95.00	62.15	XP_038903009	78.00		40.95
L. siceraria	NA	NA	NA	NA	NA	NA	NA	NA	NA	
L. acutangula	NA	NA	NA	NA	NA	NA	NA	NA	NA	
L. aegyptiaca	NA	NA	NA	NA	NA	NA	NA	NA	NA	
M. charantia	XP_022143623	97.00	79.51	XP_022142876	97.00	61.64	XP_022152070	78.00		40.95
N. tabacum	XP_016444304	99.00	89.02	NP_001312067	96.00	85.32	XP_016488795	77.00		72.84
S. lycopersicum	XP_004250396	99.00	94.42	XP_004243193	97.00	67.12	XP_004237527	77.00		79.01

Note: ID- NCBI accession IDs, %QC-percent query coverage, % NS- percent nucleotide similarity, NA- Not available

Table 3 shows the sequences with the highest query coverage, the highest amino acid (aa) sequence percent identity and the lowest Expect value (E-value) for each R gene protein and TFs. There were no R gene proteins and TFs discovered in *L. siceraria* and only one similar R gene protein of *C. annuum* LRR-III kinase was detected in *L. acutangula* (AKH61014). However, it only had 32.30 per cent aa sequence similarity and 39 per cent query coverage. *C. lanatus* has two R gene homologues, *C. annuum* LRR-III kinase and CDPK Kinase, with per cent sequence similarity of 55.90 and 67.50 and per cent query coverage of 41 and 31, respectively. *C. annum* WRKY transcription was the only protein found in *L. aegyptiaca* which shared 63.43 per cent aa similarity with 67 per cent query coverage. The homologous proteins of all the five R genes and two TFs with varying amino acid sequence per cent identity and per cent query coverage were observed in the remaining CMV hosts; *C. sativus*, *C. melo*, *C. maxima*, *B. hispida*, *M. charantia*, *N. tabacum* and *S. lycopersicum* (Table 3). Kumari *et al*. (2013) investigated the presence of the rice blast resistance gene Pi54 in wild and farmed rice types and discovered variability in the retrieved R gene sequences. The current study attempted to mine the *C. annuum* R gene proteins in different CMV hosts. However, few R gene proteins were not detected in some hosts such as *L. siceraria*, *L. acutangula*, *L. acutangula*, and *C. lanatus*, implying that the genes may not be present in these hosts or that less data about these genes has been submitted in the public domain database.

MEGA X was used to perform phylogenetic analysis for the R genes of *C. annuum* and its host to determine their ancestral connection (Fig. 1-7). According to the results of the investigation, all of the R gene proteins of *C. annuum* were shown to be closely clustered with the R gene proteins of *S. lycopersicum* and *N. tabacum*. Because these three crops are members of the *Solanaceae* family, there is a chance that they include homologous genes with significant similarity. The R genes found in the remaining crops formed a separate cluster, and they all belong to the cucurbitaceous family.

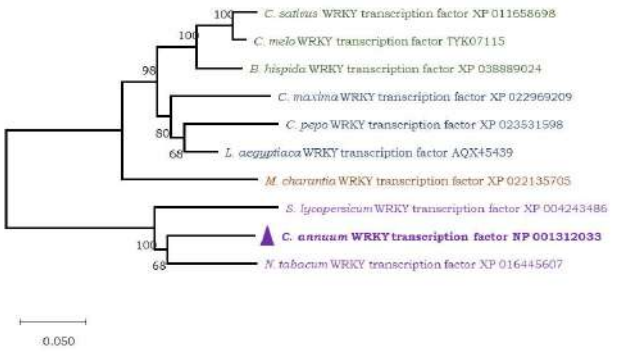

Fig. 1: Phylogenetic trees constructed for aa sequences of WRKY resistance protein using Neighbor-joining algorithm. Horizontal distances are proportional to sequence distances; vertical distances are arbitrary. A bootstrap analysis with 1000 replicates was performed

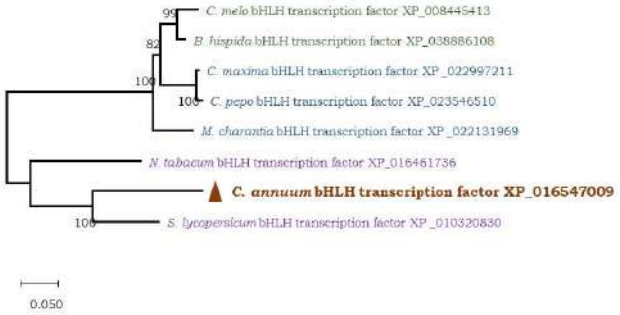

Fig. 2: Phylogenetic trees constructed for aa sequences of bHLH transcriptome factor using Neighbor-joining algorithm. Horizontal distances are proportional to sequence distances; vertical distances are arbitrary. A bootstrap analysis with 1000 replicates was performed.

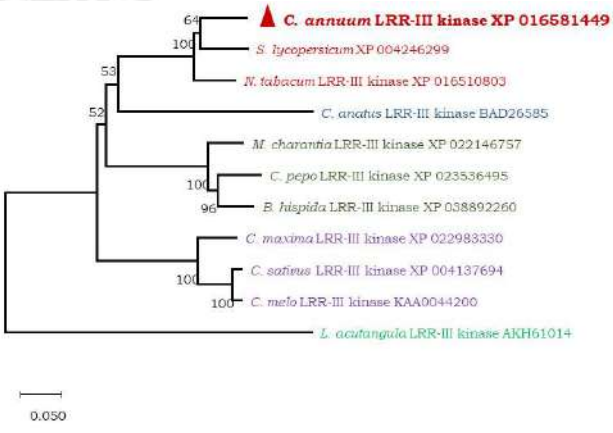

Fig. 3: Phylogenetic trees constructed for aa sequences of LRR-III kinase using Neighbor-joining algorithm. Horizontal distances are proportional to sequence distances; vertical distances are arbitrary. A bootstrap analysis with 1000 replicates was performed.

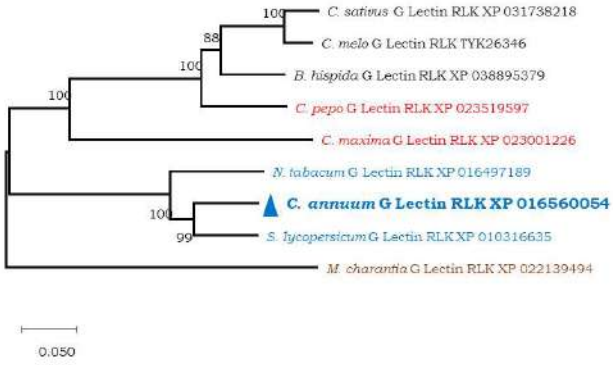

Fig. 4: Phylogenetic trees constructed for aa sequences of G-Lectin RLK using Neighbor-joining algorithm. Horizontal distances are proportional to sequence distances; vertical distances are arbitrary. A bootstrap analysis with 1000 replicates was performed

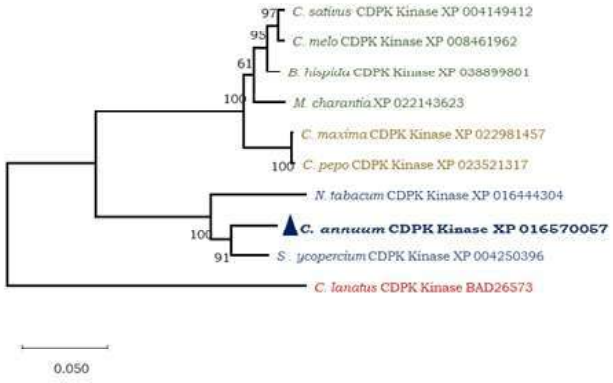

Fig. 5: Phylogenetic trees constructed for aa sequences of CDPK Kinase using Neighbor-joining algorithm. Horizontal distances are proportional to sequence distances; vertical distances are arbitrary. A bootstrap analysis with 1000 replicates was performed

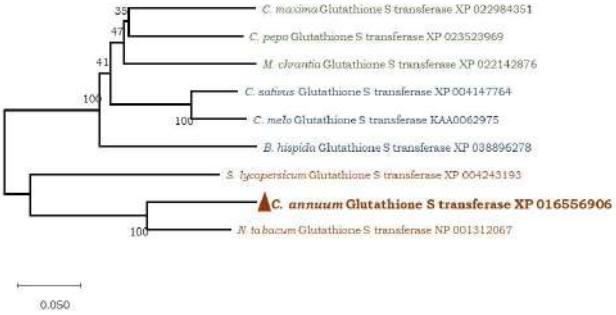

Fig. 6: Phylogenetic trees constructed for aa sequences of Glutathione-S-transferase using Neighbor-joining algorithm. Horizontal distances are proportional to sequence distances; vertical distances are arbitrary. A bootstrap analysis with 1000 replicates was performed

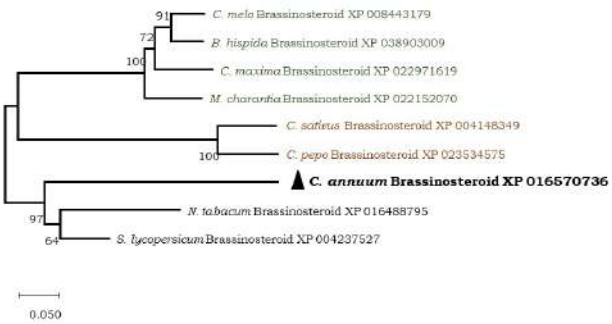

Fig. 7: Phylogenetic trees constructed for aa sequences of Brassinosteroid using Neighbor-joining algorithm. Horizontal distances are proportional to sequence distances; vertical distances are arbitrary. A bootstrap analysis with 1000 replicates was performed

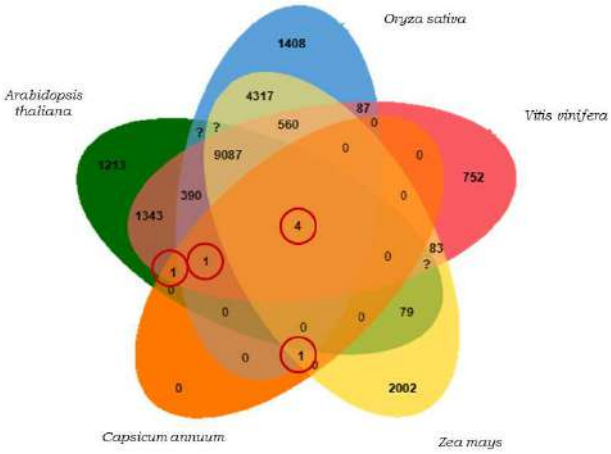

Fig. 8: Venn diagram displays the distribution of shared resistance gene orthologous clusters among the five plant species. The seven CMV resistance gene proteins of *C. annuum* were compared with the protein dataset of *Oryza sativa, Zea mays, Arabidopsis thaliana* and *Vitis vinifera* orthovenn2 webserver. The number embedded in red circle represent the shared genes

Mining of Orthologous Clusters of R Gene Proteins in selected Crop Plants

Orthologous genes are homologues that evolved from a common ancestor through mutation and recombination (Fang *et al*., 2010). The protein databases of *A. thaliana*, *O. sativa*, *V. vinifera* and *Z. mays* in OrthoVenn2 website were used to compare all five R gene proteins and two TFs (Fig. 8). Two R genes, glutathione S-transferase, LRR-III kinase and two TFs WRKY transcription factor, bHLH transcription factor were discovered to be present in

all of the crop plants after a pairwise comparison of all R genes with the protein data set homologues proteins. Except for *Z. mays*, the R gene protein, CDPK Kinase, was detected in all crops. The phytoharmone brassinosteroid protein was found in *Z. mays* and *O. sativa* but not in other crops. Similarly, homologues of G-Lectin RLK proteins were discovered in *V. vinifera* and *A. thaliana*. This information can be used in disease resistant breeding and disease resistance programs.

The mining of CMV resistance (R) genes throughout its host species will provide a wealth of information about naturally occurring resistant genes. These resistance genes' physicochemical qualities may be useful in *in-vitro* and *in-vivo* protein-protein interaction research. The ancestry of R genes found in diverse hosts gives important information on their diversity and distribution. Knowledge of R genes is extremely useful in creating CMV resistant cultivars using both standard and unorthodox breeding strategies.

REFERENCES

AGUDELO-ROMERO, P., CARBONELL, P., DE LA IGLESIA, F., CARRERA, J., RODRIGO, G., JARAMILLO, A., PEREZ-AMADOR, M. A. AND ELENA, S. F., 2008, Changes in the gene expression profile of *Arabidopsis thaliana* after infection with *Tobacco etch virus*. *Virol. J.*, **5** : 92 - 92.

ASHWATHAPPA, K. V., KRISHNA REDDY, M., VENKATARAVANAPPA, V., MADHAVI REDDY, K., HEMACHANDRA REDDY, P. AND LAKSHMINARAYANA REDDY, C. N., 2021, Genome characterization and host range studies of *Cucumber mosaic virus* belonging to the Subgroup IB infecting chilli in India and screening of chilli genotypes for identification of resistance. *Virus Dis*., 1 - 13.

BABNIGG, G. AND JOACHIMIAK, A., 2010. Predicting protein crystallization propensity from protein sequence. *J. Struct. Funct. Genom*., **11** : 71 - 80.

BABU, M., GRIFFITHS, J. S., HUANG, T. S. AND WANG, A., 2008, Altered gene expression changes in Arabidopsis leaf tissues and protoplasts in response to *Plum pox virus* infection. *BMC Genomics,* **9** : 325 - 333.

CHENG, Q., DONG, L., GAO, T., LIU, T., LI, N., WANG, L., CHANG, X., WU, J., XU, P. AND ZHANG, S., 2018, The bHLH transcription factor GmPIB1 facilitates resistance to *Phytophthora sojae* in *Glycine max*. *J. Experiment. Bot*., **69** : 2527 - 2541.

DUAN, Y., JIANG, Y., YE, S., KARIM, A., LING, Z., HE, Y., YANG, S. AND LUO, K., 2015, PtrWRKY73, a salicylic acid-inducible poplar WRKY transcription factor, is involved in disease resistance in *Arabidopsis thaliana*. *Plant cell rep*., **34** : 831 - 841.

FANG, G., BHARDWAJ, N., ROBILOTTO, R. AND GERSTEIN, M. B., 2010, Getting started in gene orthology and functional analysis. *PLoS comput. Biol*., **6** : 701 - 712.

GAMAGE, D. G., GUNARATNE, A., PERIYANNAN, G. R. AND RUSSELL, T. G., 2019, Applicability of instability index for *in vitro* protein stability prediction. *Protein Pept. Lett.*, **26** : 339 - 347.

GASTEIGER, E., GATTIKER, A., HOOGLAND, C., IVANYI, I., APPEL, R. D. AND BAIROCH, A., 2003, ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31** : 3784 - 3788.

GILDOW, F. E., SHAH, D. A., SACKETT, W. M., BUTZLER, T., NAULT, B. A. AND FLEISCHER, S. J., 2008, Transmission efficiency of *Cucumber mosaic virus* by aphids associated with virus epidemics in snap bean. *Phytopathol*., **98** : 1233 - 41.

GULLNER, G., KOMIVES, T., KIRÁLY, L. AND SCHRÖDER, P., 2018, Glutathione S-transferase enzymes in plant-pathogen interactions. *Front. Plant Sci*., **9** : 1836 - 1842.

KAVYASHREE, V. V., NAGARAJU, N. AND MOHAN RAO, A., 2018, Serological confirmation of chilli (*Capsicum annuum* L.) Lines for their resistance against *Cucumber mosaic virus. Mysore J. Agric. Sci.,* **52** (2) : 375 - 383.

KUMAR, S., STECHER, G., LI, M., KNYAZ, C. AND TAMURA, K., 2018, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol*., **35** : 1547 - 1552.

KUMARI, R., BHARDWAJ, P., SINGH, L., ZAIDI, A. A. AND HALLAN, V., 2013, Biological and molecular characterization of *Cucumber mosaic virus* Subgroup II isolate causing severe mosaic in Cucumber. *Indian J. Viro*l., **24** : 27 - 34.

LEFKOWITZ, E. J., DEMPSEY, D. M., HENDRICKSON, R. C., ORTON, R. J., SIDDELL, S. G. AND SMITH, D. B., 2018, Virus taxonomy: the database of the international committee on taxonomy of viruses (ICTV). *Nucleic acids res*., **46** : 708 - 717.

LI, K., WU, G., LI, M., MA, M., DU, J., SUN, M., SUN, X. AND QING, L., 2018, Transcriptome analysis of *Nicotiana benthamiana* infected by *Tobacco curly shoot virus*. *Virol. J.,* **15** : 138 - 145.

PALUKAITIS, P. AND GARCIA ARENAL, F., 2003, Cucumo viruses. *Adv. Virus Res*., **62** : 241 - 323.

PALUKAITIS, P., ROOSSINCK, M. J., DIETZGEN, R. G. AND FRANCKI, R. I., 1992, *Cucumber mosaic virus*. *Adv. Virus Res*., **41** : 281 - 348.

PANDA, S. AND CHANDRA, G., 2012, Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates. *Bioinformation*, **8** : 891 - 896.

ROOSSINCK, M. J., 2002, Evolutionary history of *Cucumber mosaic virus* deduced by phylogenetic analysis. *J. Virol*., **76** : 3382 - 3387.

SAHAY, A. AND SHAKYA, M., 2010, *In silico* analysis and homology modelling of antioxidant proteins of spinach. *J. Proteomics Bioinform*., **3** : 148 - 154.

SALSNKI, K., GELLERT, A., NEMES, K., DIVEKI, Z. AND BALAZS, E., 2018, Molecular modeling for better understanding of cucumovirus pathology. *Adv. Virus Res*., **102** : 59 - 88.

VAID, N., PANDEY, P. K. AND TUTEJA, N., 2012, Genome-wide analysis of lectin receptor-like kinase family from arabidopsis and rice. *Plant Mol. Biol.*, **80** : 365 - 388.

VENKATARAVANAPPA, V., REDDY, C. N. L., NANDAN, M., HIREMATH, S., ASHWATHAPPA, K. V., SHANKARAPPA, K. S., VINAY KUMAR, H. D. AND REDDY, M. K., 2021, Transmission, characterization and occurrence of recombination in Indian strain of *squash leaf curl China virus* associated with yellow mosaic and leaf curl disease of Summer squash. *3 Biotech*, **11** : 265 - 272.

VINAYKUMAR, H. D., 2020, Characterization of *cucumber mosaic virus* (CMV) and understanding the molecular mechanisms of CMV and chilli interaction through transcriptomics. *Ph.D. Thesis* (Unpub.), Univ. Agric. Sci., Bangalore.

VINAYKUMAR, H. D., REDDY, C. N. L. AND KRISHNAREDDY, M., 2018, Molecular detection of *Cucumber mosaic virus* infecting Chilli *Capsicum annuum* L. *Mysore J. Agric. Sci*., **54** (4) : 738 - 743.

XU, L., DONG, Z., FANG, L., LUO, Y., WEI, Z., GUO, H., ZHANG, G., GU, Y. Q., COLEMAN-DERR, D., XIA, Q. AND WANG, Y., 2019, OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic acids res*., **47** : 52 - 58.

YU, L., ZHANG, Y., GUTMAN, I., SHI, Y. AND DEHMER, M., 2017, Protein sequence comparison based on physicochemical properties and the position-feature energy matrix. *Sci. Rep*., **7** : 46237 - 46237.

ZHANG, H., HE, Y., TAN, X., XIE, K., LI, L., HONG, G., LI, J., CHENG, Y., YAN, F. AND CHEN, J., 2019, The dual effect of the brassinosteroid pathway on *rice black-streaked dwarf virus* infection by modulating the peroxidase-mediated oxidative burst and plant defense. *Molecular Plant-Microbe Interactions*, **32** : 685 - 696.

(*Received :* August 2021 *Accepted :* October 2021)