Indian Journal of Agricultural Sciences **87** (10): 1365–70, October 2017/Article
https://doi.org/10.56093/ijas.v87i10.75018

Comparative analysis of *Cf-4* and *Cf-19* in tomato (*Solanum lycopersicum)* – A bioinformatics study

GITANJALI TANDON[1], SATENDER SINGH[2] and SUKHDEEP KAUR[3]

Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, Uttar Pradesh 211 007

Received: 25 May 2017; Accepted: 25 August 2017

ABSTRACT

Tomatoes (*Solanum lycopersicum* L.), a model fruit crop, are largely affected by leaf mould disease, caused by an Ascomycete (*Cladosporium fulvum*) leading to significant economic loss across the globe. Though many R genes corresponding to this disease, viz. *Cf-2, Cf-4, Cf-5, Cf-6, Cf-9* and *Cf-19* are reported but such studies are confined to genomic level rather than proteomic level. Since host parasite interaction occurs at protein level thus in the present study, attempt has been made to carry out the studies of these R genes at proteomic level. Protein sequences of *Cf-4* and *Cf-19* genes were used for sequence analysis by Conserved Domain Database followed by construction of their three-dimensional models using Modeller. The obtained models were further validated and finalized protein models of both the genes were compared with the help of DALI Server. The present study revealed that there was 89% sequence identity between these proteins whereas they were confined to 78% only at structural level. Model generated from the present study can be used in interaction studies between Cf-4 and Cf-19 and their corresponding avirulence (Avr) protein. Such study would lead to better understanding of genetic basis of susceptibility of the tomato plants to the invading pathogen. This can also help in establishing them as putative candidate gene, which are desirable in development of disease resistant tomato varieties, combating pathogen attack in endeavour of tomato improvement program.

Key words: *Cf-4, Cf-19, Cladosporium fulvum*, Leaf mould disease, Tomatoes

Tomato (*Solanum lycopersicum* L.) is the one second most important vegetable crop in the world (Bhattarai *et al*. 2016). Due to its importance and consumption in various different forms, its production must be enhanced to feed the growing population. Various environmental factor, viz. biotic as well as abiotic, determine its production. It has been revealed that, there is about 162 million tonnes of production on 4.8 million harvested land area (FAOSTAT 2014). Every year there is a significant loss in its production due to both biotic as well as abiotic stress. There are about 200 disease that have been reported in tomatoes. Leaf mould is one of the most disastrous foliar diseases of field tomatoes under humid conditions (Veloukas *et al*. 2007) and believed to originate from South America. It is caused by *Cladosporium fulvum* which is a biotrophic fungal pathogen and was first described by M C Cooke (1883) on leaf samples sent from South Carolina (van Esse *et al*. 2007, Cooke 1883, Oliver *et al*. 2000). *Cladosporium fulvum* belongs to the class of fungi called imperfecti (also known as Deuteromycetes). Asexual conidia are produced on single condiospores by this fungus which lacks fruiting bodies (Joosten and de Wit. 1999). *Lycopersicon* genus acts as a host for this pathogen. Many *Lycopersicon* species are resistant to this pathogen but cultivated tomato (i.e. *Lycopersicon esculentum*) is susceptible. This fungus attaches itself to leaf but it can also attack stem, flowers as well as fruits (Butler and Jones 1949). The disease decreases the fruit quality and yield and can even lead to death of the tomato plant. During compatible interactions, between susceptible tomato plant and C. *fulvum* pathogen, the fungal spores attaches the abaxial surface of the leaves and enter the leaf through stomata. Later hypha emerges through stomata and from there it continues to grow and arborize, thus leading to the death of the infected cells (Hammond-Kosack and Jones 1884).

Resistance against C. *fulvum* is administered by gene-for-gene relationship which was given by Flor (1942) and Oort (1944). When the avirulence (Avr) proteins are released by the pathogen against the tomato plant, the resistant genes (R-genes) of the plant steers the activation of defense cascade which finally leads to hypersensitivity response and host immunity (Kruije *et al*. 2005). Plant pathogen interactions have been studied between Avr gene of *C. fulvum* and the R-genes of the tomato (*Cf* genes) (van Esse *et al*. 2009). Various *Cf* genes of tomatoes against Avr genes have also been cloned, that include *Cf-2*, *Cf-4*, *Cf-5*, *Cf-9* and *Cf-19* (Dixon *et al*. 1996, Dixon *et al*. 1998,

[1]Ph D Scholar (e mail: gitanjali.tandon@gmail.com), [2]Assistant Professor (e mail: satendra.singh@shiats.edu.in), [3]Ph D Scholar (e mail: sukhii.deep@gmail.com), Department of Computational Biology and Bioinformatics.

Grushtskaia *et al*. 2007, Jones *et al*. 1994, Takken *et al*. 1998, Thomas *et al*. 1997 and Zhao *et al*. 2016). *Cf* genes encode membrane anchored proteins largely composed of extracytoplasmic Leucine rich repeats (LRRs). Xa21 gene of rice contain component similar to both P to and *Cf* genes which encode transmembrane receptor like protein kinase with 23 extracellular LRR (Song *et al*. 1995, Dickinson *et al.* 1992). Studies reveal that *Cf-4* and *Cf-9* genes are located on chromosome 1, whereas *Cf-2* and *Cf-5* are located on Chromosome 6 of tomato (Dickinson *et al*. 1992). As far as *Cf-19* gene is concerned, its loci have been mapped on short arm of chromosome 1 (Zhao *et al*. 2016). It has also been reported that *Cf*-loci recognizes the Extracellular proteins (Ecps) released by *C. fulvum* (de Kock *et al.* 2005, Laugé *et al.* 1998). Though tomatoes have developed Cf-mediated pathways to recognize effector proteins secreted by *C. fulvum* and protect the plant from being infected, but then also effector genes are able to express during colonization in the host plant (Joosten *et al*. 1997, de Wit 1992, Thomma *et al*. 2006, Ackerveken *et al*. 1994 and van Kan *et al*. 1991). So, there is a requirement to develop more mould-resistant varieties of tomato.

Although lot of genomic information is available for the *Cf* genes but there is no information at proteomic level. Moreover, three-dimensional structural information has also not been reported yet. In our present study, two *Cf* genes, viz. *Cf-4* and *Cf-19* were studied at sequence level as well at structural level.

MATERIALS AND METHODS

To fulfill the desired aim, the protein sequences of *Cf-4* and *Cf-19* genes were obtained from NCBI-Protein (https://www.ncbi.nlm.nih.gov/protein/) (McEntyre and Ostell 2002). Both the sequences were compared with the help of Blastp program available at NCBI (Altschul *et al*. 1990). These sequences were further used to find out the domains present in them. This domain identification was performed with the help of Conserved Domain Database (CDD), a database that is used to mine the functional units in the protein sequences (Marchler-Bauer *et al.* 2017). After the functional analysis, physio-chemical properties of the both the sequences was done with ProtParam that computes various physical and chemical parameters of protein sequences (Walker 2005).

After the sequence analysis, three dimensional models of both the proteins were constructed using Modeller 9.16 based on homology modelling approach for constructing three dimensional protein models (Eswar *et al.* 2008, Fazil *et al.* 2012, Razali *et al.* 2014, Rawat *et al.* 2014 and Jha *et al.* 2014). Total ten models were constructed with Modeller, which were further validated with the help of PROCHECK. This software provides the check of stereochemical properties of a protein structure by analyzing residue by residue and overall structure geometry (Laskowski *et al*. 1993). Later RAMPAGE was used for producing the Ramachandra plots for both the proteins (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) (Lovell *et al*. 2003).

The finalized models were further compared with the help of DALI server (http://ekhidna.biocenter.helsinki.fi/dali_lite/start). This server checks whether conserved residues leads to multiple structural alignments (Holm 2010). The 3D structures of *Cf-4* and *Cf-19* were superimposed via Chimera software. Chimera is a software package with basic functions of docking, alignment and visualization of the biomolecules (Pettersen *et al*. 2004). Apart from the structural comparison, pockets for both the proteins were computed with CASTp server (http://sts.bioe.uic.edu/castp/) and then the results were compared (Binkowski *et al*. 2003).

RESULTS AND DISCUSSION

Sequence retrieval and its analysis

Proteins are the biological and functional units which coordinate for proper functioning of the body (Lesk 2001),

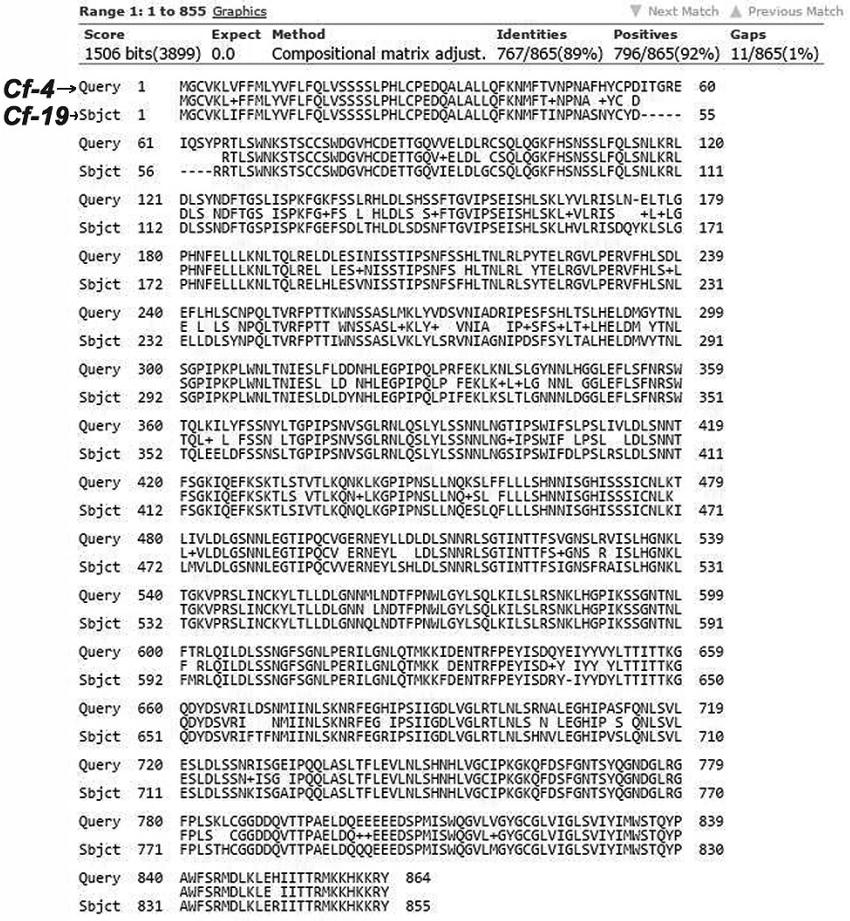


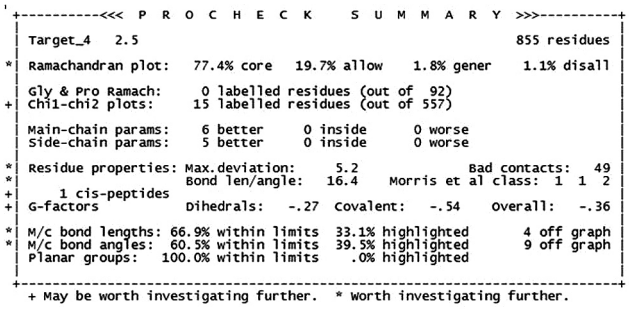
Fig 1 Pairwise sequence alignment of Cf-4 and Cf-19 gene

Table 1 Physicochemical properties

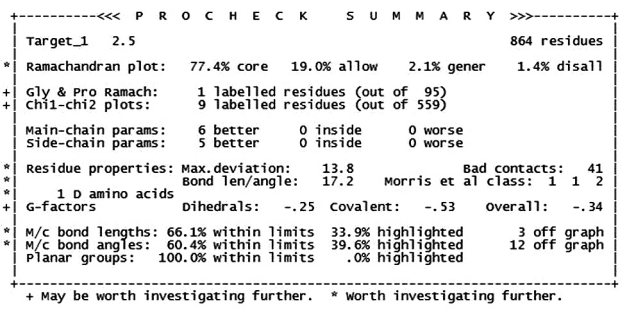
Attribute	Value	
	Cf-4	*Cf-19*
Number of amino acids	855	864
Molecular weight	96061.54	97146.13
Theoretical pI	6.45	7.11
Ext. coefficient (assuming all pairs of Cystine residues)	0.968	0.974
Instability index	37.09	36.37
Half life (*Escherichia coli, in vivo*)	>10 hours	>10 hours
Aliphatic index	100.29	101.04
Grand average of hydropathicity (GRAVY)	-0.158	-0.144

so their study can help in better understanding of the R gene related pathways. Sequence for *Cf-4* available with accession number CAA73187.1 and that of *Cf-19* available with accession AMN14930.1 was considered for the study. After performing Blastp, it was observed that both the sequences had 89% identity, 1% gaps and 1506 as bits score at 0.0 e-value. Details of this pairwise alignment is shown in Fig 1. Further, domains were mined in both the R genes. In case of *Cf-4 gene,* single domain of leucine-rich repeat receptor-like protein kinase was found from 3rd to 780th amino acid at 3.86e-78 e-value but in *Cf-19,* single domain of leucine-rich repeat receptor-like protein kinase was present in a stretch of 69 to 789 at 2.76e-81 e-value. This obviates that there is a common domain with little different length present in both the proteins. This was followed by computation of physicochemical properties of both the proteins using ProtParam interface available at Expasy server. Mostly all the properties were different in both the proteins. Details of the physio-chemical properties for both *Cf-4* and *Cf-19* are given in Table 1.

When both the proteins were compared at sequence level, it was observed that they had single LRR domain but the length of the domain was different in both the proteins. Since their amino acid number is different, thus molecular weight is also different. Other physicochemical properties were also dissimilar, although they both lie on the same

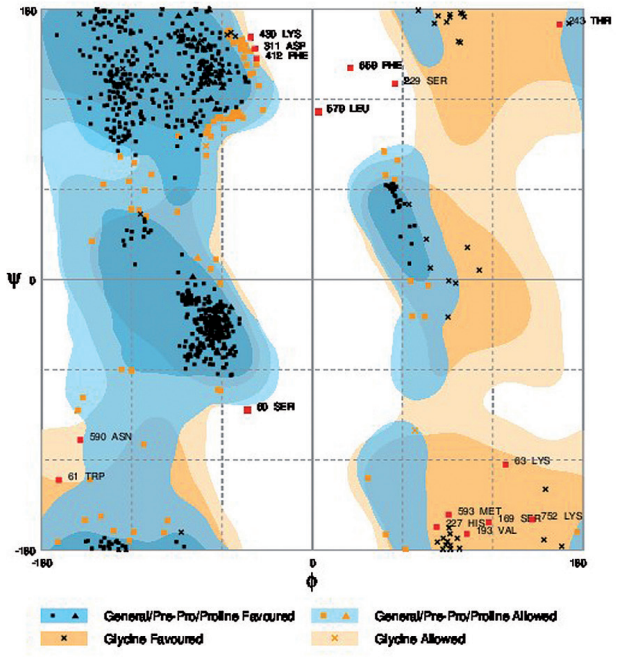


(a) *Cf-4*

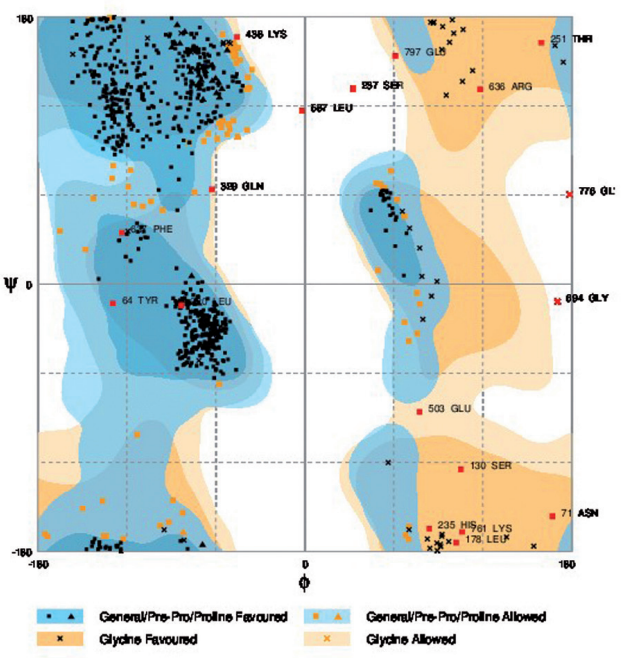


(b) *Cf-19*

Fig 2 PROCHECK results of *Cf-4* and *Cf-19*



(a) *Cf-4*



(b) *Cf-19*

Fig 3 Ramachandran plots obtained from RAMPAGE

Table 2 Structural alignment results obtained by DALI server

Z-score	RMSD	% identity
33.6	5	78
25.3	7.5	25
22.9	11.9	23
22.4	13.4	26
19.8	13.8	22
19.1	1.3	35
18.9	15.5	24
18.3	19.3	20
16.4	19.9	22
16	1.2	32

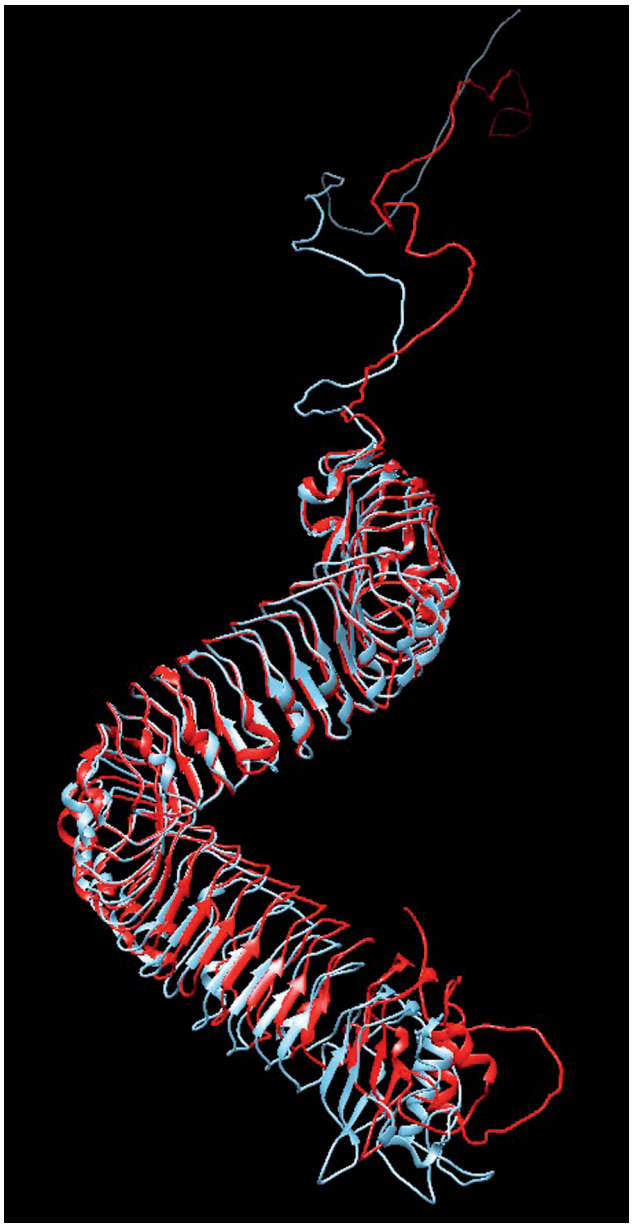


Fig 4 Superimposed structures of *Cf-4* and *Cf-19* in Chimera

chromosome 1 (Zhao *et al*. 2016, Dickinson *et al.* 1992).

Molecular modelling and structural validation

After completing the sequence analysis, molecular modeling was done with the Modeller. Based on alignment, 4MN8 and 4J0M were used as templates for *Cf-4* and *Cf-19,* respectively, for constructing models. 4MN8 had 30% identity with *Cf-4* at 3e-75 e-value, whereas 4J0M had 31% identity with e-value 7e-73. A total of ten structures were obtained via Modeller which were then further validated with the help of PROCHECK and RAMPAGE. Final model of *Cf-4* had 77.4% in favored region, 19.7% in allowed region, 1.8% generously allowed and 1.1% in disallowed region while final model of *Cf-19* had 89% in favored region, 8.9% in allowed region and 2.1% in outlier region. PROCHECK results and Ramachandran plots are given in Fig 2 and 3.

Structural and active site comparison

After finalizing three dimensional models of both the proteins, they were compared and superimposed with the help of DALI server and Chimera. Since three dimensional structures are more conserved as compared to the protein sequences, so structural comparison was done (Chothia and Lesk 1986, Murzin 1996, Murzin 1998), In DALI server, maximum identity obtained was 78% with 33.6 as Z-score and RMSD as 5Å. First ten alignments results obtained by DALI server are shown in Table 2. More the Z-score and lesser the RMSD, better is the alignment between the structures. After structural alignment, both the structures were superimposed in Chimera (Fig 4).

After comparing the structures, active sites for both the proteins were predicted with CASTp. For *Cf-4*, a total of 134 pockets were identified whereas in *Cf-19* protein, 118 pockets were observed. Details of predicted active sites have been included in supplementary sheet I.

Since, there is a difference between primary sequence of the proteins, thus their three-dimensional structures are also different as primary sequence decide the three dimensional structures of the proteins. Maximum identity between the structures obtained was 78% which was less than the sequence identity which was 89%.

Study of the interaction between *Cf-4* and *Cf-19* and their corresponding avirulence (Avr) protein is required to understand the genetic basis of the invading pathogen. This can help in developing disease resistant varieties to overcome pathogen attack in endeavor of tomato improvement program.

REFERENCES

Ackerveken G F J M, Dunn R M, Cozijnsen A J, Vossen J P M J, Broek H W J and Wit P J G M. 1994. Nitrogen limitation induces expression of the avirulence gene avr9 in the tomato pathogen *Cladosporium fulvum*. *Molecular and General Genetics* **243**(3): 277–85.

Altschul S F, Gish W, Miller W, Myers E W and Lipman D J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403–10.

Bhattarai K, Louws F J, Williamson J D and Panthee D R. 2016.

Diversity analysis of tomato genotypes based on morphological traits with commercial breeding significance for fresh market production in eastern USA. *Australian Journal of Crop Science* **10**(8): 1098.

Binkowski T A, Naghibzadeh S and Liang, J. 2003. CASTp: computed atlas of surface topography of proteins. *Nucleic Acids Research* **31**(13): 3352–5.

Butler E J and Jones S G. 1949. Tomato leaf mould, *Claosporium fulvum* Cooke. *Plant Physiology*.

Chothia C and Lesk A M. 1986.The relation between the divergence of sequence and structure in proteins. *EMBO Journal* **B**(4): 823.

Cooke M C. 1883.New American fungi. *Grevillea* **12**(61): 22–33.

de Kock M J, Brandwagt B F, Bonnema G, de Wit P J and Lindhout P. 2005. The tomato Orion locus comprises a unique class of Hcr9 genes. *Molecular Breeding* **15**(4): 409–22.

de Wit P J. 1992. Molecular characterization of gene-for-gene systems in plant-fungus interactions and the application of avirulence genes in control of plant pathogens. *Annual Review of Phytopathology* **30**(1): 391–418.

Dickinson M J, Jones D A and Jones J D. 1992. Close linkage between the *Cf-2*/*Cf-5* and Mi resistance loci in tomato. *Molecular Plant-Microbe Interactions* **6**(3): 341–7.

Dixon M S, Hatzixanthis K, Jones D A, Harrison K and Jones J D. 1998. The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell* **10**(11): 1915–25.

Dixon M S, Jones D A, Keddie J S, Thomas C M, Harrison K and Jones J D. 1996. The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* **84**(3): 451–9.

Eswar N, Eramian D, Webb B, Shen M Y and Sali A. 2008. Protein structure modeling with MODELLER. *Structural Proteomics: High-throughput Methods* pp 145–59.

Fazil M H U T, Kumar S, Naidu S R, Selvaraj C, Singh S K, Pandey H P and Singh D V. 2012. Comparative structure analysis of two proteins belonging to quorum sensing system in *Vibrio cholera*. *Journal of Biomolecular Structure and Dynamics* **30**(5): 574–84.

Grushtskaia Z E, Lemesh V A, Poliksenova V D and Khotyleva L V. 2007. Cloning of the Cf-6 tomato leaf mould resistance locus using SSR markers. *Genetika* **43**(11): 1511–6.

Hammond-Kosack K E and Jones J D. 1994. Incomplete dominance of tomato *Cf* genes for resistance to *Cladosporium fulvum*. *Molecular Plant Microbe Interactions* **7**: 58–58.

Holm L. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Research* ***38***(suppl 2): W545–W549.

Jha Y, Sablok G, Naidu S R, Sudhakar R, Fazil M H U T, Subramanian R B, Squartini A and Kumar S. 2014. Bacterial-induced expression of RAB18 protein in *Orzya sativa* salinity stress and insights into molecular interaction with GTP ligand. *Journal of Molecular Recognition* **27**(9): 521–7.

Jones D A, Thomas C M, Hammond-Kosack K E, Balint-Kurti P J and Jones J D. 1994. Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science-New York:* 789–9.

Joosten M H and de Wit P J. 1999. The tomato–*Cladosporium fulvum* interaction: A versatile experimental system to study plant-pathogen interactions. *Annual Review of Phytopathology* **37**(1): 335–67.

Joosten M H, Vogelsang R, Cozijnsen T J, Verberne M C and De Wit P J. 1997. The biotrophic fungus *Cladosporium fulvum* circumvents *Cf-4*-mediated resistance by producing unstable AVR4 elicitors. *Plant Cell* **9**(3): 367–79.

Kruijt M, de Kock M J and de Wit P J. 2005. Receptor like proteins involved in plant disease resistance. *Molecular Plant Pathology* **6**(1): 85–97.

Laskowski R A, MacArthur M W, Moss D S and Thornton J M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**(2): 283–91.

Laugé R, Joosten M H, Haanstra J P, Goodwin P H, Lindhout P and De Wit P J. 1998. Successful search for a resistance gene in tomato targeted against a virulence factor of a fungal pathogen. *Proceedings of the National Academy of Sciences* **95**(15): 9014–8.

Lesk A M. 2001. *Introduction to Protein Architecture: The Structural Biology of Proteins,* pp 217–26. Oxford University Press.

Lovell S C, Davis I W, Arendall W B, de Bakker P I, Word J M, Prisant M G, Richardson J S and Richardson D C. 2003. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function and Bioinformatics* **50**(3): 437–50.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki C J, Lu S, Chitsaz F, Derbyshire M K, Geer R C, Gonzales N R and Gwadz M. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research* **45**(D1): D200–D203.

McEntyre J and Ostell J. 2002. *The NCBI Handbook*. National Center for Biotechnology Information (US), Bethesda (MD).

Murzin A G. 1996. Structural classification of proteins: new superfamilies. *Current Opinion in Structural Biology* **6**(3): 386–94.

Murzin A G. 1998. How far divergent evolution goes in proteins. *Current Opinion in Structural Biology* **8**(3): 380–7.

Oliver R P, Henricot B and Segers G. 2000. *Cladosporium fulvum*, cause of leaf mould of tomato. (In) *Fungal Pathology*, pp 65–91. Springer, Netherlands.

Pettersen E F, Goddard T D, Huang C C, Couch G S, Greenblatt D M, Meng E C and Ferrin T E. 2004. UCSF Chimera?A visualization system for exploratory. *Interface* **8**: 1605–12.

Rawat R, Kumar S, Chadha B S, Kumar D and Oberoi H S. 2014. An acidothermophilic functionally active novel GH12 family endoglucanase from *Aspergillus niger* HO: purification, characterization and molecular interaction studies. *Antonie Van Leeuwenhoek* **107**(1): 103.

Razali N, Agarwal R, Agarwal P, Kumar S, Tripathy M, Vasudevan S, Crowston J G and Ismail N M. 2014. Role of adenosine receptors in resveratrol-induced IOP lowering in rats with steroid-induced ocular hypertension. *Clinical and Experimental Opthalmology* **43**(1): 54–66.

Song W Y, Wang G L, Chen L L and Kim H S. 1995.A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* **270**(5243): 1804.

Takken F L, Schipper D, Nijkamp H J J and Hille J. 1998. Identification and Ds-tagged isolation of a new gene at the Cf-4 locus of tomato involved in disease resistance to *Cladosporium fulvum* race 5. *Plant Journal* **14**(4): 401–11.

Thomas C M, Jones D A, Parniske M, Harrison K, Balint-Kurti P J, Hatzixanthis K and Jones J D. 1997. Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. *Plant Cell* **9**(12): 2209–24.

Thomma B P, Bolton M D, Clergeot P H and De Wit P J. 2006. Nitrogen controls in planta expression of *Cladosporium fulvum*

Avr9 but no other effector genes. *Molecular Plant Pathology* **7**(2): 125–30.

vanEsse H P, Bolton M D, Stergiopoulos I, de Wit P J and Thomma B P. 2007. The chitin-binding *Cladosporium Fulvum* effector protein Avr4 is a virulence factor. *Molecular Plant-Microbe Interactions* **20**(9): 1092–101.

vanEsse H P, Fradin E F, de Groot P J, de Wit P J and Thomma B P. 2009. Tomato transcriptional responses to a foliar and a vascular fungal pathogen are distinct. *Molecular Plant-Microbe Interactions* **22**(3): 245–58.

vanKan J A, Van den Ackerveken G F J M and De Wit P J G M. 1991. Cloning and characterization of cDNA of avirulence gene avr9 of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. *Molecular Plant-Microbe Interaction* **4**: 52–9.

Veloukas T, Bardas G A, Karaoglanidis G S and Tzavella-Klonari K. 2007. Management of tomato leaf mould caused by *Cladosporium fulvum* with trifloxystrobin. *Crop Protection* **26**(6): 845–51.

Walker J M (ed.). 2005. *The Proteomics Protocols Handbook,* pp 571–607. Humana Press, Totowa, NJ.

Zhao T, Jiang J, Liu G, He S, Zhang, H, Chen X, Li J and Xu X. 2016. Mapping and candidate gene screening of tomato *Cladosporium fulvum*-resistant gene *Cf-19*, based on high-throughput sequencing technology. *BMC Plant Biology* **16**(1): 51.