Symbiosis

www.symbiosisonline.org
www.symbiosisonlinepublishing.com

Research article | International Journal of Horticulture & Agriculture | Open Access

Computational Identification of Conserved Domains from Genomic Survey Sequences in Green Gram [Vigna radiata (L.) R. Wilczek]

Ramprasad Eruvuri[1]*, Prasad Gajula MNV[1], Durga Rani V[1], Anuradha[1], Vanisri S[1]

[1]Institute of Biotechnology, Professor Jayashankar Telangana State Agricultural University (PJTSAU), Rajendranagar, Hyderabad

Received: November 25, 2016; Accepted: January 23, 2017; Published: February 2, 2017

***Corresponding author:** Institute of Biotechnology, Professor Jayashankar Telangana State Agricultural University (PJTSAU), Rajendranagar, Hyderabad 500030, India. E-mail: rambiotech100@gmail.com*

Abstract

With recent advances in the field of genome sequencing, analysis and availability of large genomic data in the public domain, we made an attempt to survey the presence of the conserved domains, super families and multi domains having putative functions identified from green gram [*Vigna radiata* (L.) R. Wilczek] Genomic Survey Sequences (GSS) using computational tools. In this study we have been identified the various conserved domains, super families having putative functions for fundamental, metabolic, developmental, evolutionary processes and physiogenic nature from the query sequences. This study was beneficial in the area of comparativ e genomics for the identification of important genes and also development of functional molecular markers in identified genes for green gram and its related crops improvement.

Keywords: Conserved domains; Genomic Survey Sequences (GSS); Green gram; Multi domains; Putative functions; Super families

Introduction

Sequenced crop plants are the good resources in identification of important genes for quality, insect pest and disease resistance, resistance to abiotic stresses like temperature, drought, salinity etc. which are possible to transfer to cultivable background by combining traditional and molecular breeding methods [28]. Genome Survey Sequences (GSS) are nucleotide sequences similar to EST's that the only difference is that most of them are genomic in origin, rather than mRNA, While Expressed Sequence Tags sequences represent the expressed region of the genome [4]. These GSS and EST sequences used for the identification of "Functional Molecular Markers" (FMM) which are associated with trait of interest and may be transferable in closely related genera [3]. Genomic Survey Sequences of Green gram [*Vigna radiata* (L.) R. Wilczek] available online in public domain from NCBI (http://www.ncbi.nlm.nih.gov) for public use. Due to generation and availability of huge genomic information online of the crop plants, the computational studies i.e. performed on computer or via computer simulation are important areas of interest for genomics researchers for comparative genomics study [31, 21, 15].

Domains can be thought of as distinct functional and/or structural units of protein. The identification of a conserved domain footprint may be the only clue towards cellular or molecular function of a protein, as it indicates local or partial similarity to other proteins, some of which may have been characterized experimentally. Conserved Domains (CD) contain conserved sequence patterns or motifs, which allow for their detection in polypeptide sequences (www.ncbi.nlm.nih.gov/Structure/cdd). It has been suggested that domain combinations are evolutionarily conserved and evolution creates novel functions predominantly by combining existing domains [13]. The Conserved Domain Database (CDD) is a compilation of multiple sequence alignments representing protein domain conserved in molecular evolution [24]. Keeping all above points in view the computational identification study was carried out to identify the conserved domains having putative function like insect pest resistance, improved quality parameters or for resistance to abiotic stresses like temperature, drought, salinity or any other evolutionary functions from Green gram GSS and their possible use in Green gram and also its related crops improvement.

Material And Methods

Sequence Retrieval

The Genome Survey Sequences (GSS) of Green gram [*Vigna radiata* (L.) R. Wilczek] available online in public domain from NCBI (http://www.ncbi.nlm.nih.gov). These were downloaded in FASTA format to be used for further analysis.

Conserved Domain Search

Search for conserved domains within query GS sequence (nucleotide sequence) was analyzed using conserved domain search service (CDD search) available online (http://www.ncbi.nlm.nih.gov). The GS query sequences of more than 170 bp in length were used for the analysis. If a specific hit is not found on a query sequence, but the nucleotide sequence has an otherwise statistically significant hit (E-value cutoff of 0.01) to any domain model in CDD, the domain model is regarded as a non-specific hit. The E-value threshold used for filtering results was kept at 0.001,

maximum number of hits for CD search kept at 500 and the result mode kept as standard default settings.

Results and Discussion

A total of eighty two GS sequences which were chosen deliberately from the Green gram [*Vigna radiata* (L.) R. Wilczek] were analyzed for the identification of conserved domains, super families and for multi domains. The distribution of 82 hits in CDD search is represented in fig 1. As the online CDD search is inbuilt for identification of low complexity region search as determined by the SEG program of [38] or for BLASTN, by the DUST program of Tatusov and Lipman (ftp://ncbi.nlm.nih.gov/pub/tatusov/dust/versionl/), we used GS query sequences as such with more than 170 bp in length for analysis. After analysis of 82 GSS, we observed 30 non-specific hits, which contribute about 20 % sequence analyzed. The various conserved domains, super families having putative functions for fundamental, metabolic, developmental, evolutionary processes and physiogenic nature have been identified from the query sequences. From the specific hits a total of 16 conserved domains (11%) and 34 super families (23%) were observed, respectively and 29 multi domains (20%) were identified. It is interesting to note that out of 82 query sequence hits, 39 hits (26%) did not identify any specific CD or super family or multi domains indicating all these sequences are unique (unique GSS) to the Green gram species. A specific hit is a high confidence association between a nucleotide query sequence and a conserved domain, resulting in a high confidence level for the inferred function of the query sequence. In the GS query sequence AZ254242.1 (i.e sequence number 2) (Table l) the specific hit identified CD LRR_4 and have LRR_4 super family (Fig 2) with E-value of 2.62e-04. This leucine rich repeats are short sequence motifs present in a number of proteins with diverse functions and cellular locations. These repeats are usually involved in protein-protein interactions. Each Leucine Rich Repeat is composed of a beta-alpha unit. These units form elongated non-globular structures. Leucine Rich Repeats are often flanked by cysteine rich domains. LRR-containing proteins from plants have diverse overall structures and functions. Several classes contain LRR-containing receptor-like kinases (LRR-RLKs) [1, 16], LRR-containing receptor-like proteins (LRR-RLPs) [10], nucleotide binding site LRR (NBS-LRR) proteins [7, 26] and Poly-Galacturonase Inhibiting Proteins (PGIPs) [8, *9*, 30]. They provide an early warning system for the presence of potential pathogens and activate protective immune signaling in plants [20, 18, 35]. In addition, they act as a signal amplifier in the case of tissue damage, establishing symbiotic relationships and effecting developmental processes. Evolution of plant, disease resistance (R) genes that encode an LRR region has been studied by many researchers [1, 26, 27, 6, 23, 14, 25, 39, 17, 12, 32, 37, 22, 2, 40, 30, 11, 19]. The generations of R genes are proposed to be mainly due to gene duplication, genetic recombination, diversifying selection, and sequence divergence in the intergeneric region, composition of the transposable elements, gene conversion, and unequal crossover [40, 30, 11]. Another conserved sequence in the GS query sequence AZ254272.1 (i.e. sequence number 5) (Table l) the specific hit identified AspRS_cyto_N: N-terminal, anticodon recognition domain of the type found in Saccharomyces cerevisiae and human cytoplasmic aspartyl-tRNA synthetase (AspRS). This domain is a beta-barrel domain (OB fold) involved in binding the tRNA anticodon stem-loop. The enzymes in this group are homodimeric class2b aminoacyl-tRNA synthetases (aaRSs). aaRSs catalyze the specific attachment of Amino Acids (AAs) to their cognate tRNAs during protein biosynthesis. This 2-step reaction involves i) the activation of the AA by ATP in the presence of magnesium ions, followed by ii) the transfer of the activated AA to the terminal ribose of tRNA. In the case of the class2b aaRSs, the activated AA is attached to the 3'OH of the terminal ribose. Eukaryotes contain 2 sets of aaRSs, both of which are encoded by the nuclear genome. One set concerns with cytoplasmic protein synthesis, whereas the other exclusively with mitochondrial protein synthesis. This gene malfunctioning might be results in the hinderance in the protein synthesis of cytoplasm and mitochondria [34]. One more conserved domain in the GS query sequence AZ254277.1 (i.e sequence number 6) (Table l) the specific hit identified TVP38. The protein Tvp38 is conserved in yeasts and higher eukaryotes and potentially involved in

Table 1: Specific hits identified for CD, super families and multi domains.

S.No	Accession number	Length of sequence	Specific hits for the conserved domains	Specific hits for super families	Specific hits for multi domain
1	AZ254237.1	503	HATPase_c	HATPase_c	PRK11107, TMAO_torS, BaeS
2	AZ254242.1	502	LRR_4	LRR_4 super family	LRR_8, LRR, PLN03210
3	AZ254256.1	441	VQ	gal11_coact VQ	Tymo_45kd_70kd Kgd PAT1 PHA03247
4	AZ254260.1	439	LRR_4	LRR_4	LRR_8
5	AZ254272.1	451	AspRS_cyto_N	Replication protein A	PLN02850, PTZ0040, aspS_nondisc, AsnS, aspC
6	AZ254277.1	516	TVP38	SNARE_assoc	SNARE_assoc
7	AZ254281.1	506	LRR_4	LRR_4	LRR_8
8	AZ254291.1	360	SANT	SANT	SANT, PLN03212
9	AZ254293.1	532	LRR_4	LRR_RI, LRR_4	LRR_8

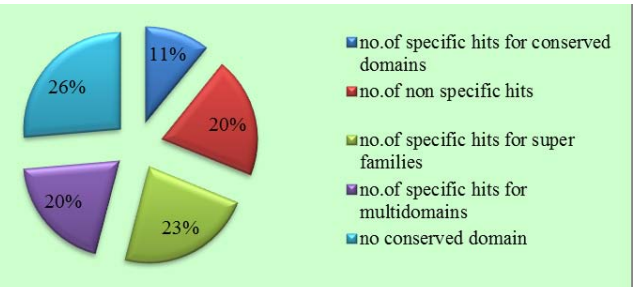

Figure 1: Distribution of 82 sequences for CDD search.

Figure 2: Distribution of 82 sequences for CD LRR_4 from AZ254242.1 GS query sequence.

vesicle transfer processes at the Golgi membrane. Members of the so-called "SNARE-associated proteins of the Tvp38-family" have also been identified in prokaryotes and those belong to the DedA protein family. Tvp38/DedA proteins are also conserved in cyanobacteria and chloroplasts. While only a single member of this family appears to be present in chloroplasts, cyanobacterial genomes typically encode multiple homologous proteins. Mainly based on our understanding of the DedA-homologous proteins of Escherichia coli, it appears likely that the function of these proteins in chloroplast and cyanobacteria involves stabilizing and organizing the structure of internal membrane systems. Another domain name SANT (AZ254291.1 table 1) found in regulatory transcriptional repressor complexes where it also binds DNA. Based on the putative function of the conserved domains for the gene or gene families identified from the nucleotide query sequence, they could be used for the identification of gene-targeted markers (GTMs) (Varshney and Tuberosa, 2007) and for the development of Functional Markers (FM) (Andersen and Lu" bberstedt, 2003). After validation, these markers can be used in the Marker Assisted Breeding (MAB) while transferring the gene of interest from wild species under the cultivated background of Green gram. In the present analysis, it has been observed that most of the identified CD regions are common in large groups of living beings *viz.,* rice, maize, human, arabidopsis, tortula, fungus, bacteria, protozoan, drosophila etc. Thus during the evolution process, these sequences have been highly conserved in nature among different living organisms and additional information can be generated for the identified CDs or gene families or for the gene of interest by means of comparative genomics tools. The generated information for the genes of interest can be utilized for the improvement of Green gram and this new approach may help the plant breeders for identification of functional markers in crop plants.

Conclusion

It with From this study we observed the various conserved domains, super families having putative functions for fundamental, metabolic, developmental, evolutionary processes and physiogenic nature have been identified from the query sequences which can be further utilized for the development of functional markers in green gram and its related crop plants.

Acknowledgements

We express our gratitude to the Department of Biotechnology, Government of India for providing fellowship to the seniour author and to the Professor Jayashankar Telangana State, Agricultural University, Hyderabad for providing the opportunity to work on the above topic.

References

1. Afzal AJ, Wood AJ, Lightfoot DA. Plant receptor-like serine threonine kinases: Roles in signaling and plant defense. Molecular Plant-Microbe Interactions. 2008;21(5):507–517. doi: 10.1094/MPMI-21-5-0507.
2. Baumgarten A, Cannon S, Spangler R, May G. Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics. 2003;165(1):309–319.
3. Bhawna G, Bonthala VS, Gajula MNVP. PvTFDB: a Phaseolus vulgaris transcription factors database for expediting functional genomics in Legumes. Database (Oxford). 2016;2016:baw114. doi: 10.1093/database/baw114.
4. Bhawna G, Chaduvula PK, Suresh V, Siddiq EA, Polumetla AK, MNV Prasad Gajula MNV . CmMDb: A versatile database for Cucumis melo microsatellite markers and other horticulture crop research. PLOS ONE. 2015;10(4):e0118630. doi:10.1371/ journal.pone.0118630.
5. Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, McAdams SA, et al. tA single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi-ta. Plant Cell. 2000;12(11): 2033–2046.
6. Couch BC, Spangler R, Ramos C, May G. Pervasive purifying selection characterizes the evolution of I2 homologs. Molecular Plant-Microbe Interactions. 2006;19(3):288–303.
7. DeYoung BJ, Innes RW. Plant NBS-LRR proteins in pathogen sensing and host defense. Nature Immunology. 2006;7(12): 1243–1249.
8. Di Matteo A, Bonivento D, Tsernoglou D, Federici L, Cervone F. Polygalacturonase inhibiting protein (PGIP) in plant defence: A structural view. Phytochemistry. 2006;67(6):528–533. DOI:10.1016/j.phytochem.2005.12.025.
9. Di C, Zhang M, Xu S, Cheng T, An L. Role of poly-galacturonase inhibiting protein in plant defense. Critical Reviews in Microbiology. 2006;32(2): 91–100.
10. Dievart A, Clark SE. LRR-containing receptors regulating plant development and defense. Development. 2004;131(2): 251–261.
11. Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JD. The tomato Cf-5 disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell. 1998;10(11):1915–1925.
12. Dodds PN, Lawrence GJ, Catanzariti AM, The T, Wang CIA, Ayliffe MA, et al. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proceedings of the National Academy of Science USA. 2006;103(23):8888–8893.
13. Dudhe MY, Meena HP, Ranganatha APR, Muktha N, Lavanya C. Insilco identification of conserved domains from EST database in Safflower. Journal of oilseeds research. 2012;29:178-181.

14. Friedman AR, Baker BJ. The evolution of resistance genes in multi-protein plant resistance systems. Current Opinion in Genetics and Development. 2007;17(6):493–499.

15. Gajula MP, Kumar A, Ijaq J. Protocol for Molecular Dynamics Simulations of Proteins. Bio-protocol. 2016;6(23): e2051. DOI: 10.21769/BioProtoc.2051

16. Gish LA, Clark SE. The RLK/Pelle family of kinases. Plant Journal. 2011;66(1): 117–127. doi: 10.1111/j.1365-313X.2011.04518.x.

17. Hulbert SH, Webb CA, Smith SM, Sun Q. Resistance gene complexes: Evolution and utilization. Annual Review of Phytopathology. 2012;39:285–312.

18. Jaillais Y, Belkhadir Y, Balsemao-Pires E, Dangl JL, Chory J. Extracellular leucine-rich repeats as a platform for receptor/coreceptor complex formation. Proceedings of the National Academy of Science USA. 2011;108(20): 8503–8507. doi: 10.1073/pnas.1103556108.

19. Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO Journal. 2000;19(15):4004–4014.

20. Jones D, Jones J. The role of leucine-rich repeat proteins in plant defenses. Advances in Botanical Research. 1997;24: 89–167.

21. Kumar A, Kumar S, Kumar U, Suravajhala P, MNVP Gajula. Functional and structural insights into novel DREB1A transcription factors in common wheat (Triticum aestivum L.): A molecular modeling approach. Computational Biology and Chemistry. 2016;64: 217-226.

22. Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. Trends in Genetics. 2004;20(3): 116–122.

23. Liu J, Liu X, Dai L, Wang G. Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. Journal of Genetics and Genomics. 2007;34(9): 765–776.

24. Marchler-Bauer, Aron, Panchenko, Anna R, Shoemaker, Benjamin A, Thiessen Paul A, et al. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Research. 2002;30(1): 281-283.

25. McDowell JM, Simon SA. Molecular diversity at the plant-pathogen interface. Developmental and Comparative Immunology. 2008;32(7):736–744. doi: 10.1016/j.dci.2007.11.005.

26. McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: Adaptable guards. Genome Biology. 2006;7(4):212.

27. Michelmore RW, Meyers, BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research. 1998;8(11):1113–1130.

28. MNVP Gajula, Vanisree SG, Dastagiri MB. Outlook on application of bioinformatics in agriculture. World Research Journal of Bioinformatics. 2014;2(1):33-40.

29. MNV Prasad Gajula. Multi-Omics Data Integration: A Modular Approach. J Mol Genet Med. 2016;10(4):232 doi:10.4172/1747-0862.1000232

30. Mondragon-Palomino M, Gaut BS. Gene conversion and the evolution of three leucine-rich repeat gene families in Arabidopsis thaliana. Molecular Biology and Evolution. 2005;22(12): 2444–2456.

31. Nagar Laxman, Kumar Anuj Y, Vimala, MNV Prasad Gajula. Sequence to Structure Analysis of DOPA Protein from Mucuna pruriens: A Computational Biology Approach. International Journal of Emerging Trends & Technology in Computer Science. 2016;2(8): 3083-3089. DOI: 10.18535/ijetst/v2i8.12

32. Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, et al. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell. 1997;91:821–832.

33. Ramprasad E, Rakesh G, Durga Rani Ch V, Vanisri S, MNV Prasad Gajula. In Silico Analysis of Chalcone Synthase 1 Protein Sequences from Different Plant Species. International Journal of Science, Environment and Technology. 2016;5 (4):1968–1979.

34. Stitzinger SM, Pellicena-Palle A, Albrecht EB, Gajewski KM, Beckingham KM, Salz HK. Mutations in the predicted aspartyl tRNA synthetase of Drosophila are lethal and function as dosage-sensitive maternal modifiers of the sex determination gene Sex-lethal. Molecular Genetics and Genomics. 1999;261(1):142-151.

35. Tor M, Lotze MT, Holton N. Receptor-mediated signalling in plants: Molecular patterns and programmes. Journal of Experimental Botany. 2009;60(13):3645–3654.

36. Wang G, Fiers M. Receptor-like proteins: Searching for functions. Plant Signaling and Behavior. 2010;5(5):540-542.

37. Wei F, Wing RA, Wise RP. Genome dynamics and evolution of the Mla (powdery mildew) resistance locus in barley. Plant Cell. 2002;14(8):1903–1917.

38. Wootton JC, Federhen S. Statistics of Local Complexity in Amino Acid Sequences and Sequence Databases Computer. Chemistry. 1993;17(2):149-163. DOI: 10.1016/0097-8485(93)85006-X

39. Wulff BB, Chakrabarti A, Jones DA. Recognitional specificity and evolution in the tomato Cladosporium fulvum pathosystem. Molecular Plant-Microbe Interactions. 2009;22(10):1191–1202. doi: 10.1094/MPMI-22-10-1191.

40. Zhou B, Dolan M, Sakai H, Wang GL. The genomic dynamics and evolutionary mechanism of the Pi2/9 locus in rice. Molecular Plant-Microbe Interactions. 2007;20:63–71.