

Isolation and Characterization of CBP-Coding Gene *NPGR2* from *Dimocarpus longan*

Min Kyaw Thu^{1,2,*}, Qilin Tian¹, Manman Yang¹, Yukun Chen¹, Xin Feng¹, Ruilian Lai¹ and Zhongxiong Lai^{1,*}

Abstract

In many cellular processes, Ca^{2+} and Calmodulin (CaM) play essential roles. The CaM-binding proteins (CBP) are also involved in important role in those processes. For the better understanding of the characteristics of the CBP-coding gene, we isolated a CBP-coding gene *NPGR2* (*No Pollen Germination Related 2*) from embryogenic callus (EC) of longan. After isolation of the gene, it was cloned and then was bioinformatically analyzed. Several bioinformatics tools such as DNAMAN, ProtParam, Euk-mPLoc 2.0, SignalP, COILS, TMPRED, MEGA 6.06, SMART and NetWheels were used to perform bioinformatic analysis. The full length cDNA of *DINPGR2* was 2,448 bp and Open Reading Frame (ORF) was 1,389 bp. coding 462 amino acids. The molecular weight was 51.66 kDa and theoretical isoelectric point (pI) was 7.07. It was not a transmembrane protein. The predicted cellular localization of the protein was nucleus. *DINPGR2* possessed six TPR (Tetratricopeptide Repeat) and one CBD (Calmodulin-Binding Domain). In phylogenetic tree, *DINPGR2* was grouped with other *NPGR2* proteins from different plant species. Taking together, it was concluded that *DINPGR2* might be a CBP-coding gene with six TPRs and one CBD.

Keywords: Bioinformatic characterization, *Dimocarpus longan*, *No Pollen Germination Related 2* (*NPGR2*), Calmodulin-binding protein (CBP), Tetratricopeptide Repeat (TPR)

Introduction

Secondary messengers such as cyclic AMP (cAMP), cyclic GMP (cGMP), diacylglycerol (DAG), inositol-1,4,5-triphosphate (IP3), and calcium (Ca^{2+}) play important roles in plant signal transduction (Srivastava, 2002). Among them, Ca^{2+} is a common second messenger in plant cell signaling. A calcium-modulated protein, calmodulin (CaM), is a calcium-binding protein which plays essential role in cellular signaling pathways via interaction with many target proteins (Ranty et al., 2006). There are CaM-binding proteins (CBP) in the cytoplasm, and they are involved in many cellular processes. In order to extend our knowledge related to CBP, in this study, we isolated a CBP-coding gene *NPGR2* (*No Pollen Germination Related 2*) from embryogenic callus (EC) of longan. Then, we sequenced it and bioinformatically analyzed. This study might help the researchers who work in further studies upon the gene such as expression pattern of the gene under different treatments.

¹Institute of Horticultural Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

²Department of Botany, University of Yangon, Yangon 11041, Yangon Region, Myanmar

*Corresponding authors:

Materials and Methods

Plant tissue sample

We cultured embryogenic callus (EC) from *Dimocarpus longan* cv. Honghezi Lour. growing in the garden of Fujian Agriculture and Forestry University, Fuzhou, Fujian, China by using the method reported by Lai and Chen (1997) and Lai et al. (2000).

Isolation of nucleic acid, cloning and synthesis of cDNA

We isolated total RNA from longan EC by using TriPure isolation reagent (Roche Diagnostics GmbH, Mannheim, Germany) and other supplementary chemicals. Using 0.8% agarose gel electrophoresis and a NanoDrop Lite spectrophotometer (Thermo Fisher Scientific Inc., Madison, WI, USA) the quality and concentration of the isolated total RNA was analyzed. Then, it was reverse-transcribed into complementary DNA (cDNA) using a GeneRacer™ Kit (Invitrogen). In order to isolate the interested gene longan NPGR2, we designed the gene specific primers based on the longan genomic data (Lin et al., 2017) and our longan transcriptome data (NCBI SRA database accession number SRA315202 (Lin et al., 2017)). We cloned the full-length cDNA of longan *NPGR2* using PCR and RACE methods. The PCR products with expected size were cut from the gel, purified and sub-cloned into a Blank-5 vector (TransGen Biotech, China). In order to confirm the product, it was finally sequenced.

Table 1. The list of primers used to clone full length of longan *NPGR2* cDNA

Primer	Sequence (5' to 3')	Purpose
<i>DINPGR2</i> -ORF-F	AAACTTGCTGATTGTCCACG	ORF amplification
<i>DINPGR2</i> -ORF-R	GGGATTGGTCATCTGAAGG	
GeneRacer™ 5' Primer (5P)	CGACTGGAGCACGAGGACACTGA	1 st round of 5'
<i>DINPGR2</i> -5' RACE GSP-1	CCGTCTTGTGGCCCTCTCAAACCTCAGC	RACE
GeneRacer™ 5' Nested (5NP)	GGACACTGACATGGACTGAAGGAGTA	2 nd round of 5'
<i>DINPGR2</i> -5' RACE GSP-2	CAGCATCAAGATTCCAATAATGAAGG	RACE
<i>DINPGR2</i> -3' RACE GSP-1	GGAATTGGATGTGTGGCATGATCTGGC	1 st round of 3'
GeneRacer™ 3' Primer (3P)	GCTGTCAACGATACGCTACGTAAACG	RACE
<i>DINPGR2</i> -3' RACE GSP-2	TAATCTCGGCCTGCTTCACATATCC	2 nd round of 3'
GeneRacer™ 3' Nested (3NP)	CGCTACGTAACGGCATGACAGTG	RACE

Bioinformatic analysis of the sequence

The software DNAMAN (v.6.0, Lynnon Corporation, Quebec) was used to get deduced amino acid sequence from *DINPGR2*. Moreover, we used several bioinformatic online tools to analyze the sequence. In order to determine the physical and chemical parameters, we used ProtParam (<http://web.expasy.org/protparam/>). We also used an online tool Euk-mPLOC 2.0 (<http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/>) for the prediction of subcellular localization of *DINPGR2*. Another two online software, SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) and TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html) were used to predict the signal peptide and transmembrane topology, respectively. Thirty homologous amino acid sequences were downloaded from NCBI database and TAIR database. Then, we constructed an unrooted phylogenetic tree using the neighbor-joining method with MEGA 6.06 software (<http://www.megasoftware.net/>). Simple Modular Architecture Research Tool (SMART) (<http://smart.embl-heidelberg.de/>) (Schultz et al., 1998) was

used to predict the number and position of tetratricopeptide repeat (TPR) domains. We identified CBD by performing multiple alignment the sequence with Arabidopsis version of NPGR2. Moreover, we used NetWheels program (<http://lbqp.unb.br/NetWheels/>) to analyze CBD.

Results

DINPGR2 cDNA

We obtained the *DINPGR2* with full length of 2,448 bp and Open Reading Frame (ORF) was 1,389 bp. coding 462 amino acids. Then, the full length of the cDNA was deposited at NCBI (GenBank accession number KP402185.1) (Fig. 1).

Dimocarpus longan cultivar Honghezi No pollen germination related 2 protein mRNA, complete cds

GenBank: KP402185.1

>KP402185.1 *Dimocarpus longan* cultivar Honghezi No pollen germination related 2 protein mRNA, complete cds

```
TTGCTTGAAGATGTGAAAAGATATGAAGAGTAAAAATAAGATAATTTGGCAGAGTAGAAGTGGAAGTAAGCA
GAGTTTTTTTTGAAAAATAATGAATTGCGTTTGCTCAGGAGAGCAATTGGAAGCAGTAGATGAAATGGTACC
ATCGTCTGAATCCCGTGCACAGGGATTATTATTCGGTGGGTGGTCATTCTCTAGACCTGGTGATGTTGA
GAAGCAGCAGCCAGATAACGGCAATATAGAAGAAGCCGAGTGGTCTCTGCGAGAGACTAGTTCTTTGAATTA
TGAGGAAGCAAGAGCATTGTTAGGAAAAATCGAATATCAGAAAGGAAATATAGAAGCTGCTCTTCATGTATT
TGAAGGAATAGATATAGCTGCAGCAACTCCAAAAATGAAAACAACACTTGCTAGTAAAAGAGTAGAACGGCA
TAGAAGGCATTCTCACAATTATGCTTCACCACCAATGTCTATCCATACTGTCAGTTTGCTCTTGGAAGCAAT
CTTTCTTAAAGCTAAATCATTGCAGGGTCTGGGGAGATTCAAAGAAGCTGCACAATCTTGTAAGATTATTCT
GGACATTCTAGAATCTTCATTGCCAGAAGGCCTACCTGAAAACCTTGGTGCTGATTGTAAATTGCAGGAGAC
TCTGAACAAAGCTGTTGAACTGCTTCCAGAATTGTGGAACCTTGCTGATTGTCCACGTGAGGCAATCTTGTC
ATACCGGGCGGGCCCTCCCTTCATTATTGGAATCTTGATGCTGAAATGACTGCAAGGATTCAAAAAGAATTTGC
TGTTTTTTTATTGTATTGTGGAGGTGAAGCAAGCCCCCAAACCTTCGCTCTCAAATGGACAGTGCATTTGTT
CCCAGAAAACAACATTGAAGAGGCTATACCTTTTGTTAATGATTCTGTAAAGAAAAATTTCTCTCAGAAGAATT
GAGTGGGATCCATCAATTTTGAATCACCTTTTCATTTGCTCTTTCTATTTCGGGTGATATGAGGTCTTTGGCT
ACTCGGATTGAAGAACTTCTCTCGGGATTATTAATCGGAAAGAAAGGTATTACCTTCTAGCTCTCTGTTAC
CATGGAGCAGGTGAGGACCAGGTGCTTTGAATTTGCTGAGAAAATTTGTTGAGTAGCGGAGAGGATCCAAAT
TGTGTTCCAGCTTTATTGATGGCGTCACAGCTTTGTGGGGAGAAGCCTGATCTTGACAGAGGAAGGAATCGGT
TTTGCTCAGAGGGCCCTTGAATGCTTGGACGGCGGATGTGATCAGTTAGATGGTACTGTGAACCTCTTGTTG
GGCGTATCACTTTCAACACAATCTAAATTAGAGGTTGCTGAGTTTGAGAGGGCCACAAGACGGTCTCAGGCA
CTTCAGGCTTTGCAAAGTGCTGCAATAAGTACAGGAATGAAAGAACCAATATCTTTACCACCTTAGTCTA
GAATGTGCCGAGCAGAGAAAGTTGGATGCAGCACTTTATTATGCAAAGTCTTTGTTAAACTAGAAAGATGAT
TCTAATGTTAAAGGATGGTTATTGATAGCCCGGATGTCAGCTCAGAAACGGTTTGGGGATGCTGAAACT
GTTATAAATGCTGCTTTTAGAACAACTGGGAAATGGGATCAGGGAGAAATTAATGCGAACAAAGCTAAGCTT
CAAATTGCACAGGGTCAGTTAAAGAGTGCCATGGAGACATATACTCAACTTCTCGCAGTTTTTCAAGTCCAG
AGTAAAACCTTTGTTACTGGGAGGAAGTTTCGTAAGGACTGTGCAAATCCTCCTAGAAGTTTGGAATTGGAT
GTGTGGCATGATCTGGCATTGTCTTCATAAAGCTGTACAGTGGCATGATGCTGAGATTTGTCTTTCCAAA
TCTAAGGCCATCAGTGCTTACTCTGCTACAAGATGTCATGCATATGGTGTATTATGAAAAAAGGGCCTC
TTCAAAGAAGCTTTACAAGCCTTCAGGAGAGCTTGGGATATTGATCCCAACCATGTACCAAGTTTGATATCC
ATAGCTTTGGTTCTAAGACGGCTTGGGGGGCAATTTGCAGTTATAAGAACTTTTGTGATGGCTGCTCGTCAA
CTTGATAGAATGAATGCATCTGCCTGGTATAATCTCGGCCTGCTTCACATATCCGAGGGTACTCAATCAGCA
AACTAGAAGCAGCTGAATGTTTGCAGGTTGCATACCTTCTTGAAGAGACAACACAGTTGAACCCTTCAGA
TGAACCAATCCCCGCACTTTTTGGTCTATGTAACGTAAGCCTTGAAAAATTTTTTTTACTATCCTCTACAA
ACATGTATCTTATAAATTATATGGTCCATACACATTGTATAACGAATTGAAATATTAATAAGAAGGGAAAAA
AAAAAAGGAAAAGGAAAAAAAACAGTTTCTTTCTTGATGTGTATGAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 1. The full length cDNA of *DINPGR2*. Highlighted codons ATG and TGA are start codon and stop codon, respectively.

Bioinformatic analysis of *DINPGR2*

The predicted molecular weight of deduced polypeptide *DINPGR2* was 51.66 kDa and theoretical isoelectric point (pI) was 7.07. The instability index (II) is

computed to be 49.18. It was not a transmembrane protein and has no signal peptide. The predicted subcellular localization of the protein was nucleus.

MDSAFVPRNNIEEAILLLMILLRKISLRRIEWDPSILNHLISFALSISGDMRSLATRIEELLPGIINRKERY
LLALCYHGAGEDQVALNLLRKLSSGEDPNCVPALLMASQLCGEKPDLAEEGIGFAQRALECLDGGCDQLDG
TVNFLGLVSLSTQSKLEVAEFERATRRSQALQALQSAAISTGMKEPNILYHLSLECAEQRKLDAAALYYAKSL
LKLEDDSNVKGWLLIARILSAQKRFGDAETVINAALQETGKWDQGEIMRTKAKLQIAQGQLKSAMETYTQLL
AVFQVQSKTLVTGRKFRKDCANPPRSLELDVWHDLAFFVFIKLSQWHDAEICLSKSKAISAYSATRCHAYGVI
YEKKGLFKEALQAFRRAWDIDPNHVPISLISIALVLRRLGGQFAVIRNFLMAARQLDRMNASAWYNLGLLHIS
EGTQSAKLEAAECLQVAYLLEETTPVEPFR

Fig. 2. The deduced polypeptide DINPGR2. The putative CBD (NVKGWLLIARILSAQKRF) is highlighted.

There were three distinct clusters, namely, NPG1, NPGR1 and NPGR2. The DINPGR2 was closely related with the homologous proteins of *Pistacia vera* and *Citrus clementina* (Fig. 3). A putative CBD (NVKGWLLIARILSAQKRF) of DINPGR2 between 223 and 242 amino acid residues was identified by comparing it with CBD sequence of AtNPGR2 (DLEVWLLARVLSAQKRF) where the identity was 66.67%.

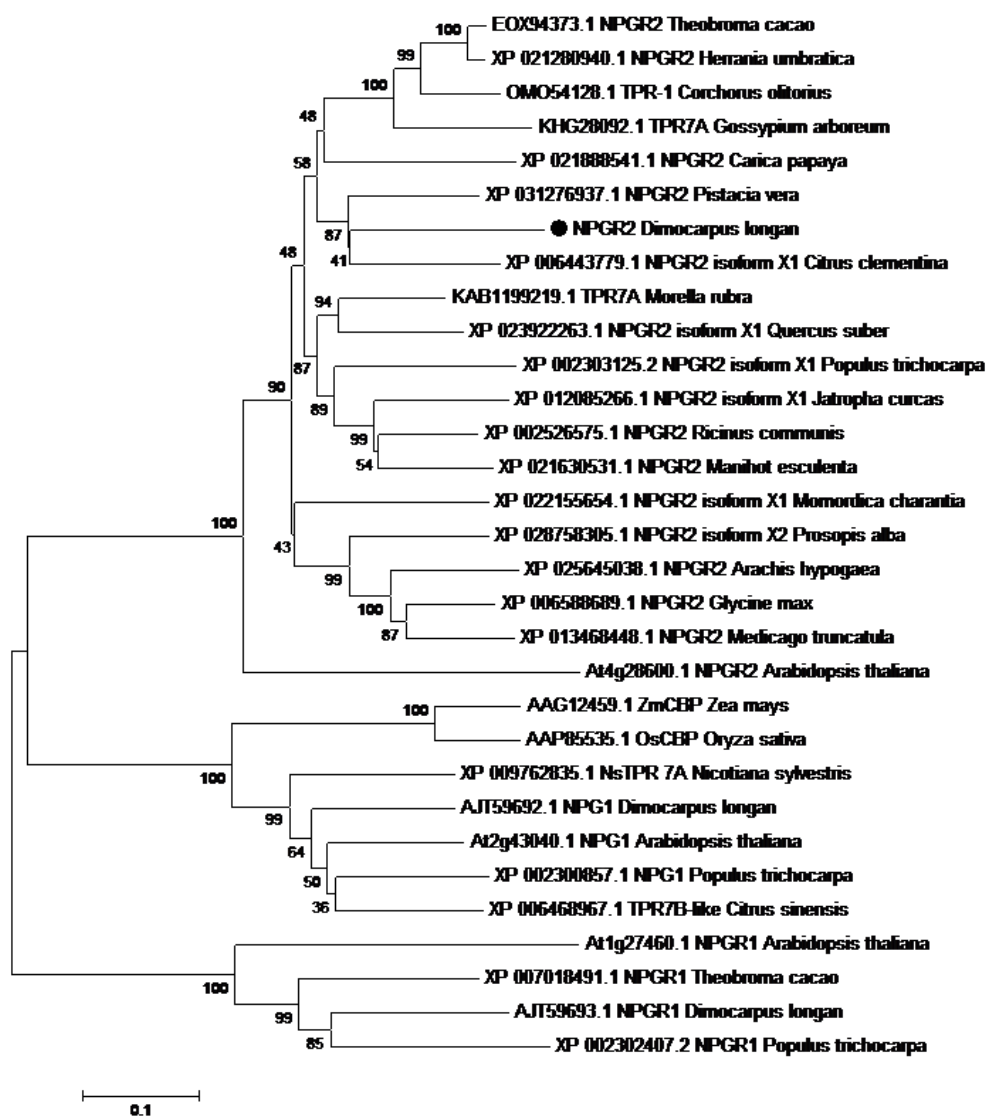


Fig. 3. An unrooted phylogenetic tree of 31 proteins was related to No Pollen Germination 1 (NPG1) using the neighbor-joining method. The protein sequences were

downloaded from the NCBI and TAIR (their accession numbers or locus IDs are given in brackets). Number at each node represents the bootstrap values based on 1000 replications.

The position of the polar and non-polar amino acid residues was predicted by contracting helical wheel structure of CBD domain (Fig. 4). Moreover, it was predicted that DINPGR2 possesses 6 TPR domains (Fig. 4).

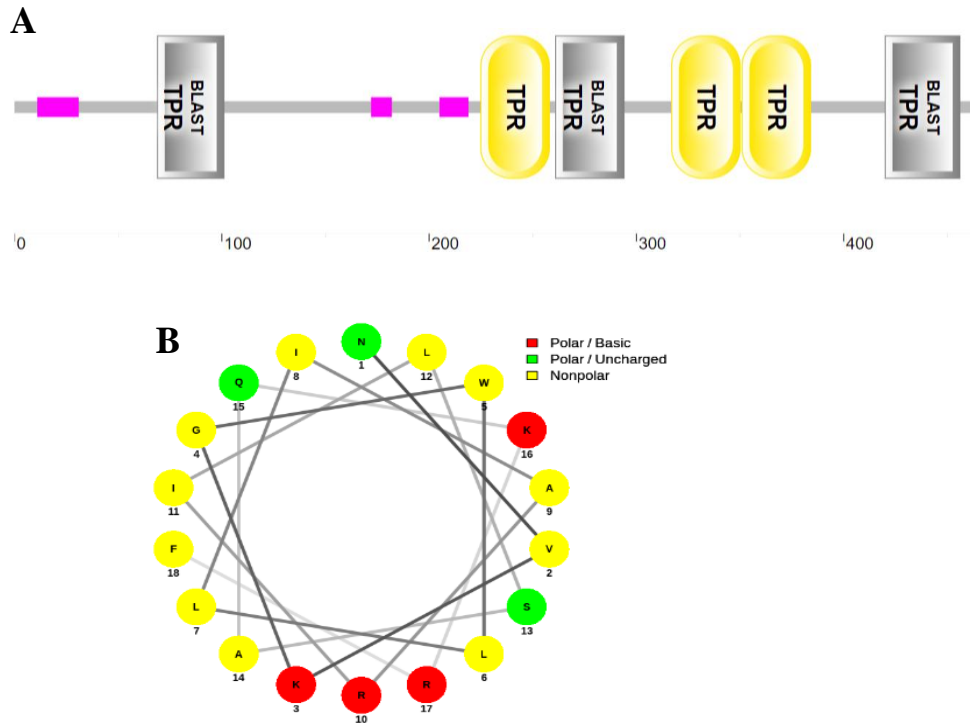


Fig. 4. Domains of DINPGR2. (A) Six tetratricopeptide repeat (TPR) domains were predicted by SMART. (B) Prediction of Calmodulin-binding domain (CBD) of DINPGR2 (NVKGWLLIARILSAQKRF) between amino acid residues 223 and 242 by a helical wheel program.

DINPGR2 showed high identity with the closely homologous proteins from other plant species namely *Pistacia vera*, *Citrus clementina*, *Theobroma cacao* and *Herrania umbratica* (Table 2).

Table 2. Identity of the DINPGR2 with closely homologous proteins from other plant species

	<i>Pistacia vera</i> NPGR2 [XP_031276937.1]	<i>Citrus clementina</i> NPGR2 isoform X1 [XP_006443779.1]	<i>Theobroma cacao</i> NPGR2 [EOX94373.1]	<i>Herrania umbratica</i> NPGR2 [XP_021280940.1]
DINPGR2	73.65%	72.35%	71.24%	70.63%

Discussion

According to the predicted instability index, the protein was unstable. Base to the prediction, DINPGR2 possessed one CBD. It has already proved that the Arabidopsis AtNPG1 polypeptide binds CaM in a calcium-dependent manner (Golovkin & Reddy, 2003). Considering altogether, DINPGR2 might involve in the process of cell regulation by Ca^{2+} /CaM in plants. Moreover, DINPGR2 also has six

TPR domains suggesting that DINPGR2 might interact with different kinds of protein through those TPR domains.

One group of the proteins namely CBP binds to calmodulin in a Ca^{2+} -dependent and reversible manner and are involved in multitude of processes in which Ca^{2+} and calmodulin play crucial roles (Sharma & Parameswaran, 2018).

According to numerous studies, the formation of protein-protein complexes is involved in many biological functions. In fact, protein-protein interactions can be regarded as the hub of all living processes (Zeytuni & Zarivach, 2012). TPR-containing proteins is a common group of proteins. Those are involved in protein-protein interaction. TPR domain is widespread in evolution, and it can be found from bacteria to human, and in various subcellular locations. TPR-containing proteins are involved in many cellular processes such as cell cycle and stress response. Altogether, it seems that there is no common cellular process connecting to TPR-containing proteins (Lamb et al., 1995). Therefore, longan version of TPR-containing protein DINPGR2 was also one of the important proteins which involved in many cellular processes.

In conclusion, the high identity of the DINPGR2 with homologous TPR-containing CBP proteins from other plant species suggesting that it belonged to the plant TPR-containing CBP protein family.

Acknowledgments

We would like to thank our lab members who helped us during this study. Special thanks go to the Research Funds for the National Natural Science Foundation of China (31572088 and 31672127), Science and Technology Plan Major Projects of Fujian Province (2015NZ0002-1), and Horticultural Postdoctoral Research Funding of Fujian Agriculture and Forestry University (150797).

References

- Day IS and Reddy ASN. 2011. Elucidation of Calcium-Signaling Components and Networks. Pages 147–175: In Sheng Luan, ed, *Coding and Decoding of Calcium Signals in Plants*. Springer-Verlag Berlin Heidelberg, NY.
- Golovkin M and Reddy AS. 2003. A calmodulin-binding protein from Arabidopsis has an essential role in pollen germination. *Proc Natl Acad Sci U S A*, 100: 10558–10563.
- Lai Z and Chen Z. 1997. Somatic embryogenesis of high frequency from longan embryogenic calli. *J Fujian Agric Univ.*, 26: 271–6.
- Lai Z, Chen C, Zeng L and Chen Z. 2000. Somatic embryogenesis in longan (*Dimocarpus longan* Lour.). In: Jain SM, Gupta P, Newton R, editors. *Somatic Embryogenesis in Woody Plants*. Netherlands: Springer, 67: 415–31.
- Lamb JR, Tugendreich S, Hieter P. 1995. Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci*, 20:257–259.
- Lin Y, Min J, Lai R, Wu Z, Chen Y, Yu L, Cheng C, Jin Y, Tian Q, Liu Q, Liu W, Zhang C, Lin L, Zhang D, Thu M, Zhang Z, Liu S, Zhong C, Fang X, Wang J, Yang H, Varshney RK, Yin Y and Lai Z. 2017. Genome-wide sequencing of longan (*Dimocarpus longan* Lour.) provides insights into molecular basis of its polyphenol-rich characteristics. *GigaScience*, 6:1–14: doi: 10.1093/gigascience/gix023
- Ranty B, Aldon D and Galaud JP. 2006. Plant calmodulins and calmodulin-related proteins: multifaceted relays to decode calcium signals. *Plant Signal. Behav.* 1:96. doi: 10.4161/psb.1.3.2998
- Schultz J, Milpetz F, Bork P and Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A*, 95: 5857–64.
- Sharma RK and Parameswaran S. 2018. Calmodulin-binding proteins: A journey of 40 years. *Cell Calcium*, 75:89-100. doi: 10.1016/j.ceca.2018.09.002.
- Srivastava LM. 2002. *Plant Growth and Development: Hormones and Environment*. Academic Press. London. ISBN: 012660570X, 9780126605709
- Zeytuni N and Zarivach R. 2012. The TPR Motif as a Protein Interaction Module - A Discussion of Structure and Function, *Protein Interactions*. Pages 103-118: In J Cai and RE Wang ed, *Protein Interactions*, InTech, Rijeka, Croatia.