Australian Journal of Crop Science AJCS
AJCS 7(12):1884-1892 (2013) ISSN:1835-2707

Molecular cloning, characterization and bacterial overexpression of D-*myo*-inositol 3-phosphate synthase (*MIPS1*) gene from soybean (*Glycine max* [L.] Merr.)

Swati Kumari and Archana Sachdev*

Division of Biochemistry, Indian Agricultural Research Institute, New Delhi 110012, India

***Corresponding author: arcs_bio@yahoo.com**

Abstract

D-*myo*-inositol-3-phosphate synthase (EC 5.5.1.4; MIPS) is an isomerase that catalyzes the conversion of glucose-6-phosphate to D-*myo*-inositol-3-phosphate, a sole synthetic source of *myo*-inositol. Phytic acid (*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate) which is the principal storage form of phosphorus (60-80%) in plant seeds, is further generated by a stepwise phosphorylation of *myo*-inositol. Poorly digested by monogastrics, phytic acid, chelates essential mineral cations and proteins; thereby, reducing their bioavailability and hence classified as an anti-nutrient. In soybean, the transcripts encoding *MIPS1* isoform are expressed early during the cotyledonary stage of seed development to function in phytic acid biosynthesis. In the present study, we report the cloning and molecular characterization of the *MIPS1* gene from developing seeds of soybean (*GmMIPS1*). A full-length *GmMIPS1* cDNA (*Glycine max* cv. Pusa 16) of 1,791bp revealed an ORF of 1,533 bp predicting 510 amino acids. *In silico* analysis further revealed its striking homology (80-99%) at both the nucleotide and the deduced amino acid sequence levels; with other plant *MIPS*, particularly with the dicots, *Vigna radiata* and *Phaseolu svulgare*. The protein sequence analysis of the predicted *GmMIPS1* cDNA indicated the absence of a signal peptide in the N-terminal region. To validate the expression of the *GmMIPS1* coding gene, nucleotide sequence residues from 131 to 1,556 bp were amplified by high fidelity PCR and fused in frame to a 19 amino acid N-terminal region of 6X His-tag in expression vector pET-28a (+). The *E.coli* cell strain BL21 (DE3) transformed with the recombinant plasmid resulted in the production of a 52 kDa fusion protein under optimized induction and expression conditions as confirmed by SDS-PAGE and Western blot analysis. Results of the present study demonstrated that down-regulation of *GmMIPS1* using a seed specific promoter can be targeted as a great potential for developing of low-phytate soybean without affecting the critical aspects of inositol metabolism in other tissues of the plant.

Keywords: Anti-nutrient, *MIPS1*, phytic acid, prokaryotic expression vector, soybean.
Abbreviations: DAB_3,3′-diaminobenzidine tetrahydrochloride, IPTG_Isopropyl β-D-thiogalactopyranoside, *MIPS*_D-*myo*-inositol-3-phosphate synthase, ORF_open reading frame, PVDF_Polyvinylidenedifluoride, SDS-PAGE_Sodium Dodecyl Sulphate-Polyacrylamide gel Electrophoresis, UTR_untranslated region.

Introduction

Soybean (*Glycine max* [L.] Merr.), one of the world's most important economic crops, has a steadily increasing agronomical value because of its high protein and vegetable oil content suitable for human and animal nutrition. While soybean is an important source of protein, its potential to provide energy and minerals has not fully reached in non-ruminants animals including humans due to their inability to digest certain compounds such as phytates (Sebastian et al., 2000). Phytate (*myo*-inositol 1,2,3,4,5,6-hexa*kis*phosphate), also known as phytic acid (PA) or phytin, is a major form of phosphorus (P) storage in seeds, comprising over 75–80% of the total P in plant seeds (Cosgrove 1966; Raboy et al., 2001). In soybean seeds, phytic acid accounts for up to 2% of the seed dry weight (Raboy et al., 1984). It begins to accumulate in seeds after the cellular phosphate levels have reached maximum level and continues to increase linearly throughout seed development and seed filling (Raboy and Dickenson, 1987). It is usually deposited in protein bodies as a mixed salt (phytin), bound to mineral cations such as Fe^{3+}, Ca^{2+}, Mg^{2+}, Zn^{2+} and K^{+} (Prattley and Stanley 1982, Lott 1984). Additionally, stored phytate in seed is broken down by the enzyme phytase into inorganic phosphate and *myo*-inositol, which are then available for seedling growth. Although, it is an important storage molecule for growing seedlings, PA poses severe nutritional consequences because it acts as an antinutrient by forming indigestible complexes with minerals and proteins and decreases the seeds nutritional quality. It chelates mineral cations, including calcium, zinc, magnesium and iron from the diet and affects the bioavailability of these essential minerals (Raboy et al., 2001). It also has the potential to bind charged amino acid residues of proteins resulting in a concomitant reduction of protein availability and digestibility. This antinutritional quality of phytate can be further extended to human health as it contributes to the iron deficiency suffered by over 2 billion people worldwide (Bouis, 2000). Also the excretion of unused P in the waste makes its way into the waterways causing environmental hazards. The economic, nutritional, and environmental problems associated with phytate in animal or human feed can be reduced by developing low phytate soybean (Raboy, 2007). Therefore, the development of low phytic acid *(lpa*) crops is an important goal in genetic engineering programs aiming at improving the nutritional quality as well as at developing environment friendly and sustainable production. One of the approaches for reduction of plant seed phytate levels involves the reduction of the

expression of the rate limiting enzymes in the biosynthetic pathway of phytic acid. D-*myo*-inositol 3-phosphate synthase (MIPS, E.C. 5.5.1.4) catalyzing the NADH-dependent conversion of D-glucose 6-phosphate (G-6-P) to D-*myo*-inositol 3-phosphate (MIP) is the first and rate-limiting step of *myo*-inositol biosynthesis (Biswas et al., 1984; Loewus and Murthy, 2000). Further a stepwise phosphorylation of *myo*-inositol generates phytic acid. *MIPS* has been isolated and characterized from both prokaryotic and eukaryotic organisms. The structural gene coding for the *MIPS* was first identified in yeast (Donahue & Henry, 1981; Majumder et al., 1981). Subsequently, *MIPS* coding sequences have been cloned and characterized from widely different organisms, including plants such as *Spirodela polyrrhiza* (Smart & Fleming, 1993), *Citrus paradisi* (Abu-abied and Holland, 1994), *Arabidopsis thaliana* (Johnson, 1994; Johnson and Sussex, 1995), *Mesembryanthemum crystallinum* (Ishitani et al., 1996), wild halophytic rice *P. coarctata* (Majee et al., 2004), *Xerophyta viscosa* (Majee et al., 2005), *Passiflora edulis* (Abreu &Aragao, 2007), *Cicer arietinum* (Kaur et al., 2008) etc. Several plants have been found to possess multiple isoforms of MIPS enzyme, suggesting that each gene copy may be differentially controlled and expressed. Soybean contains four MIPS isoforms and one of the *MIPS* cDNAs (*GmMIPS1*) was shown to express mainly in developing seeds (Hegeman et al., 2001; Chappell et al., 2006). Using immunolocalization techniques, a specialized area of *GmMIPS-1* expression has been identified in the outer integumentary layer during early soybean seed development (Chiera and Grabau, 2007). A number of genes homologous to *GmMIPS1* have been reported till date and a "core catalytic structure" conserved across evolutionary divergent taxa has been identified (Majumdar et al., 2003). Down regulation of *MIPS* gene expression in seeds offer a potential approach for developing low-phytate soybean (Hitz and Sebastian, 1998).

In the present study, we report the isolation, cloning and characterization of full length *GmMIPS* cDNAs from developing seeds of soybean and validation of its expression in *Escherichia coli*, especially with respect to its involvement in phytic acid biosynthesis. The fully functional *GmMIPS1* gene can further be targeted for genetic manipulation by advanced gene silencing strategies to develop low phytate soybean seeds with improved nutritional value.

Results

Isolation of MIPS cDNA from soybean seeds

A full-length *GmMIPS* cDNA of 1,791 bp was amplified from the developing soybean seeds (4 to 6 mm) and cloned into pGEMT-Easy vector (Fig 1a.). The positive clones were confirmed by restriction analysis (Fig 1b.) with *Eco*RI and the reliability of the method was further confirmed by sequencing of the cloned products and entered into the NCBI GenBank database under the accession number HM46196. The *GmMIPS* cDNA containing an open reading frame (ORF) of 1,533 bp was predicted to encode a polypeptide of 510 amino acids with a 61-bp 5'- UTR and a 173-bp 3′-UTR.

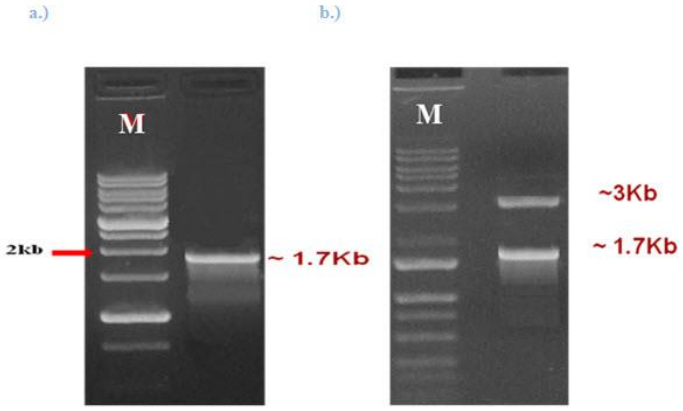
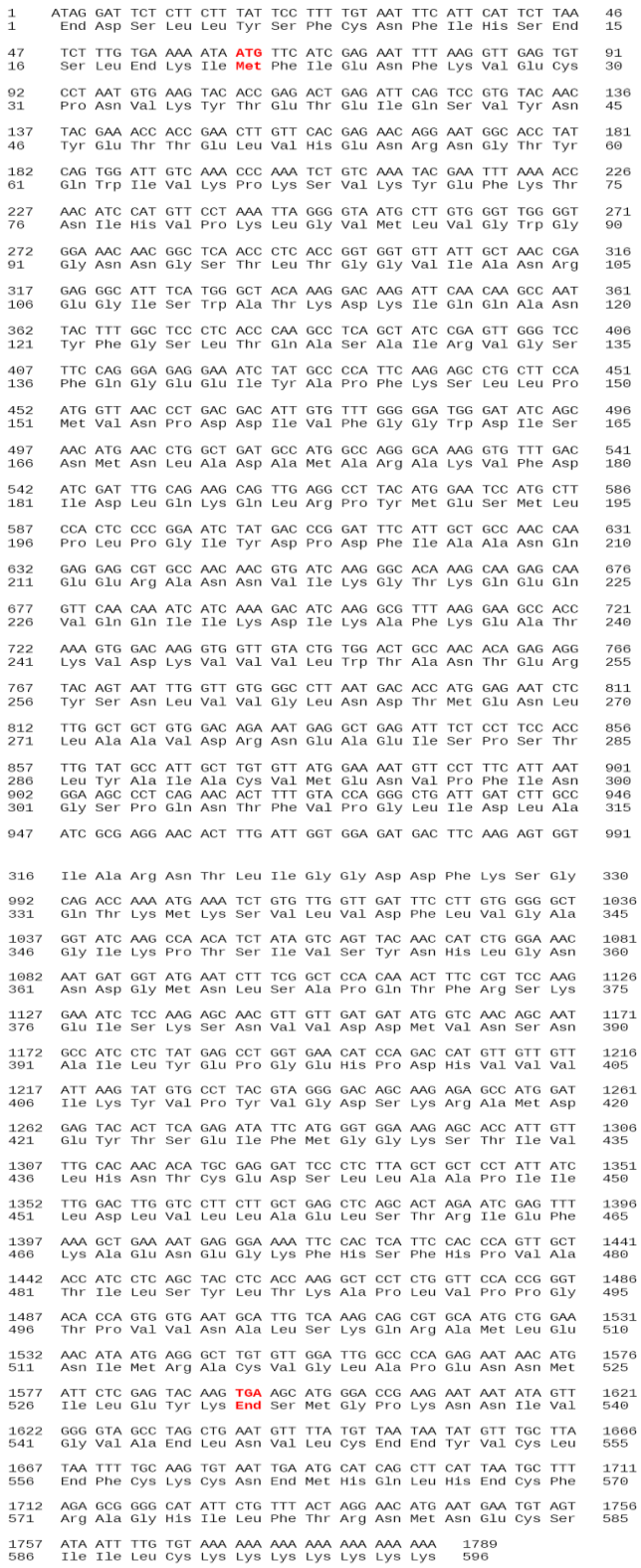


Fig 1. PCR amplification and cloning of full length MIPScDNA (~1.7 Kb) from developing soybean seeds. a.) PCR amplified target gene (1.7kb) against marker (M); b.) EcoRI digested product (1.7kb) from the recombinant pGEMT-Easy (3 kb) clones against 1Kb DNA marker (M).

Characterization of Glycine max MIPS cDNA(GmMIPS1)

The nucleotide sequence and deduced amino acid sequence of full-length *GmMIPS* cDNA are shown in Fig 2**.** The nucleotide sequence analysisusing the BLASTNsoftware at NCBI server (http://www.ncbi.nlm.nih.gov/BLAST) revealed 99% homology to the previously reported soybean *MIPS* gene (Accession No.AF293970)**.** Further BLASTP analysis revealed a high level sequence conservation with both monocots as well as dicots. Amongst dicots, highest homology scores were observed with *Vigna radiata*(96%), *Phaseolus vulgare* (95%) and *Cicer arietinum* (94%). SignalP program algorithms predicted no signal for chloroplast transit or mitochondrial targeting peptides in the N-terminal region of the *GmMIPS1*suggesting it to be a non-secretory protein**.** The nucleotide composition of 429 A (27.98%), 343 C (22.37%), 369 G (24.07%) and 392 T (25.57%) residues including 46.44% G+C and 53.56% A+T content was reflected in the sequence**.** The hydropathy plot of the GmMIPS protein sequence revealed the presence of about 250 hydrophilic amino acid residues (residues 200-450) close to the COOH-terminus**.** The hydropathy index thus predicted its hydrophilic nature which suggested it to be a cytosolic protein. The whole protein composition analysis of this protein revealed 37 Ala (7.3%), 14 Arg (2.7%), 38 Asn (7.5%), 25 Asp (4.9%), 4 Cys (0.8%), 17 Glm (3.3%), 34 Glu (6.7%), 33 Gly (6.5%), 8 His (1.6%), 37 Ile (7.3%), 41 Leu (8%), 33 Lys (6.5%), 17 Met (3.3%), 19 Phe (3.7%), 26 Pro (5.1%), 31 Ser (6.1%), 28 Thr (5.5%), 5 Trp (1%), 18 Tyr (3.5%) and 45 Val (8.8%) residues. Using the ProtParam package of ExPASy web server, the computation of various physical and chemical parameters of GmMIPS protein were carried out. The predicated molecular weight of the GmMIPS protein was 52 kDa with a theoretical pI of 5.31. A total number of negatively charged (aspartate and glutamate) and positively charged (arginine and lysine) residues predicted for GmMIPS were 59 and 47 respectively. Fairly high concentration of aspartate and glutamate residues as revealed in the amino acid composition of GmMIPS protein probably contributed to its theoretical pI of 5.31. The deduced formula of GmMIPS protein using this program was $C_{2527}H_{3985}N_{661}O_{761}S_{21}$.



Accession No. : HM461969

Length of ORF encoding soybean MIPS: 1533bp

Length of predicted protein: 510aa

Fig 2. Nucleotide sequence and deduced amino acid sequence of *Glycine max* full length (~ 1.7 kb) *MIPS1*cDNA submitted to GenBank (Accession No. : HM461969). The start (ATG) and end (TGA) points of coding sequence in the open reading frame (ORF) are coloured in red.

Total numbers of atoms present in the protein were 7955 and the extinction coefficient calculated was 54570 M^{-1} cm^{-1}. Instability index depicted a value of 32.77 which suggested the stable nature of the protein. Aliphatic index of this protein was 92.49 while the grand average of hydropathicity was -0.137 which confirmed the hydrophilic nature of protein. The predicted structural class of the protein was alpha. The high scoring amino acid sequences of MIPS proteins from eukaryotes were obtained from the NCBI using the BLAST program (Altschul et al., 1990) to generate a phylogenetic tree (Fig 3) by neighbor joining method in the MEGA 5 program (Tamura et al., 2011). Phylogenetic analysis divided the *MIPS* gene into clusters in accordance with taxonomic differentiation. Phylogenetic analysis showed three branches, with the main branch comprising of the sequences from dicotyledonous species. Surprisingly, *Triticum aestivum* rooted with the dicots. The branch of the dicotyledonous species further split into four conspicuous observable sub-branches. The main sub-branch comprised of species from the order Fabales (*Phaseolus vulgaris*, *Vigna radiata*, *Tripolium pretense* and*Cicer arietinum*). *Passiflora edulis* MIPS rooted with *N. tabacum*; both species were from different orders, Malpighiales and Solanales respectively. The minor sub-branch comprised of orders Brassicales (*Arabidopsis thaliana* and *Brassica napus*). The second main branch however encompassed the sequences of monocots (*Zea mays, Oryza sativa, Hordeumvulgare* and *Avena sativa). Brassica juncea* from Brassicales rooted with monocots. The third branch was of animal kingdom, *Homo sapiens.* Sequence alignment performed using the Boxshade software (University of California San Diego**)** revealed that there were several regions of high homology shared among MIPS proteins. The GmMIPS1 polypeptide contains four highly conserved motifs found in other plant species: GWGGNNG (domain 1), VLWTANTER (domain 2), NGSPQNTFVPGL (domain 3) and SYNHLGNNDG (domain 4) (red box in Fig 4). Moreover, the program PSORT (Prediction of Protein Sorting Signals and Localization Sites in Amino Acid Sequences) predicted a conserved transmembrane motif (CEDSLLAAPIILDLVLLAELSTR), located approximately 68 amino acids from the C-terminus of MIPS protein in both the monocots as well as the dicots (shown by arrow in Fig 4). This provided evidence for a high level of sequence conservation throughout the plant kingdom.

Bacterial overexpression of the GmMIPS1

To attain overexpression of *MIPS* in *E.coli*, 1425 bp (131-1556 bp) sequence with *Bam*H1 and *Hind*III sites was amplified by high fidelity PCR from *MIPS*cDNA template (Fig 5a) and cloned into pGEMT-Easy vector. Recombinant pGEMT-Easy clone was digested with *Bam*H1 and *Hind*III and the released fragment (Fig 5b) was sub-cloned into the prokaryotic expression vector pET28a (+) with the*MIPS1* coding sequence (131-1556 bp) fused in frame to sequences encoding a 6X histidine (HIS) tag of pET28a (+) vector. The recombinant pET28a (+) clones were confirmed through insert release by restriction digestion with *Bam*H1 and *Hind*III (Fig 6a) and the nucleotide sequence of the cloned fragment confirmed by sequencing. The recombinant pET-MIPS plasmid was transformed into *E.coli* BL21 (DE3) and confirmed by colony PCR (Fig 6b.). The positive transformants (BL21/pET-MIPS) were induced by 1 mM IPTG at 37°C for 3 h. and the induced protein band of 52 KDa was resolved on 12% SDS-PAGE (Fig 7a). Western blot analysis using anti-His antibody confirmed the induced

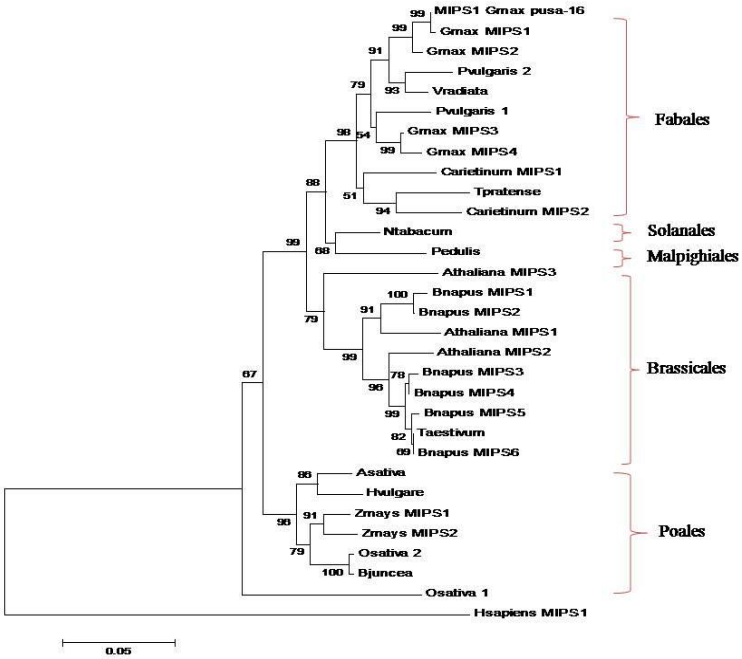


Fig 3. Phylogenetic tree showing relationship amongst the eukaryotic MIPS proteins. The tree was constructed by MEGA 5 using the neighbor-joining (N-J) method with 1000 bootstrap replicates. The bar indicates an evolutionary distance of 0.05%. The GenBank database accession numbers of the MIPS sequences used are as follows: GmPusa-16 (MIPS1GmaxPusa-16:ADJ38521), *Glycine max* (Gmax MIPS1: ABC55420; Gmax MIPS2: DQ323906; Gmax MIPS3: ABC55421; Gmax MIPS4: ABC55422), *Phaseolus vulgaris* (P vulgaris1: Q41107; P vulgaris 2: CAH68559), *Vigna radiata* (Vradiata: ABW99093), *Cicer arietinum* (Carietinum MIPS1: ACH87552; Carietinum MIPS2: ACH87553), *Trifolium pretense* (Tpratense: BAE71281), *Nicotiana tabacum* (Ntabacum: Q9LW96), *Passiflora edulis* (Pedulis: ABF51620), *Arabidopsis thaliana* (Athaliana MIPS1: NP_195690; Athaliana MIPS2: NP_179812; Athaliana MIPS3: NP_196579), *Brassica napus* (Bnapus MIPS1: Q96348; Bnapus MIPS2 : ACJ65004; Bnapus MIPS 3: ACJ65005; Bnapus MIPS 4: ACJ65006; Bnapus 5: ACJ65007; Bnapus6: ACJ65008), *Triticum aestivum* (Taestivum: AAD26330), *Avena sativa* (A sativa: BAB40956), *Hordeum vulgare* (Hvulgare: O65195), *Zea mays* (Zmays MIPS1: AAG40328; Zmays MIPS2: Q9FPK7), *Oryza sativa* (Osativa 1: NP_001049242; Osativa 2: NP_001064454), *Brassica juncea* (Bjuncea: ABY74556), *Homo sapiens* (Hsapiens MIPS1: NP_057452).

BL21/pET-MIPS expression as a His-tagged fusion protein (Fig 7b). The uninducedtransformants harboring the empty pET-28a(+) vector showed no induction of the fusion protein. This study is a foundation for production of low phytate varieties, resulting in improved nutrient availability for animal feed and reduced environmental impact of livestock production.The results thus confirmed that the enzyme GmMIPS1 catalyzing the first step of phytate biosynthesis in developing cotyledons can be targeted successfully by gene silencing strategies for developing low phytate transgenic soybean.

Discussion

The development of low phytate crops is considered an important goal in plant transgenic programs aimed at improving nutritional quality as well as at developing environment friendly and sustainable production (Raboy, 2007). There are several strategies on which scientists are working to develop low phytate soybeans with increased levels of available phosphorus. An alternative to phytase engineering for decreased seed phytic acid levels is the reduction of the expression of enzymes in the biosynthetic pathway of phytic acid. Down regulation of the *MIPS* gene encoding the first and rate-limiting enzyme in the biosynthetic pathway offers a potential approach for the development of low-phytic acid soybean (Hitz et al., 2002). Due to high degree of similarity between the *MIPS* isoforms, a constitutive expression of an antisense construct to block *MIPS* expression could affect other cellular processes dependent on *myo*-inositol such as cell wall biosynthesis, hormonal homeostasis, and signal transmission and could down-regulate *MIPS* activity throughout the plant if it blocks the expression of the multiple isoforms. To decrease the *MIPS* activity in developing seeds to obtain low phytate seeds without a significant impact on other aspects of inositol metabolism, it is necessary to limit the regulation of expression to a seed specific *MIPS* gene isoform. Soybean has four *MIPS* gene family isoforms with differential expression patterns (Chappell et al., 2006). Expression studies on the four isoforms have confirmed the role of *GmMIPS1* isoform in generating D-*myo*-inositol-3-phosphate as a substrate for phytic acid biosynthesis (Hegeman et al., 2001; Chappell et al., 2006 and Kumar et al., 2012). In the present study, a full length *GmMIPS1* gene was isolated from cotyledonary stage of seed development as it was revealed through quantitative expression analysis, to have the maximum expression in developing seeds of size 4-6 mm. Compared to developing cotyledons, the expression levels of *GmMIPS1* were much lower in other vegetative tissues like leaves, flowers, roots, and stems where mainly *MIPS2*, *MIPS3* and *MIPS4* showed significantly higher transcript levels (Chappell et al., 2006). *MIPS*cDNA sequence was

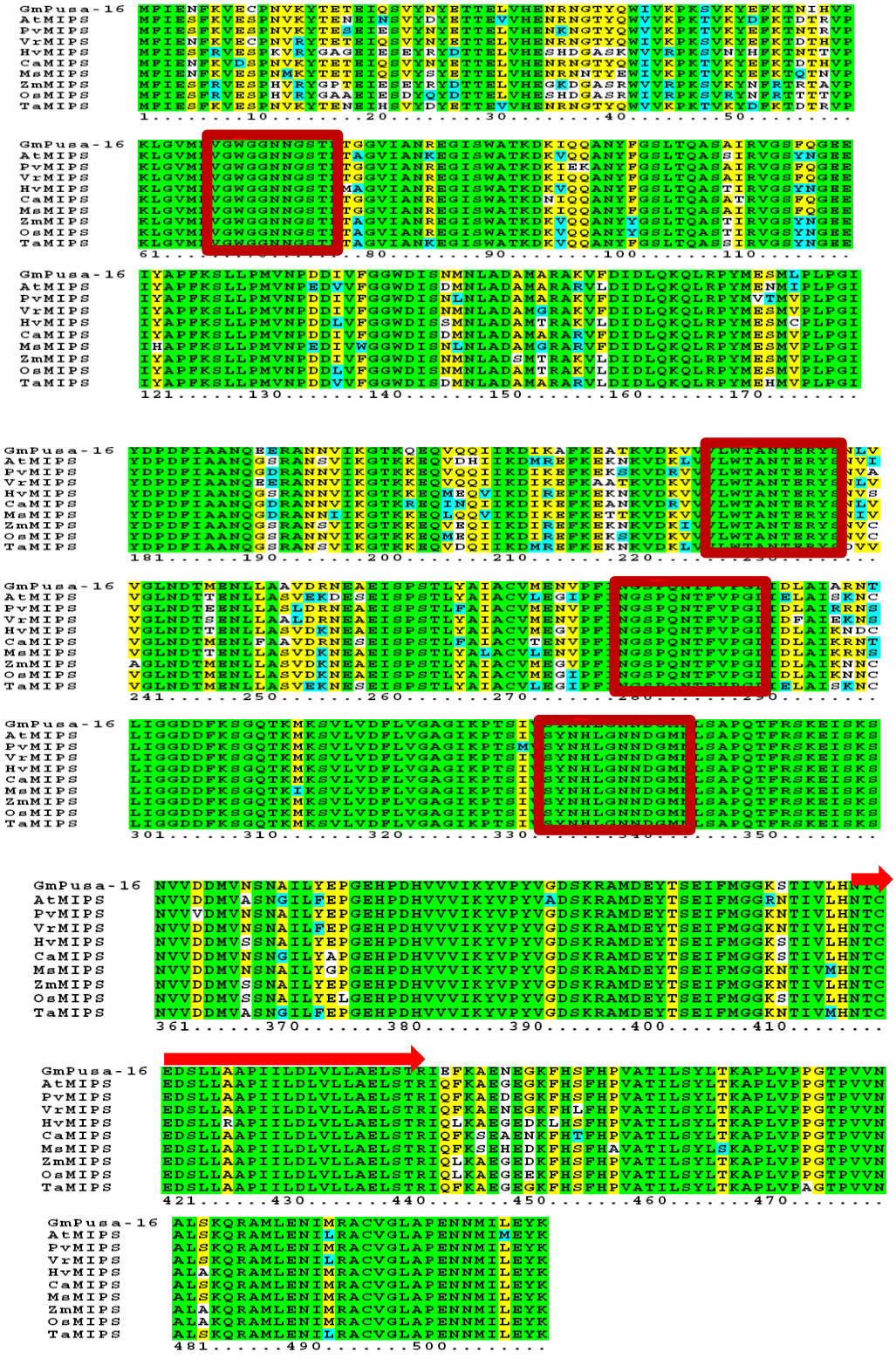


Fig 4. Sequence alignment of the deduced MIPS proteins from plant species. GenBank database accession numbers of the displayed MIPS proteins are as follows: *G. max* (GmPusa-16:ADJ38521), *A.thaliana* (AtMIPS:AAD23618), *P.vulgaris* (PvMIPS: ACN12926), *V. radiata* (VrMIPS:ABW99093), *H.vulgare* (HvMIPS:O65195), *C. arietinum* (CaMIPS: ACH87552), *M. sativa* (MsMIPS:ABO77439), *Z. mays* (ZmMIPS:ACG33827), *O. sativa* (OsMIPS:NP_001049242), *T. aestivum* (TaMIPS:AAD26330). Red box indicates four highly conserved putative domains. A conserved transmembrane motif located at C-terminus of MIPS proteinin both monocots and dicots is marked with an red arrow.

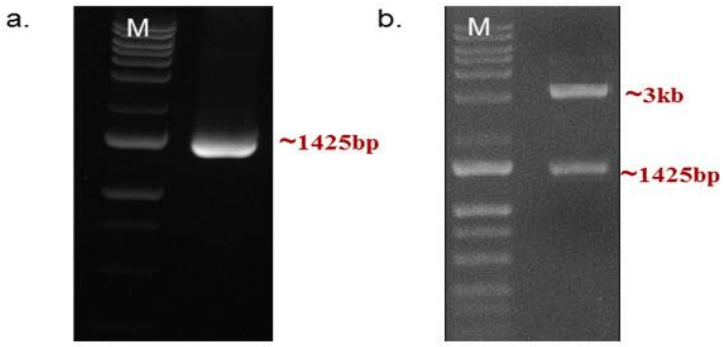


Fig 5. Amplification and cloning of MIPS coding sequence from ORF, a) PCR amplified MIPS coding sequence (1425bp) from ORF; B.*Bam*H1 and *Hind*III digested product (1425bp) from recombinant pGEM-T Easy (3kb) clone containing the MIPS coding sequence on a 1% agarose gel against 1Kb DNA marker (M).

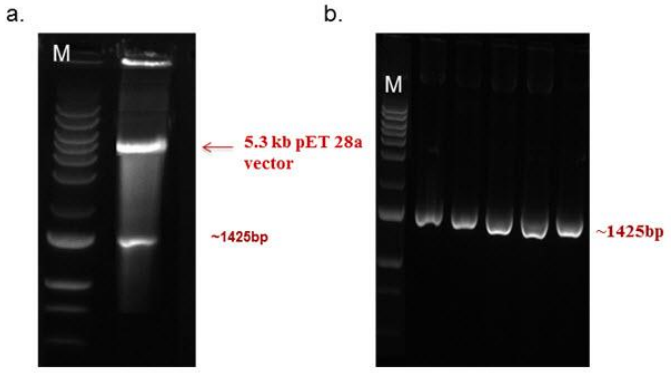


Fig 6. Cloning of the MIPS coding sequence in a prokaryotic expression vector, a) Electrophoresis of insert released using *Bam*H1 and *Hind*III from recombinant pET28a(+) clone; b) Electrophoresis of the *E.coli* BL21(DE3) colonies transformed with recombinant pET-MIPS in 1% agarose gel as against a 1Kb DNA marker (M).

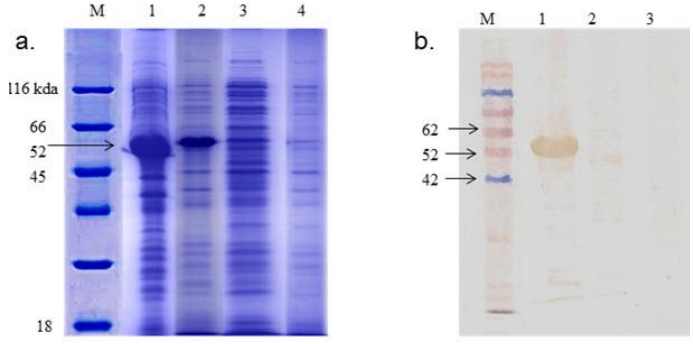


Fig 7. Bacterial expression of GmMIPS1 recombinant protein. a) 12% SDS-PAGE analysis of GmMIPS1, overexpressed protein (~30 µg of protein/lane). Lane 1-induced cell pellet, Lane 2- induced cell supernatant, Lane 3-control cell pellet (recombinant plasmid pET-28a-MIPS without IPTG induction), Lane 4-non-recombinant induced pellet. Molecular weight of the SDS-PAGE markers are shown at the extreme left lane; b) Western Blot analysis of corresponding SDS-PAGE showing a strong hybridization signal at 52kD.

isolated from developing soybean seeds using reverse transcriptase (RT) PCR. The 1,791-bp cDNA sequence submitted to NCBI (Accession No. HM46196) contained a 1,533-bp open reading frame sequence encoding a protein of 510 amino acids. The putative soybean MIPS protein had a predicted molecular mass of 56.5 kD. The expression was consistent with previous results which indicated high expression of *MIPS1* in developing seeds (Hegeman et al., 2001). BLAST database searches performed with the amplified *MIPS*cDNA sequence and the predicted protein sequence (Altschul et al., 1997) revealed a high degree of identity between *GmMIPS1* and *MIPS* genes from other plants. The highest scoring match was with *Vigna radiata* and *Phaseolus vulgare* MIPS sequence, which showed 93% homology to GmMIPS1 at the amino acid level. With yeast, a lower eukaryote, a homology of 40.2% at the nucleotide level and of 41.5% at the amino acid level (Johnson and Henry, 1989), further confirmed a reasonable degree of conservation amongst *MIPS* sequences from different sources. Multiple alignment of deduced amino acid sequences of the *GmMIPS* cDNA and other plant species showed high degree of conservation amongst them. The four highly conserved domains of amino acid residues present in the four *MIPS* isoforms of soybean, were also reported in other monocots and dicots (Majumder et al., 2003; Chun et al., 2003; Kaur et al., 2008 and Abid et al., 2012). The four domains having a role in MIPS protein binding are also essential for MIPS functions, such as cofactor NAD+ binding and reaction catalysis (Majumder et al., 1997 and Norman et al., 2002). It was also predicted that a conserved transmembrane motif (CEDSLLAAPIILDLVLLAELSTR) located about 68 amino acids from the C-terminus observed in *GmMIPS* is also reported in other plant species. The sequence analysis of the *GmMIPS*indicated the abcence of signal for chloroplast transit or mitochondrial targeting peptides in the N-terminal region, thereby suggesting that *GmMIPS1* encodes for a non-secretory cytosolic protein rather than an organelle-specific MIPS protein. However, the absence of a convincing transit peptide in GmMIPS1 does not preclude its targeting to a plastidic site for inositol synthesis as a MIPS from *M. crystallinum* that does not contain a transit peptide for chloroplast import presented enhanced activity only in chloroplasts from light-grown, salt-tolerant rice plants (RayChaudhuri and Majumder, 1996). The analysis was also supported by Lackey et al. (2003) and Abid et al. (2012) who found that the *P. vulgaris* MIPS enzyme was present in plasma membranes, plastids, mitochondria, endoplasmic reticulum, nuclei, and cell walls although it did not have a recognizable transit peptide. A phylogenetic tree constructed to include few representative MIPS sequences from diverse organisms presented an overall evolutionary divergence of this enzyme in the biological kingdom. The higher plants constituted one close subgroup, while the higher animals formed the other sub-groups in the eukaryotic cluster. The amino acid sequence of GmMIPS showed a high degree of homology and clustered strongly in a subgroup belonging to species Fabales. The data indicated that the MIPS protein sequences show a remarkable evolutionary conservation of the primary structure (Majumder et al., 2003). *Glycine max* MIPS protein showed high degree of sequence identity with MIPS of different plant sources from both monocots as well as dicots. It shared 96% identity with *Vigna radiata*, 95% with *Phaseolu svulgare*, 94% with *Cicer arietinum*, 93% with *Nicotiana tabacum*, 92% with *Sesamum indicum*, 89% with *Triticum aestivum*, 88% with *monocots Zea mays, Oryza sativa* and 87% with *Arabidopsis thaliana* which was also confirmed by phylogenetic analysis. On the other hand it

shared least similarity (6%) with *Saccharomyces cerevisiae* and *Homo sapiens*. For expression in bacterial host BL21(DE3), 1425 bp (131bp to 1556bp of ORF) was amplified with restriction site *Bam*H1 and *Hind*III. The amino acid sequences encoded by this fragment were higly specific to GmMIPS1 as compared to other isoforms GmMIPS 2/3/4 as shown by protein sequence alignment. To validate the expression of amplified *GmMIPS1* gene in a prokaryotic system, a MIPS fusion protein construct was generated by insertion of the MIPS amplicon into the expression vector pET28a(+) containing a His tag. Induction of the transformed recombinant BL21/pET-MIPS produced a 52 kDa fusion protein as revealed by SDS-PAGE. Characterizations of the recombinant MIPS proteins in *Cicer arietinum* was done by bacterial expression and were expressed as expected 56 kDa protein when resolved by SDS–PAGE (Kaur et al., 2008). Western-blot analysis was also performed to further corroborate the MIPS expression as a his-tagged fusion protein using anti-his antibody.

Materials and methods

Bacterial strains and plant materials

Escherichia coli BL21 (DE3) and DH-5α strains were cultured on LB medium at 37^{o}C. Cells containing recombinant plasmids, pGEMT-Easy and pET-28a(+) were supplemented with 100 mg ml^{-1} ampicillin and 50 mg ml^{-1} kanamycin. Soybean seeds (*Glycine max* L. var. Pusa16) were collected from the Division of Genetics, Indian Agricultural Research Institute, New Delhi, India. Mercuric chloride (0.02%, 5 min.) sterilized seeds were sown in pots maintained under controlled environmental conditions at the National Phytotron Facility, I.A.R.I., New Delhi. The developing seeds (4 to 6 mm) were harvested and rapidly frozen in liquid nitrogen at -80^{o}C.

RNA isolation and RT-PCR amplification

Total RNA was isolated from developing cotyledons (4 to 6mm seeds) of soybean samples (100 mg) using the RNeasy Plant Mini Kit (Qiagen) according to manufacturer's instructions. Frozen plant tissues were homogenized using pestle and mortar with liquid nitrogen and one ml of Qiagenlysis buffer added per 100 mg of tissue in 2 ml micocentrifuge tubes. First strand of cDNA was synthesized from RNA by using oligo(dT) primer and reverse transcriptase from RevertAidTM H Minus first strand cDNA synthesis kit (Fermentas, Life Sciences). The full length cDNA for *MIPS1* was amplified using oligonucleotide primers designed by BioEdit software based on the published soybean *MIPS* sequences (GenBank Accession Number AF293970) available in NCBI GenBank (forward primer: 5'-ATAGGATTCTCTTCTTTATTCCT-3´; reverse primer: 5´-TACACAAAATTATAC TACATTCAT-3´). The PCR thermal cycling parameters used were 94°C denaturation for 4 min followed by 35 cycles of 94°C for 30 sec., 57.6°C for 30 sec, 72°C for 1 min 30 sec, with a final extension at 72°C for 10 min.

Cloning and sequencing of full-length MIPS1 gene from Glycine max

The amplified product (~1.7 kb) was eluted and purified from the gel using PCR Gel Elution Kit (Bioscience), and ligated overnight at 4^{o}C to the pGEMT-Easy vector (Promega) using manufacturer's protocol. The ligation mixture was used for DH5α transformation and transformants selected on ampicilin/IPTG/X-gal plates according to Sambrook et al., 1989. Plasmid DNA isolated from the transformants in large quantities was purified and the resulting recombinants analyzed by *Eco*RI digestion followed by sequencing of the insert at the DNA Sequencing Facility of South Campus, Delhi University (GenBank Accession number HM461969).

In silico sequence analysis

The nucleotide sequence of *Glycine max MIPS* cDNA (*GmMIPS1*) was subjected to BLAST analysis. The protein prediction and analysis was performed using the ExPASy server and ClustalW tool. Prediction of signal peptide sequence was conducted using SignalP 3.0 server. Identity matrix between different eukaryotic MIPS proteins was obtained by BioEdit Software. The phylogenetic analysis were carried out using the neighbor-joining (N-J) method with 1000 bootstrap replicates implemented in MEGA (Molecular Evolutionary Genetic Analysis) version 5 software program (Tamura et al., 2011). ClustalW and BOXSHADE software were used for multiple alignment of the MIPS protein in plants.

Bacterial overexpression of the GmMIPS1

To achieve bacterial expression and subsequent protein characterization, the *MIPS* coding
sequence was fused in frame to a 19 amino acid N-terminal region of 6X His-tag in expression vector pET-28a (+). From the ORF of 1,533 bp, a 1,425 bp fragment containing majority of *MIPS1* coding nucleotide sequence residues (131 to 1556 bp) was amplified by high fidelity PCR. The restriction sites *Bam*HI and *Hind*III were included in the upstream primer MIPSf*Bam*HI (5–GCAAGGATCCTACAACTACGAAACCACCGA AC T-3') and downstream primer MIPSR *Hind*III (5'-CCATAAGCTTCATCCAACACAA GCCC TCA TTAT-3') respectively to facilitate directional cloning into the expression vector. To reduce the possibility of undesired mutations during PCR, DNA polymerase with proofreading activity, phusion DNA polymerase (Finnzymes) was used and the PCR reaction mixture was prepared according to the instruction manual. The thermal cycling conditions were 94°C denaturation for 4 min followed by 35 cycles of 94°C for 30 sec., 57.6°C for 30 sec, 72°C for 1 min 30 sec, with a final extension at 72°C for 10 min. The amplified MIPS coding sequence was initially cloned into pGEMT-Easy vector (Promega) using manufacturer's protocol. The insert excised from the recombinants using *Bam*HI and *Hind*III was subcloned into the pET-28a(+) vector digested with the same enzymes. The recombinant plasmid, pET-MIPS, was transformed into *E.coli* host strain BL21 (DE3) (Novagen) using standard protocols as described by Sambrook and Russell (2001). The reading frame of recombinants designated as BL21/pET-MIPS were checked by sequencing. Hundred milliliters of Luria Bertani (LB) medium containing kanamycin (50µg mL^{-1}) was inoculated with 1ml of the overnight culture of the BL21/pET-MIPS and agitated at 220 rev min^{-1} at 37°C. Once the OD_{600} reached 0.6-0.8 (log phase), a non-metabolisable lac-inducer, IPTG was added at a concentration of 1 mmolL^{-1} to induce the culture of recombinant protein. Standardization was done for optimization of the length of IPTG induction and the temperature for maximum expression of the fusion protein. Maximum expression of the protein was found to be induced by 1 mM IPTG at 37°C for 3 h. The cells were harvested by

centrifugation at 5000 x g at 4°C and were stored at -80°C. The cells were lysed by sonication with 2x SDS buffer, the lysate was centrifuged at 12,000 x g and the resulting insoluble pellet was resuspended in 2x SDS buffer. The soluble and insoluble fractions of the recombinant protein were resolved by electrophoresis on a SDS- polyacrylamide gel (Laemmli, 1970) in Tris-Glycine buffer (pH 8.0) with a 5% stacking gel (pH 6.8) and a 12% separating gel (pH 8.8) along with a standard protein marker. Coomassie brilliant blue R-250 (0.1%) in a solution of methanol: water: acetic acid (4:5:1) was used to stain the proteins.

Western blot assay

To confirm the recombinant MIPS expression as a His-tagged fusion protein, western blot analysis was carried out using mouse monoclonal anti His-Tagged HRP (horseradish peroxidase as secondary antibody) conjugated antibody (Santa Cruz Biotechnology). After resolving on SDS-PAGE, the protein was transferred onto a PVDF membrane by electroblotting (Invitrogen electroblotter) for 30 min as described in the manufacturer's protocol. The PVDF membrane was incubated in blocking buffer (3% BSA in 1x TBST buffer) for 1 h at 37°C and washed three times for 10 min in 0.05% TBST buffer (pH 7.4). Subsequently, the membrane was incubated with the anti His-Tag HRP conjugated antibody at a dilution of 1:2000 in 1x TBST buffer for 1 h at 37°C, further the blot was washed three times for 10 min each with 0.05% TBST buffer (pH 7.4). The blot was finally developed using the chromogenic substrate DAB (3,3´-diaminobenzidine tetrahydrochloride) and hydrogen peroxide.

Conclusion

Since understanding phytic acid biosynthesis is the most important step in the long-term goal of targeted pathway modification for development of low phytate soybean, the present study resulted in the successful isolation, characterization and expression of the seed specific *MIPS1* gene isoform to be further used in gene silencing strategies for genetic manipulation.

Acknowledgements
Financial support provided by Indian Agricultural Research Institute, New Delhi, in the form of SRF is duly acknowledged.

References

Abid G, Sassi K, Muhovski Y, Jacquemin JM, Mingeot D, Tarchoun N, Baudoin JP (2012) Comparative expression and cellular localization of *myo*-inositol phosphate synthase (*MIPS*) in the wild type and in an EMS mutant during common bean (*Phaseolus vulgaris* L.) seed development. Plant Mol Bio Rep. 30:780–793

Abreu EFM, Aragao FJL (2007) Isolation and characterization of a *myo*-inositol phosphate synthase gene from yellow passion fruit (*Passifloraedulis*f.*flavicarpa*) expressed during seed development and environmental stress. Ann Bot. 99(2):285–292

Abu-abied M, Holland D (1994) Two newly isolated genes from citrus exhibit a different pattern of diurnal expression and light response. Plant Mol Biol. 26:165–173

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nuc Acids Res. 25:3389–3402

Biswas BB, Ghosh B, Majumder AL (1984) *Myo*–inositol polyphosphates and their role in cellular metabolism: a proposed cycle involving glucose 6-phosphate and *myo*-inositol phosphates. In: Roodyn DB (ed.) Subcellular biochemistry. Plenum Press, New York 10:237–280

Bouis HE (2000).Special issue on improving human nutrition through agriculture. Food Nutr Bull. 21: 351-576

Chappell AS, Scaboo AM, Wu X, Pantalone VR, Nguyen H and Bilyeu KD (2006) Characterization of the MIPS gene family in *Glycine max*. Plant Breed. 125: 493- 500

Chiera JM, Grabau EA (2007) Localization of *myo*-inositol phosphate synthase (GmMIPS-1) during the early stages of soybean seed development. J Exp Bot. 58(8):2261-2268

Chun JA, Jin UH, Lee JW, Yi YB, Hyung NI, Kang MH, Pyee JH, Suh MC, Kang CW, Seo HY, Lee SW, Chung CH (2003) Isolation and characterization of a *myo*-inositol 1-phosphate synthase cDNA from developing sesame (*Sesamum indicum* L.) seeds: Functional and differential expression, and salt induced transcription during germination. Planta. 216:874–880

Cosgrove DJ (1966) The Chemistry and Biochemistry of inositol polyphosphate. Rev Pure Appl Chem. 16:209-216

Donahue TF, Henry SA (1981) *Myo*-Inositol-1 phosphate synthase. J Biol Chem. 256:7077–7085

Hegeman CE, Good LL, Grabau EA (2001) Expression of D-*myo*-inositol-3-phosphate synthase in soybean.Implications for phytic acid biosynthesis. Plant Physiol. 125:1941-1948

Hitz WD, Carlson TJ, Kerr PS, Sebastian SA (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiol. 128:650–660

Ishitani M, Majumder AL, Bornhouser A, Michalowski CB, Jensen RG, Bohnert HJ (1996) Coordinate transcriptional induction of *myo*-inositol metabolism during environmental stress. Plant J. 9:537-548

Johnson MD (1994) The *Arabidopsis thalianamyo*-inositol 1-Phosphate Synthase (EC 5.5.1.4). Plant Physiol. 105:1023-1024

Johnson MD, Henry SA (1989) Biosynthesis of inositol in yeast: primary structure of *myo*-inositol-1-phosphate synthase (EC 5.5.1.4) and functional analysis of its structural gene, the INO1 locus. J Biol Chem. 264:1274–1283

Johnson MD, Sussex IM (1995) 1-L-*myo*-inositol 1-phosphate synthase from *Arabidopsis thaliana*. Plant Physiol. 107:613-619

Kumar A, Kumari S, Jolly M, Rai RD, Sachdev A (2012) Cloning and characterization of *myo*-inositol-1-phosphate synthase (MIPS1) gene in developing seeds of *Glycine max*. Indian J Plant Physiol. 17(2):121-127

Kaur H, Shukla R, Yadav G, Chattopadhyay D, Majee M (2008) Two divergent genes encoding L-*myo*-inositol 1-phosphate synthase1 (CaMIPS1) and 2 (CaMIPS2) are differentially expressed in chickpea. Plant Cell and Env. 31:1701–1716

Lackey KH, Pope PM, Johnson MD (2003) Expression of 1L-*myo*-inositol-1-phosphate synthase in organelles. Plant Physiol. 132:2240–2247

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227(5259):680-685

Loewus FA, Murthy PPN (2000) *Myo*-Inositol metabolism in plants. Plant Sci. 150:1–19

Lott JNA (1984) Accumulation of seed reserves of phosphorus and other minerals. In: Murray, D.R. (Ed). Seed Physiol Acad Press, New York pp 26-166

Majee M, Maitra S, Dastidar KG, Pattnaik S, Chatterjee A, Hait NC, Das KP, Majumder AL (2004) A novel salt-tolerant L-*myo*-inositol-1-phosphate synthase from *Porteresi acoarctata* (Roxb.) Tateoka, a halophytic wild rice. J Biol Chem. 279(27):28539–28552

Majee M, Patra B, Mundree SG, Majumder AL (2005) Molecular cloning, bacterial overexpression and characterization of L-*myo*-inositol 1-phosphate synthase from a monocotyledonous resurrection plant, *Xerophytaviscosa* Baker. J Plant Biochem Biotech. 14(2):95-99

Majumder AL, Duttagupta S, Goldwasser P, Donahue TF, Henry SA (1981) The mechanism of interallelic complementation at the INO-1 locus in yeast: immunological analysis of mutants. Mol General Gen. 184:347–354

Majumder AL, Johnson MD, Henry SA (1997) 1L-myo-Inositol-1-phosphate synthase. Biochem Biophy Acta. 1348:245–256

Majumder AL, Chatterjee A, Dastidar KG, Majee M (2003) Diversification and evolution of L-*myo*-inositol 1 phosphate synthase. FEBS Letters. 533: 3–10

Norman RA, McAlister MSB, Murray Rust J, Movahedzadeh F, Stoker NG, McDonald NQ (2002) Crystal structure of inositol 1-phosphate synthase from *Mycobacterium tuberculosis*, a key enzyme in phosphatidyl inositol synthesis. Structure. 10:393–402

Prattley CA, Stanley DW (1982) Protein-phytate interactions in soybeans: I. Localization of phytate in protein bodies and globoids. J Food Biochem. 6: 243-253

Raboy V, Dickinson DB (1987) The timing and rate of phytic acid accumulation in developing soybean seeds. Plant Physiol. 85:841–844

Raboy V, Dickinson DB, Below FE (1984) Variation in seed total P, phytic acid, zinc, calcium, magnesium, and protein among lines of *Glycine max* and *G. sojae*. Crop Sci. 24: 431-434

Raboy V (2001) Seeds for a better future: 'low phytate' grains help to overcome malnutrition and reduce pollution. Trends in Plant Sc. 6:458-462

Raboy V (2007) Seed phosphorus and the development of low-phytate crops. In: Turner, B.L., Richardson, A.E., Mullaney, E.J. (Eds.), Inositol Phosphates. Linking Agriculture and the Environment. CAB International, Wallingford, Oxfordshire, UK. pp. 111-132

RayChaudhuri A, Majumder AL (1996) Salinity-induced enhancement of L-*myo*-inositol 1-phosphate synthase in rice (*Oryza sativa* L.). Plant Cell Env. 19: 1437–1442

Sambrook J, Russell DW (2001) Molecular Cloning: a Laboratory Manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA

Sebastian SA, Kerr PS, Pearlstein RW, Hitz WD (2000) Soybean germplasm with novel genes for improved digestibility.*In* JK Drackley, ed, Soy in Animal Nutrition. Federation of Animal Sci Societies, Savoy, IL 56–74

Smart CC, Fleming AJ (1993) A plant gene with homology to D-*myo*-inositol-3-phosphate synthase is rapidly and spatially upregulated during an abscisic acid induced morphogmic response in Spirodelapolyrrhiza. Plant J. 4: 279-293

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 28(10):2731–2739