

# ANÁLISE *IN SILICO* DO METABOLISMO DO FITATO E DA FERRITINA EM *Coffea*

Vera QUECINI<sup>1</sup>, E-mail: vquecini@iac.sp.gov.br; Paula M. NOBILE<sup>1</sup>; Carlos COLOMBO<sup>1</sup>; Paulo MAZZAFERA<sup>2</sup>

<sup>1</sup>Instituto Agronômico de Campinas- Genética, Campinas, SP; <sup>2</sup>UNICAMP – Instituto de Biociências, Campinas, SP

## Resumo:

O presente trabalho teve por objetivo empregar ferramentas de bioinformática em análises *in silico* para investigar o metabolismo do fator antinutricional fitato e da família de proteínas associadas a este composto por ação quelante, a ferritina. A maior parte dos componentes caracterizados do metabolismo fitato-ferritina foi encontrada nos bancos de dados do Genoma do Café. Transcritos semelhantes às duas principais famílias do metabolismo do fitato foram encontrados em *Coffea*. Interessantemente, conforme descrito em *Arabidopsis*, estas proteínas de café também parecem ter uma divergência funcional, estando algumas associadas ao acúmulo de reservas energéticas nas sementes e as demais, induzidas em condições de estresse biótico e abiótico. Em café, a família das ferritinas é divergente da família de *Arabidopsis*, sendo os componentes mais semelhantes entre si do que aos ortólogos da planta-modelo. As descobertas do presente trabalho demonstram o poder e a transferibilidade de informações funcionais de sistemas-modelo para culturas de importância econômica.

Palavras-chave: anti-nutricional, ferritina, ferro, fitato, *mio*-inositol.

## *IN SILICO* ANALYSIS OF PHYTATE AND FERRITIN METABOLISM IN *Coffea*

### Abstract:

The main goal of the present research work was to introduce bioinformatics and *in silico* analysis tools to investigate metabolism of an anti-nutritional factor, phytate, and its associated iron-rich protein family, ferritin. High phytate contents trigger its chelating effects on ferritin, thus making the iron unavailable for animal nutrition. For the most part, the components of phytate-ferritin metabolism were found in *Coffea* Genome databases. Interestingly, as described in *Arabidopsis*, coffee proteins also appear to be functionally divergent, with the main branching point being energy storage and stress responses. *Coffea* ferritin family shares restricted sequence similarity to its *Arabidopsis* counterparts. They present higher degrees of similarity among each other to the model plant corresponding sequences. Taken together, our analyses have demonstrated the power and the high transferability of model-system functional information to economically significant crops.

Key words: anti-nutritional, ferritin, iron, *myo*-inositol, phytate.

### Introdução

O acúmulo de amido, lipídios, proteínas e minerais essenciais durante o desenvolvimento das sementes é responsável pelo alto valor nutricional das sementes vegetais. Durante o desenvolvimento, outros compostos também são depositados na semente, incluindo o derivado fosforilado da glicose, *mio*-inositol-1,2,3,4,5,6-hexaquisfosfato (Ins P6) ou fitato, que é responsável por aproximadamente dois terços do fósforo total da semente (Raboy, 2003). Recentemente, foi demonstrado que o fitato participa de outros processos e tem outras funções biológicas, além de seu papel na armazenagem de nutrientes (Raboy, 2003). Ins P6 e seus derivados estão envolvidos na exportação de RNA, reparo e recombinação de DNA (Hanakahi and West, 2002; Luo et al., 2002; York et al., 1999), na endocitose e no transporte vesicular (Saiardi et al., 2002), e como antioxidante (Graf et al., 1987). O mutante *fieryl* de *Arabidopsis thaliana* é defeitivo na sequência codificadora da Ins polifosfato 1-fosfatase e apresenta um significativo aumento nas respostas reguladas por ABA ou induzidas por estresse (Xiong et al., 2001), associando Ins P6 a respostas desencadeadas por estresses. Além disso, quantidades sub-micromolares de Ins P6 (16) desencadeiam alterações rápidas em células-guarda de estômatos de *Solanum tuberosum* através de fluxo iônico (Lemtiri-Chlieh et al., 2000). Nas sementes maduras Ins P6 (16) é frequentemente encontrado na forma de depósitos de sais mistos de “fitato” e cátions minerais, tais como K, Mg, Ca, Mn e Zn (Lott, 1984). As altas concentrações de fitato nas sementes têm efeitos quelantes sobre íons metálicos, tais como ferro e zinco. Portanto, há uma relação próxima entre o metabolismo do ferro e do *mio*-inositol. Os metais de transição, tais como o ferro, são micronutrientes essenciais para a nutrição animal. Sua retenção em quelatos com fitato reduz severamente a disponibilidade para os animais e pode levar à desnutrição. O principal objetivo do presente trabalho foi investigar o metabolismo do ácido fítico e da ferritina em *Coffea*, empregando ferramentas de análise *in silico*, visando fornecer ferramentas para que os programas de melhoramento possam abordar caracteres complexos responsáveis pelas propriedades nutraceuticas da bebida de café.

### Material e Métodos

Foram realizadas buscas por BLAST (Altschul et al., 1997) de homólogos dos componentes do metabolismo do ácido fítico e da ferritina caracterizados em *Arabidopsis thaliana* nas bases de dados do Projeto Genoma Café, constituídas de aproximadamente 160,000 ESTs obtidos do sequenciamento de 35 bibliotecas. A validação dos dados foi feita por

buscas tBLASTx e tBLASTn das sequências recuperadas contra a base de dados GenBank database inserida no servidor LGE-UNICAMP. As sequências que não retornaram a sequência-isca original foram eliminadas dos projetos. Os alinhamentos resultantes foram filtrados por um valor E de  $1e-15$  e os homólogos (*hits*) foram analisados mais detalhadamente de acordo com os domínios funcionais caracterizados. As sequências validadas foram traduzidas e alinhamentos de proteínas (sequência deduzida de amino ácidos) foram construídos empregando-se ClustalX (Thompson et al., 1997). Quando necessário, os alinhamentos foram ajustados manualmente usando o *software* Lasergene MegAlign (DNASTAR, Madison, WI, USA).

Os homólogos identificados foram investigados detalhadamente para a presença e conservação de sequência nos domínios funcionais descritos em várias bases de dados de análise de proteínas e de função gênica (EMBL EBI [www.ebi.ac.uk/interpro/](http://www.ebi.ac.uk/interpro/); ExPaSy do SIB <http://www.expasy.org/prosite/> e <http://www.us.expasy.org/sprot/>; GO, <http://www.godatabase.org/cgi-bin/amigo/go.cgi>; Pfam <http://www.sanger.ac.uk/Software/Pfam/>).

A funcionalidade dos genes identificados em *Coffea* em comparação aos ortólogos de *Arabidopsis* foi avaliada por estudos de distância genética e de filogenia. As análises filogenéticas foram realizadas empregando-se métodos de distância e de parcimônia no *software* PAUP\* 4.0b10 (<http://paup.csit.fsu.edu/>), usando os parâmetros *default* do *software*. Árvores de re-amostragem por *bootstrap* contendo 1000 amostras aleatórias foram construídas usando o programa PSIGNFIT (<http://www.bootstrap-software.org/>). Os domínios funcionais modulares foram usados para os estudos de distâncias genéticas para os genes anteriormente caracterizados no caso de proteínas constituídas de regiões divergentes e blocos conservados.

Os perfis qualitativos de expressão gênica foram realizados por análise *in silico* empregando as bases de dados de *Coffea* mencionadas anteriormente, por análise de *northern blot* virtual. O gene de interesse foi utilizado em buscas contra as bases de dados de referência, gerando um alinhamento do gene-isca e de seus parálogos. O alinhamento resultante foi então empregado na busca de sequências de mRNA específicas (sondas). Os alinhamentos resultantes foram usados coletivamente para investigar novamente a base de dados de EST usando BLAST. Esta abordagem heurística foi empregada visando eliminar falsos positivos, ou ESTs de genes parálogos à isca mas não correspondentes ao próprio gene-isca. Os números dos ESTs correspondentes às sondas foram identificados e as bases de dados do projeto Genoma Café foram empregadas para identificar as bibliotecas em que os ESTs foram encontrados. A frequência de *reads* para cada EST *contig* em uma determinada biblioteca foi calculada e normalizada para o número total de *reads* da biblioteca e o número total de *reads* em todas as bibliotecas. Foi gerada uma matriz de correlação entre os EST *contigs* e as bibliotecas e os padrões de expressão dos ESTs nas bibliotecas foram obtidos por clusterização hierárquica de acordo com a matriz de correlação e classificação de Spearman usando o *software* Cluster v.2.11 (Eisen et al., 1998; <http://rana.lbl.gov/EisenSoftware.htm>), substituindo-se os clusters por seu padrão médio de expressão. Os gráficos foram gerados usando-se o programa Tree View v.1.6 (<http://rana.lbl.gov/EisenSoftware.htm>) e apresentados na forma de escala de cinza.

## Resultados e Discussão

A estrutura das vias metabólicas centradas em Ins P6 consiste de duas etapas: as vias iniciais que compreendem a síntese e provisão de Ins P6, e as vias finais de Ins fosfato, que se iniciam a partir de Ins fosfatos solúveis em direção a Ins(3,4,6)P3 (5) ou via intermediários fosfatidil-Ins fosfato em direção à produção de Ins(1,4,5)P3 (7) (Raboy 2003). Nas bases de dados transcricionais de *Coffea*, foram identificados 89 transcritos que apresentam sequências similares aos mRNAs de *A. thaliana* mRNAs que codificam genes envolvidos no metabolismo do fitato e ferritina (Tabela 1).

Tabela 1. Componentes do metabolismo do fitato e ferritina em plantas de *Coffea*.

Nome do gene	Código <i>At</i> da proteína	% Similaridade (similaridade, cobertura)	Domínio Funcional e Processo Biológico	Referência
<i>FIERY1</i>	AT4G21670	11 contigs e singlets (99 a 12.7%, 100 a 27.0%)	nucleotidase bisfosfato, ligação a RNA fita dupla, síntese de fitato	Xiong et al, 2001
Família da Ferritina	AT2G40300	3 contigs (66 a 25.6%, 97 a 75%)	ferritina, homeostase de ferro, transporte de ferro, resposta a ferro iônico	Petit et al, 2001
Família IK	AT3G09920	25 contigs e singlets (100 a 24.5%, / 100% a 33.0%)	inositol quinase	Lou et al, 2007
Família IP/PI	AT1G34120	10 contigs e singlets (100 a 19.0%, 85 a 21.0%)	inositol monofosfatase	Sanchez and Chua, 2001
Família MIOX	AT1G14520	2 contigs (60.2 a 17.7%, / 87 a 23%)	<i>mio</i> -inositol oxidase	Kanter et al., 2005
Família <i>AtPIS1</i> /PGP	AT1G68000	35 contigs e singlets (100% a 24.5, / 100 a 33%)	fosfatidil glicerol fosfato sintase	Qin et al., 1997
Família PCP	AT1G07230	2 contigs (95 a 24.4% / 100 a 25%)	fosfolipase C	na

A única fonte biossintética de esqueleto Ins é a atividade da enzima *mio*-inositol(3)P1 sintase, que converte a glicose-6P a Ins(3)P1 (Loewus & Murthy, 2000). Nos tecidos de plantas de café, foram identificados 25 transcritos que apresentam sequência semelhante à família de proteínas responsável pela biossíntese de In P6 (Tabela 1).

A via inicia-se a partir da adição de P a derivados de glicose por uma família de quinases de ação sequencial: 3-quinase, 6-quinase, 4-quinase, 5-quinase, 1-quinase, e finalmente 2-quinase. No transcriptoma de *Coffea*, foram identificadas seis famílias de fosfatidil inositol quinases. Notavelmente, as formas predominantes de fosfatidil inositol quinases em café foram 3-4- e 4-5-diquinase (Figura 1). O domínio MORN, presente na família de fosfatidil inositol quinases funcionalmente caracterizadas, é abundante em bibliotecas de *Coffea* (Figura 1), sendo preferencialmente expresso ao longo do desenvolvimento dos frutos e em resposta a estresses (Figura 3A), consistente com seu papel no acúmulo de fitato e nas respostas aos estresses (Xiong et al., 2001).

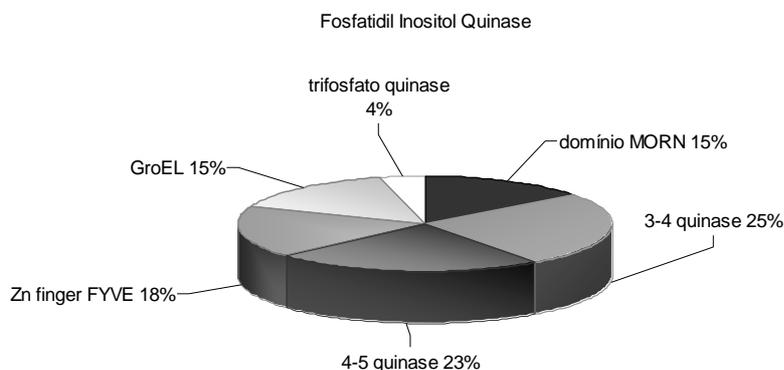


Figura 1 - Categorias funcionais das proteínas fosfatidil inositol quinase de *Coffea* baseado em esquema de ontologia gênica de função molecular. As atribuições foram baseadas nos dados disponíveis no Índice Gênico de *Arabidopsis thaliana* versão 13.0 no TIGR.

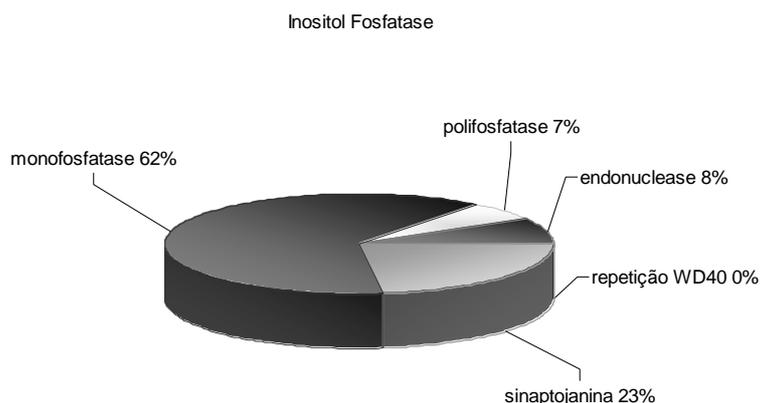


Figura 2 - Categorias funcionais das proteínas fosfatidil inositol fosfatase de *Coffea* baseado em esquema de ontologia gênica de função molecular. As atribuições foram baseadas nos dados disponíveis no Índice Gênico de *Arabidopsis thaliana* versão 13.0 no TIGR.

A via reversa na biossíntese de fitato ocorre pela ação de uma família de proteínas fosfatase, constituída de quatro membros (Figura 2). O grupo mais abundante é o grupo de monofosfatases (Figura 2), cuja expressão também é fortemente induzida nas bibliotecas que constituem o Projeto Genoma Café, principalmente em bibliotecas derivadas de botões florais e culturas de calos embriogênicos (Figura 3B). Os resultados das análises *in silico* realizadas neste trabalho indicam que a família IK/ *AtPIS1* tem funções não-redundantes no metabolismo de fitato em *Coffea*. Nossas observações também sugerem que a divisão funcional ocorra entre as proteínas envolvidas no acúmulo de fitato e as responsáveis pelas respostas ao estresse.

A família das ferritinas de *Coffea* apresenta similaridade mediada com relação à família multigênica de quatro ferritinas de *Arabidopsis*. A similaridade mais elevada foi observada dentro das famílias de cada espécie, sendo que a similaridade entre as espécies é menor (Figura 4).

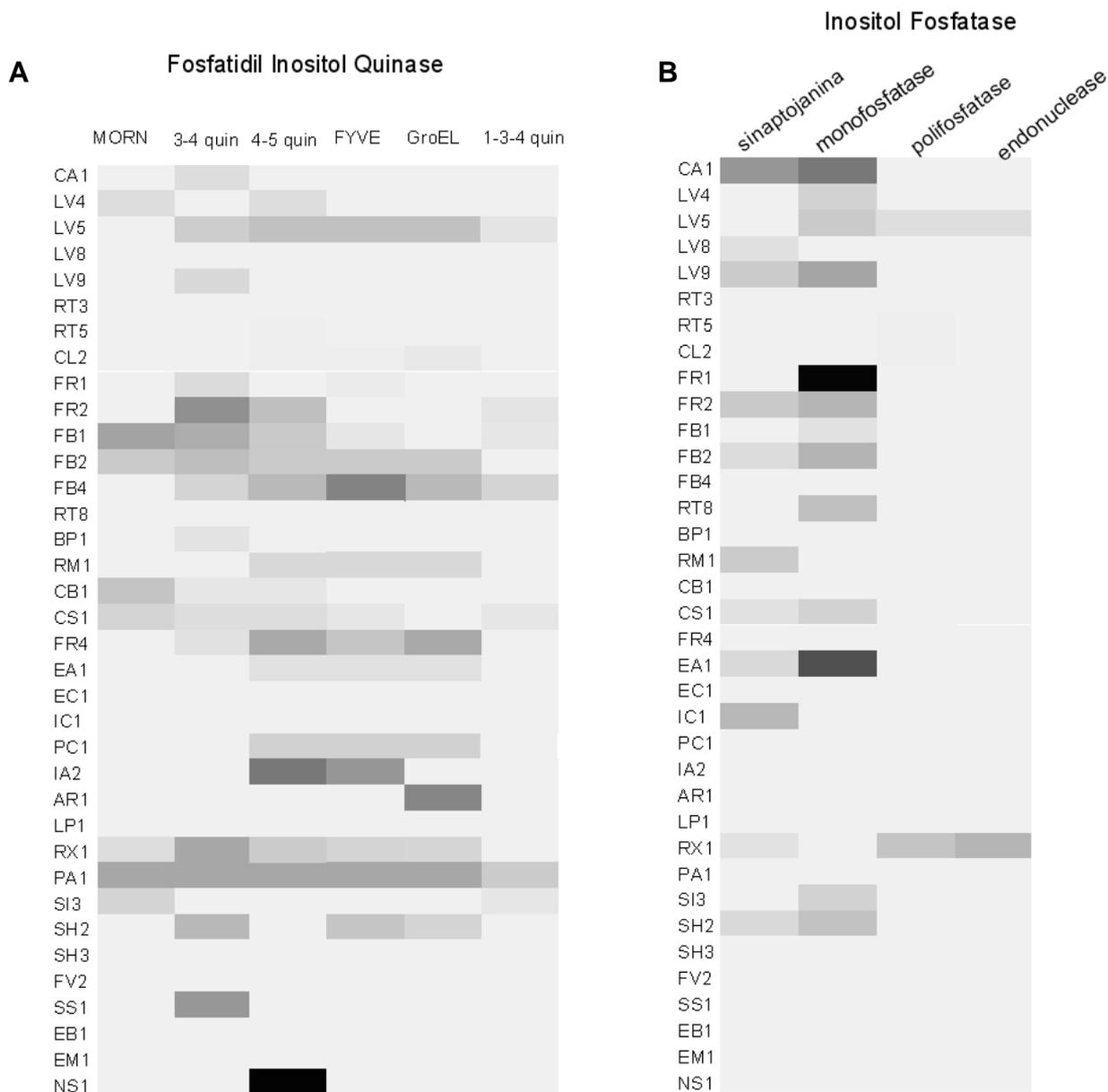


Figura 3 - Perfil de expressão *in silico* dos components do metabolismo do fitato-ferritina em *Coffea*. A: família fosfatidil inositol quinase e B: família inositol fosfatase. O número normalizado de reads correspondents aos transcritos em cada biblioteca está representado pela escala de cinza. As bibliotecas de *Coffea* estão nas linhas e as famílias nas colunas. Os códigos para as bibliotecas são: CA1: calos não-embriogênicos, LV4: folha; LV5: folha; LV8: folha; LV9: folha; RT3: raiz; RT5: raiz com análogo de ácido salicílico (SA); CL2: hipocotilo; FR1: botão floral; FR2: botão floral; FB1: botão floral; FB2: botão floral; FB4: botão floral; RT8: suspensão de células radiculares; BP1: suspensão de células radiculares e foliares; RM1: folha infectada com bicho mineiro e ferrugem; CB1: suspensão celular com análogo de SA e brassinosteróides; CS1: suspensão celular com manose, NaCl e KCl; FR4: frutos de *racemosa*; EA1: calo embriogênico de *Coffea arabica*; EC1: calo embriogênico de *Coffea canephora*; IC: linhagem de *C. arabica* não-embriogênica e não-induzida; PC1: linhagem de *C. arabica* não-embriogênica induzida; IA2: linhagem embriogênica de *C. arabica* induzida com 2,4-D; AR1:folhas tratadas com ácido araquidônico; LP1: plântulas tratadas com ácido araquidônico; RX1: ramos infectados com *Xyllela*; PA1: linhagem embriogênica de *C. arabica*; SI:sementes em germinação; SH2:plantas de campo sob estresse hídrico; SH3: plantas de *C. canephora* tolerantes sob estresse hídrico; FV2: furtos de *C. racemosa*; SS1: plantas no campo sob irrigação; EB1: embrião de fruto imaturo; EM: embrião de semente madura; NS1: raiz infectada com nematóide.

### Conclusões

Este estudo preliminar do metabolismo do fitato e da ferritina em *Coffea* permitiu a identificação de componentes conservados e o estabelecimento de bases para estudos futuros destas vias associados às qualidades nutracêuticas da bebida de café. A transferabilidade de descobertas da espécie-modelo *A. thaliana* para várias espécies de *Coffea* demonstrou extensa conservação de sequência em componentes do metabolismo básico em plantas.

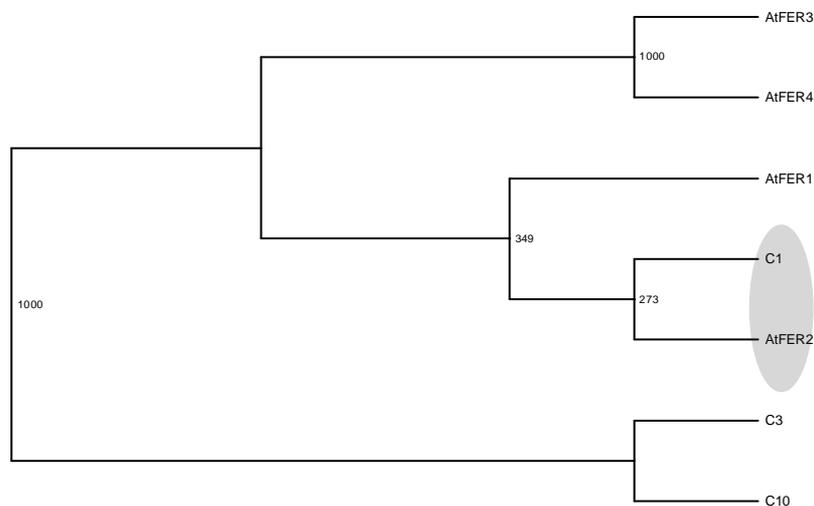


Figura 4- Análise filogenética da família da ferritina de *Coffea* e *A. thaliana*. A árvore foi construída a partir das sequências deduzidas de amino ácidos completas de *Coffea* (C) e das sequências de proteína *Arabidopsis* (At) por Neighbor-joining a partir de alinhamento gerado por ClustalX.

### Referências Bibliográficas

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Research*, 25:3389-3402.
- Eisen MB, Spellman PT, Brown PO and Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences (USA)*, 5:14863-14868.
- Graf E, Epton KL, Eaton JW (1987) Phytic acid: a natural antioxidant. *Journal of Biological Chemistry*, 262:11647-11650.
- Hanakahi LA, West SC (2002) Specific interaction of IP6 with human Ku70/80, the DNA-binding subunit of DNA-PK. *EMBO Journal*, 21:2038-2044.
- Kanter U, Usadel B, Guerineau F, Li Y, Pauly M, Tenhaken R (2005). The inositol oxygenase gene family of *Arabidopsis* is involved in the biosynthesis of nucleotide sugar precursors for cell-wall matrix polysaccharides. *Planta.*, 221:243-254.
- Lemtiri-Chlieh F, MacRobbie EAC, Brearley CA (2000) Inositol hexakisphosphate is a physiological signal regulating the K<sup>+</sup>-inward rectifying conductance in guard cells. *Proceedings of the National Academy of Sciences (USA)* 97:8687-8692.
- Lott JNA, Ockenden I, Raboy V, Batten GD (2000) Phytic acid and phosphorus in crop seeds and fruits: a global estimate. *Seed Sciences Research*, 10:11-33.
- Lou Y, Gou JY, Xue HW (2007) PIP5K9, an *Arabidopsis* phosphatidylinositol monophosphate kinase, interacts with a cytosolic invertase to negatively regulate sugar-mediated root growth. *Plant Cell*, 19:163-181.
- Luo HR, Saiardi A, Yu H, Nagata E, Ye K, Snyder SH (2002) Inositol pyrophosphates are required for DNA hyper recombination in protein kinase C1 mutant yeast. *Biochemistry*, 41:2509-2515.
- Petit JM, van Wuytswinkel O, Briat JF, Lobreaux S (2001) Characterization of an iron-dependent regulatory sequence involved in the transcriptional control of *AtFer1* and *ZmFer1* plant ferritin genes by iron. *Journal of Biological Chemistry*, 276:5584-9550.
- Qin W, Pappan K, Wang X (1997) Molecular heterogeneity of phospholipase D (PLD). Cloning of PLD-gamma and regulation of plant PLD-gamma, -beta, and -alpha by polyphosphoinositides and calcium. *Journal of Biological Chemistry*, 272:28267-28273.
- Raboy V (2003) Myo-Inositol-1,2,3,4,5,6-hexakisphosphate. *Phytochemistry*, 64:1-33-1043.
- Saiardi A, Nagata E, Luo HR, Sawa A, Luo X, Snowman AM, Snyder SH (2001) Mammalian inositol polyphosphate multikinase synthesizes inositol 1,4,5-trisphosphate and an inositol pyrophosphate. *Proceedings of the National Academy of Sciences (USA)* 98:2306-2311.
- Sanchez JP, Chua NH (2001) *Arabidopsis* PLC1 is required for secondary responses to abscisic acid signals. *Plant Cell*, 13:1143-1154.
- Xiong L, Lee Bh, Ishitani M, Lee H, Zhang C, Zhu JK (2001) FIERY1 encoding an inositol polyphosphate 1-phosphatase is a negative regulator of abscisic acid and stress signaling in *Arabidopsis*. *Genes and Development*, 15:1971-1984.
- York JD, Odom AR, Murphy R, Ives EB, Went SR (1999) A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science* 285:96-100.