

ANÁLISE *IN SILICO* DA VIA BIOSINTÉTICA DO TETRAHIDROFOLATO EM *Coffea*

Paula M. NOBILE¹, E-mail: pmnobile@iac.sp.gov.br; Vera QUECINI¹; Carlos A. COLOMBO¹; Paulo MAZZAFERA²

¹Instituto Agrônomo de Campinas, Campinas, SP; ²UNICAMP – Instituto de Biociências, Campinas, SP.

Resumo:

O objetivo deste estudo foi minerar os dados de seqüências expressas do café (ESTs), depositadas na plataforma de bioinformática do Laboratório de Genômica e Expressão (LGE) da Universidade Estadual de Campinas, relacionadas à biossíntese do tetrahydrofolato. Devido às plantas serem a principal fonte de folatos na dieta humana, há um grande esforço para selecionar plantas que tenham uma alta concentração desta família de vitaminas ou para modificar seu metabolismo nas espécies vegetais, visando aumentar sua quantidade inicial. Nós relatamos sete famílias gênicas que foram caracterizadas por seu envolvimento na produção de folato em café. Folhas, cultura de suspensão de células, calos de tecidos embriogênicos (ou não) e botões florais representaram os tecidos e órgãos que mais acumulam transcritos relacionados ao folato. As informações apresentadas neste trabalho serão valiosas em futuros experimentos que considerem a análise da biossíntese do folato em café.

Palavras-chave: bioinformática, genômica, café, tetrahydrofolato

IN SILICO ANALYSIS OF BIOSYNTHETIC PATHWAY OF TETRAHYDROFOLATE IN *Coffea*

Abstract:

The aim of this study was to mine the data from coffee expressed sequence tags (ESTs), at the bioinformatics platform of Laboratório de Genômica e Expressão (LGE) of the State University at Campinas, concerning tetrahydrofolate biosynthesis. Due to the fact that plants are the main source of dietary folates, there is great concern to breed plant feedstock with high concentrations of this vitamin or to engineer folate metabolism to increase its initial amount. We have report seven gene families that are characterized as been involved in coffee folate production. Leaves, cell-suspension cultures, embryogenic (or not) calli tissues and flower buds represented the coffee tissues and organs accumulating higher contents of folate-associated transcripts. We believe that the information presented in this article will be valuable for further approaches to analyze folate biosynthesis in coffee.

Key words: bioinformatic, genomic, coffe, tetrahydrofolate.

Introdução

O Projeto Brasileiro do Genoma Café (PBGC) inaugurou uma nova fase da pesquisa em genômica do café. Diversas linhas de interesses estão sendo investigadas a partir dos dados gerados pelo PBGC. A disponibilidade de 153.739 seqüências expressas (ESTs) de três espécies de *Coffea* analisadas, agrupadas e anotadas (Vieira *et al.*, 2006) por meio de ferramentas de bioinformática desenvolvidas pelo Laboratório de Genômica e Expressão (LGE - <http://www.lge.ibi.unicamp.br/cafe>) permite o uso de diversas estratégias e ferramentas para o estudo funcional dos genes. Em uma primeira fase, a principal estratégia investigativa é a análise *in silico* da expressão dos genes de interesse. Desta forma elimina muitas etapas experimentais e auxilia no delineamento de experimentos futuros.

O interesse em agregar valores nutracêuticos à bebida do café motivou o estudo do metabolismo do folato. Os folatos são vitaminas (B9) envolvidas na intermediação de várias reações de transferência do carbono-um (C1). Suplementam unidades de C1 para vias metabólicas que são de extrema importância para a viabilidade celular (Hanson e Roje, 2001). A deficiência desta vitamina na dieta humana pode causar doenças sérias.

Todas as enzimas envolvidas na biossíntese do tetrahydrofolato (THF) em plantas foram identificadas ao longo dos últimos dez anos (Rébeillé *et al.* 2006). A via metabólica desta síntese apresenta uma organização espacial complexa que envolve três compartimentos sub-celulares. No citosol ocorre a síntese da hidroximetildihidropterina a partir do GTP, enquanto que no plastídio ocorre a síntese do ácido *p*-aminobenzoico (*p*ABA) e finalmente a última etapa da síntese é a formação do tetrahydrofolato (THF), na mitocôndria. As respectivas enzimas participantes de cada etapa de síntese são: aminodeoxychorismate (ADC) sintase e ADC liase; GTP ciclohrolase I; nudix hidrolase; dihidroneopterina aldolase; hidroximetildihidropterina pirofosfoquinase e dihidropteroato sintase (HPPK-DHPS) bifuncional; dihydrofolate sintetase (DHFS) e folilpoliglutamato sintase (FPGS) e por último dihydrofolate reductase (DHFR).

O objetivo deste trabalho foi identificar e caracterizar as seqüências disponíveis no banco de dados do PBGC homólogas as seqüências funcionalmente caracterizadas em organismos modelos na biossíntese do tetrahydrofolato, além de avaliar a expressão dos genes de interesse *in silico* de acordo com a freqüência proporcional (normalização) encontrada em cada uma das 35 bibliotecas de cDNA que representam diferentes tecidos, estágios de desenvolvimento e condições ambientais.

Material e Métodos

Busca em Bancos de dados e alinhamentos

Foram identificadas ESTs (seqüências expressas) homólogas a genes caracterizados funcionalmente envolvidos na biossíntese do tetraidrofolato (THF) por busca no BLAST (Altschul *et al.*, 1997) contra contigs e EST do bando de dados PBGC, constituído aproximadamente de 122.274 ESTs obtidas de 35 bibliotecas e disponibilizadas pelo Laboratório de Ge Genômica e Expressão (LGE - <http://www.lge.ibi.unicamp.br/cafe>) (Vieira *et al.*, 2006). A validação dos dados foi obtida pela busca tBLASTx e tBLASTn com matrix de pontuação BLOSUM80 da seqüência recuperada do PBGC contra o banco de dados do NCBI (National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/BLAST/>), construído pelo LGE. Os resultados de alinhamento foram filtrados por um limite de corte de *e-value* de 1e-15 para a seleção das ESTs e posterior análise. As seqüências validadas foram traduzidas (amino-ácido deduzido) e alinhadas usando o ClustalX (Thompson *et al.*, 1997).

Análise de domínios e caracterização in silico

A identificação dos genes homólogos foi posteriormente investigada pela presença de seqüência conservada de domínios funcionais reconhecidos e descritos em diversas análises de função de genes em banco de dados (European Bioinformatics Institute-European Molecular Biology Laboratory –EMBL-EBI www.ebi.ac.uk/interpro/; Expert Protein Analysis System – ExPaSy do Swiss Institute of Bioinformatics - SIB <http://www.expasy.org/prosite/> e <http://www.us.expasy.org/sprot/>; Protein Families database – Pfam <http://www.sanger.ac.uk/Software/Pfam/>).

Análise de expressão de genes in silico

O perfil quantitativo de expressão dos genes foi avaliado por análise *in silico* do banco de dados de EST, gerando uma matriz de relação entre o número de ESTs correspondentes para um determinado gene dado em uma biblioteca e normalizando-o ao número de ESTs de uma biblioteca. O padrão de expressão do gene dos contigs, EST e bibliotecas foram determinados por agrupamento hierárquico, baseado na matriz de correlação de Spearman e os resultados de agrupamento mostrados como sua média de expressão padrão.

Resultados e Discussão

Foram identificadas sete famílias gênicas que participam da biossíntese do THF em plantas (Tabela 1). Não foram encontradas seqüências significativamente similares ao gene homólogo codificante da nudix hidrolase, que participa da síntese da hidroximetildihidropterina. No total, foram identificadas 84 ESTs com similaridade significativa aos genes em estudo e, após a caracterização dos domínios e a análise de alinhamento, foram selecionadas 45 ESTs para a análise de expressão *in silico*, distribuídas em 12 *contigs* e 2 *singlets* (Tabela 1).

A identificação de duas ESTs no bando de dados PBGC, que formaram um contig, homólogas ao GTP ciclohidrolase I (GCHI) (Tabela 1) merece atenção, uma vez que esta enzima catalisa o primeiro passo da biossíntese de derivados de pterina, sendo considerada como controle de fluxo da produção de THF, como foi demonstrado em dois trabalhos. No primeiro, a expressão do gene GCH de *E. coli* em *Arabidopsis* aumentou 3,3 vezes o total de folato e 1250 vezes a pterina (Hossain *et al.*, 2004). No segundo, fruto maduro de tomate durante a fase de consumo para a alimentação humana foi modificado para o aumento do nível de expressão do GCHI, por meio da transformação com o gene sintético GCHI, em cassete com o promotor *ripening-specific* E8. O impulso do aumento de suplemento de *pABA* somado ao aumento da produção de pterina produziu ganhos adicionais na produção de folato (Diaz de la Garza *et al.*, 2004).

A quantidade relativa de ESTs por biblioteca foi avaliada para a análise de expressão de genes *in silico*. Em dez bibliotecas das 35 disponíveis no LGE não houve nenhuma seqüência que fosse significativamente similar aos genes em estudo por comparação utilizando o tBlastn e Blastn. Em geral, estas bibliotecas apresentaram um número menor de ESTs que as demais bibliotecas depositadas no LGE, sendo que cinco delas têm menos que 400 ESTs cada uma. Este resultado corrobora o fato de que, em geral, há uma concentração baixa de THF em plantas (Cossins, 1984). As 25 bibliotecas que apresentaram pelo menos uma seqüência com elevada similaridade aos genes de interesse foram categorizadas de acordo com os tecidos e ou órgãos de origem. As categorias das bibliotecas e os respectivos percentuais da expressão das ESTs normalizadas envolvidos na biossíntese do THF foram: Suspensão de Células (SC) com diferentes tratamentos- 19%; Folhas (FL) em diferentes estágios de desenvolvimento- 27% Botão Floral (BF) em diferentes estágios de desenvolvimento- 16%; Frutos em diferentes estágios de desenvolvimento, Botão floral e “pinhead” frutos (FR/BF)- 8%; Calos (CL) embriogênicos ou não- 17%; uma biblioteca (SI3) de germinação de semente- 4% e outras categorias misturadas (OT) que abrangem LP1, RX1 e SH2, respectivamente, plântulas e folhas tratadas com ácido araquidônico, nervuras inoculadas com patógeno e *pool* de tecidos de plantas tratadas com estresse hídrico- 9% (Figura 1).

Para melhor compreensão do perfil de expressão das famílias gênicas envolvidas na síntese do THF, os genes foram agrupados de acordo com a participação dos mesmos nas vias de síntese do THF (Figura 2A). A primeira via da síntese da hidroximetildihidropterina (produção de pterinas) e a segunda via da síntese da *pABA* apresentaram um nível bem menor de expressão de genes quando comparadas à última via da síntese do THF na mitocôndria (hidroximetildihidropterina, *pABA* e glutamato são combinados para produzir H₄F-Glu₁). Nesta última etapa, a família gênica da proteína bifuncional dihidrofolate reductase (DHFR), que é responsável pela redução do (dihidrofolato) H₂F-Glu₁ em H₄F-Glu₁ (tetraidrofolato), destacou-se pelo número de contigs formados (Tabela 1) bem como pelo nível de expressão (Figura 2B). Em *Arabidopsis* foram identificados três genes que codificam as proteínas DHFR-TS. Outra enzima presente nesta última etapa, e de interesse em estudo funcional, é a HPPK-DHPS, pois apresenta o nível de expressão

diretamente e comprovadamente relacionado com o aumento dos níveis de folato (Jabrin *et al.*, 2003). Esta enzima só foi encontrada nas bibliotecas de fruto e botão floral (FR2) e ramos infectados com *Xylella* (RX1).

As folhas em diferentes fases de desenvolvimento (FL) formaram a categoria de destaque quanto ao maior nível de expressão dos genes THF. Segundo Rébeillé *et al.* (2006), em folhas há uma maior quantidade de THF devido ao envolvimento no ciclo fotorrespiratório.

O fato dos níveis de folato serem bastante variáveis em diferentes espécies de plantas implica que há um potencial natural para o melhoramento do folato. A biofortificação do folato pode ser atingida explorando as variações naturais ou por meio de engenharia metabólica. Uma vertente para esta última estratégia seria modificar a biossíntese do folato.

Conclusão

A análise *in silico* do presente estudo revelou os genes específicos de café presentes na base de dados do PBGC e envolvidos na biossíntese do tetrahydrofolato. Este trabalho também contribuiu para inferir sobre o nível de expressão destes genes nos órgãos e tecidos do café, o que está relacionado com o nível de folato produzido. Novas perspectivas em agregar valores nutracêuticos a bebida do café estão sendo vislumbradas a partir do estudo funcional do metabolismo do folato em frutos de café.

Tabela 1– ESTs de *Coffea* homólogas a genes caracterizados funcionalmente, envolvidos no metabolismo do folato de *Arabidopsis thaliana* e outras espécies modelos.

<i>Arabidopsis thaliana</i>	EST de <i>Coffea</i>		Motifs de proteína ^c e		Referências	
Nome ^a	Gene	EST	% ^b	e value		processo biológico
Síntese Tetrahydrofolate em mitocôndria						
<i>DHFR</i>	At1g78680	C2 (At4g34570) ^a	70.4	1e-121	Dihydrofolato redutase bifuncional, timidilato reutase-sintase, glutamine amido transferase	Orsomando <i>et al.</i> , 2005
		Read(At1g78680)	60.2	1e-103		
		C4 (At4g34570) ^b	60.0	1e-133		
		C5 (At4g34570) ^c	45.0	1e-154		
<i>DHFS-FPGS</i>	At3g10160	C1 (At1g78680) ^d	42.5	9e-70	Di-hidrofolato sintase, folipoliglutamato di-hidropeteroato sintase	Ravel <i>et al.</i> , 2001
		C5 (At5g05980) ^e	53.9	1e-72		
		C4 (At5g05980) ^f	40.4	4e-72		
<i>HPPK-DHPS</i>	At1g69190	C2 (At3g10160) ^g	39.8	1e-55	Di-hidropeteroato sintase	Talarico <i>et al.</i> , 1992
		C1 (At4g30000) ^h	72.8	1e-125		
Síntese Hidroxymethyldihydropterin por meio do GTP						
<i>DHNA</i> <i>GTP</i>	At3g11750	Read (At5g62980)	39.7	1e-20	di-hidropterina aldolase	Thomas <i>et al.</i> , 1998
	*Le17979018	C1(Le17979018) ⁱ	71	e-118	Hidrolase de ligaçõs carbono-nitrogênio	Basset <i>et al</i> 2002
Síntese pABA- p-aminobenzoate no plastídeo						
<i>ADC synthase</i> <i>ADC lyase</i>	*Le40218381	C3(Le40218381) ^j	62	2e-100	Aminodeoxicorismato sintase domínios PabA e PabB	Basset <i>et al.</i> , 2004a
	At5g57850	C2 ^k	59.8	1e-120	Atividade 4-amino-4-deoxicorismato, biossíntese de THF, localizada no cloroplasto	Basset <i>et al.</i> , 2004b
	C1 ^l	44.8	3e-92			

^aAbreviações de nome de genes: *DHFS-FPGS*: dihydrofolate synthetase- folylpolyglutamate synthetase; *DHNA*: dihydroneopterin aldolase; *HPPK-DHPS*: hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase; *DHFR*: dehydrofolate reductase; *GTP*: GTP cyclohydrolase I; *ADC*: aminodeoxychorismate. ^bPorcentagem de identidade em nível de amino-ácido. ^cAbreviações do domínio de proteínas: THF: tetrahydrofolate. ^aContig2- 5 reads, ^bContig4 - 7 reads, ^cContig5- 3 reads, ^dContig1- 3 reads, ^eContig5- 2 reads, ^fContig4- 7 reads, ^gContig2- 2 reads, ^hContig1- 2 reads, ⁱContig1- 2 reads, ^jContig 3- 2 reads, ^kContig2- 3 reads, ^lContig1- 5 reads. **Lycopersicon esculentum*

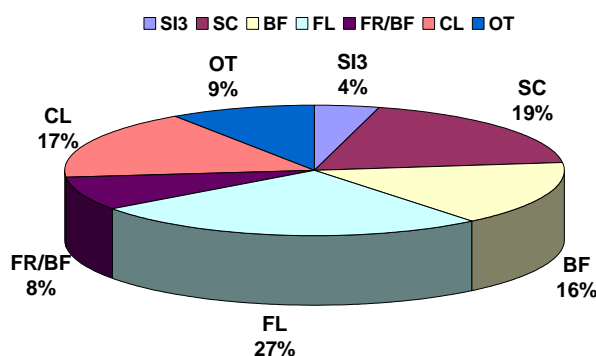


Figura 1 – Percentagem de ESTs funcionalmente caracterizadas na biossíntese do THF a partir dos dados de uma matriz de relação entre o número de ESTs correspondentes para um determinado gene ou família gênica e normalizando-o ao número de ESTs de uma biblioteca, agrupadas em categorias de acordo com o tecido e órgãos oriundos do café. **FL**- folhas

em diferentes estágios de desenvolvimento; **BF**- Botão floral em diferentes estágios de desenvolvimento, Frutos em diferentes estágios de desenvolvimento e botão floral; **SC**- Suspensão de células submetidos a diferentes tratamentos; **CI**- Calos embriogênicos ou não, **SI3**- germinação de semente e **OT**- outros .

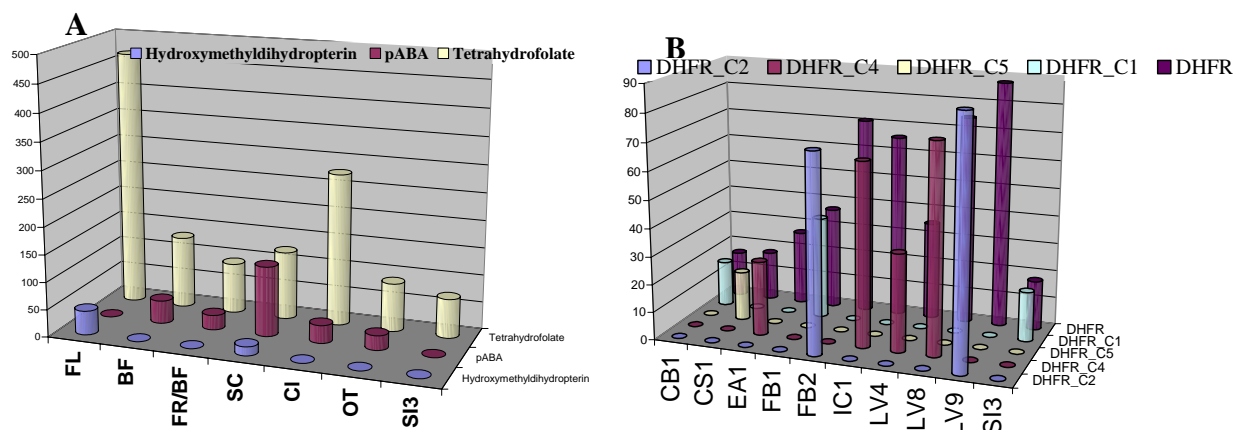


Figura 2 – Gráfico da distribuição da expressão das ESTs correspondentes a famílias gênicas em diferentes tecidos e órgãos do café. Os eixos estão representando, respectivamente, y - número de ESTs normalizadas, x - categorias das bibliotecas e z- famílias gênicas. **A**- Expressão das ESTs divididas em três grupos constituídos pelas famílias gênicas que se seguem: **1**- hidroximetilhidropterina- GTP ciclohrolase I e dihidroneopterina aldolase; **2**- pABA-aminodeoxychorismate (ADC) sintase e ADC liase e **3**- THF- hidroximetildihidropterina pirofosfoquinase e dihidropteroato sintase (HPPK-DHPS) bifuncional; dihidrofolate sintetase (DHFS) e folilpoliglutamato sintase (FPGS) e por último dihidrofolate reductase (DHFR). **B**- Análise da expressão de ESTs pertencentes a família gênica da dihidrofolate reductase (DHFR). As categorias das bibliotecas são: **FL**- folhas em diferentes estágios de desenvolvimento; **BF**- Botão floral em diferentes estágios de desenvolvimento, Frutos em diferentes estágios de desenvolvimento e botão floral; **SC**- Suspensão de células submetidos a diferentes tratamentos; **CI**- Calos embriogênicos ou não, **SI3**- germinação de semente e **OT**- outros .

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.
- Basset G.J., Quinlivan, E.P., Ravel, S., Rébeillé, F., Nichols, B.P., Shiozaki, K., Seki, M., Adams-Phillips, L.C., Giovannoni, J.J., Gregory, J.F III, Hanson, A.D. (2004a) Folate synthesis in plants: the p-aminobenzoate branch is initiated by a bifunctional PabA-PabB protein that is targeted to plastids. *Proc. Natl. Acad. Sci. USA* 101: 146-1501.
- Basset GJ, Ravel S, Quinlivan EP, White R, Giovannoni JJ, Rebeille F, Nichols BP, Shiozaki K, Seki M, Gregory JF III and Hanson AD (2004b) Folate synthesis in plants: the last step of the p-aminobenzoate branch is catalyzed by a plastidial aminodeoxychorismate lyase. *Plant J* 40:453-461
- Basset, G., Quinlivan, E.P., Ziemak, M.J., Díaz de la Garza, R., Fischer, M., Bacher, S.S. A., Gregory III, J.F. e Hanson, A.D. (2002) Folate synthesis in plants: The first step of the pterin branch is mediated by a unique bimodular GTP cyclohydrolase I. *PNAS*. 99 (19): 12489–12494 www.pnas.org/cgi/doi/10.1073/pnas.192278499.
- Diaz de la Garza, R., Quinlivan, E.P., Klaus, S.M., Basset, G.J. (2004). Folate biofortification in tomatoes by engineering the pteridine branch of folate synthesis. *PNAS*. 101 (38) 13720-13725.
- Hanson, T. e Roje, S. (2001) One-carbon metabolism in higher plants. *Annu Rev. Plant Physiol. and Plant Mol Biol* 52:119-137.
- Hossain, T., Rosenberg, I., Selbhu, J., Kishore, G., Beach, R., Schubert, K. (2004) Enhancement of folate in plants through metabolic engineering. *Proc Ntl Acad Sci USA* 101: 5158-5163.
- Jabrin, S., Ravanel, S., Gambonnet, B., Douce, R. Rébeillé, F. (2003) One-Carbon metabolism in plants. Regulation of tetrahydrofolate synthesis during germination and seedlings development. *Plant. Physiol.* 131: 1431-1439.
- Orsomando G, de la Garza RD, Green BJ, Peng M, Rea PA, Ryan TJ, Gregory JF 3rd and Hanson AD (2005) Plant gamma-glutamyl hydrolases and folate polyglutamates: characterization, compartmentation, and co-occurrence in vacuoles. *J Biol Chem* 280:28877-28884
- Ravel S, Cherest H, Jabrin S, Grunwald D, Surdin-Kerjan Y, Douce R and Rebeille F (2001) Tetrahydrofolate biosynthesis in plants: molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 98:15360-15365

Rébeillé F., Ravanel, S., Jabrin, S., Douce, R., Storozhenko, S., Van Der Straeten, D. (2006) Folates in plants: biosynthesis, distribution, and enhancement. *Physiologia Plantarum*. 126: 330-342.

Talarico TL, Ray PH, Dev IK, Merrill BM and Dallas WS (1992) Cloning, sequence analysis, and overexpression of *Escherichia coli folK*, the gene coding for 7,8-dihydro-6 hydroxymethylpterin-pyrophosphokinase. *J Bacteriol* 174:5971-5977

Thomas MC, Ballantine SP, Bethell SS, Bains S, Kellam P and Delves CJ (1998) Single amino acid substitutions disrupt tetramer formation in the dihydroneopterin aldolase enzyme of *Pneumocystis carinii*. *Biochemistry* 37:11629-11636

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 24:4876-4882.

Vieira L.G.E., Andrade, A.C, Colombo, C.A, Moraes, A.H. de A., *et al.* (2006) Brazilian coffee genoma project: an EST-based genomic resource. *Braz. J. Plant Physiol.*, 18(1):95-108.