Universidade Estadual de Campinas Instituto de Biologia Departamento de Genética e Evolução

Análise Filogenética de Fatores de Transcrição bZIP em Angiospermas

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas, para obtenção do título de Mestre em Genética e Biologia Molecular, área de concentração Genética Vegetal e Melhoramento.

Luiz Gustavo Guedes Corrêa

Orientador: Prof. Dr. Michel Georges Albert Vincentz

Campinas – SP 2004

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA - UNICAMP

Corrêa, Luiz Gustavo Guedes

C817a Análise filogenética de fatores de transcrição bZIP em angiospermas / Luiz Gustavo Guedes Corrêa. --Campinas, SP:[s.n.], 2004.

> Orientador: Michel Georges Albert Vincentz Dissertação (mestrado) – Universidade Estadual de Campinas. Instituto de Biologia.

1. Angiospermas. 2. Filogenia. 3. Proteína. I. Vincentz, Michel Georges Albert. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

iii

Campinas, 02 de julho de 2004.

Banca Examinadora

Prof. Dr. Michel Georges Albert Vincentz (Orientador)

Prof. Dr. Carlos Frederico Martins Menck

Profa. Dra. João Paulo Fumio Whitaker Kitajima

Suplente:

Prof. Dr. José Andrés Yunes

Assinatura

Assinatura

Assinatura

Assinatura

O Rio e o Oceano

Antes de um rio cair no oceano, ele treme de medo.Olha para trás, para os cumes, as montanhas, o longo caminho sinuoso através das florestas e vê a sua frente um oceano tão vasto que, entrar nele, nada mais é do que desaparecer para sempre.

Mas não há outra maneira! O rio não pode voltar! Ninguém pode voltar! Voltar é impossível na existência. O rio pode apenas ir em frente. O rio precisa se arriscar e entrar no oceano. E somente quando ele entra no oceano é que o medo desaparece, porque então o rio saberá que não se trata de desaparecer no oceano, mas tornar-se oceano. Por um lado é desaparecimento e por outro lado é renascimento.

Assim somos nós... Voltar é impossível na existência. Você pode ir em frente e se arriscar.

Coragem, torne-se oceano.

(Autor desconhecido)

O primeiro dever da inteligência é desconfiar de si mesma.

(Stanislaw Kersy Lec, escritor polonês)

Para Ana Paula e Fabiana, meus dois anjos da guarda.

Agradecimentos

Ao meu orientador, Michel, pela enorme paciência e horas de dedicação. Obrigado por ter apostado em mim. *Merci beaucoup*!

Aos doutores Carlos Menck e João Paulo Kitajima, membros da banca.

Aos membros da pré-banca João Paulo Kitajima, João Meidanis e Gonçalo Guimarães pelas críticas, revisões e sugestões.

Ao estimado Dr. Adilson Leite (in memorium) por seu grande empenho e exemplo.

Às meninas da secretaria, Tânia e Sandra, por conseguirem resolver nossos problemas e conseguirem liberar minhas senhas sempre de última hora.

Ao pessoal da limpeza, em especial Rosilda e Solange, por terem cuidado tão bem do laboratório e pela grande amizade desenvolvida nestes anos.

Aos três grandes "Joãos" (Kitajima, Meidanis e Setúbal) que me iniciaram neste mundo da bioinformática, não poderia ter tido melhores mentores. E um quarto João (Piazza) pela ajuda na montagem dos bancos de dados e pela resposta aos mais de 400 emails intitulados "help" ou "urgente".

À todos os meus professores de graduação e pós por minha formação e espírito crítico.

Ao amigo Rodrigo Marins, por sua paciência em nossos ensaios com proteínas e me fazer acreditar na entidade do fentomol.

À Syngenta e ao Instituto de Genômica de Pequim pelas seqüências genômicas de arroz.

Ao Martin Lomas (Universidade de Leeds, Inglaterra) pelas seqüências EST de *Physcomitrella patens*.

Ao Kazuza DNA Research Institute e ao Riken Institute of Physical and Chemical Research pelos cDNAs de AtbZIP76 e AtbZIP78.

Ao Prof. Dr. Robert Grebenok por nos ter cedido os controles positivos para os ensaios de localização nuclear.

Aos amigos Renato e Fisch pela ajuda em todos os momentos e discussões que tivemos durante todos estes anos. Parte deste trabalho aconteceu graças a vocês.

Ao Dr. Nilson Zanchin (LNLS) pelo auxílio nas análises microscópicas.

A Sandra Camargo pela grande força que nos deu nas transformações de cebola.

À Sandra Martins por seu auxílio técnico.

Às minhas amigas Adriana, Letícia e Luciane que me orientaram e desorientaram durante meu mestrado.

À Luciane, obrigado por seus conselhos e sua amizade. Você é uma pessoa impar.

À Roberta, minha primeira orientada, desculpe pelas minhas falhas. A Amanda e Mina, coragem no caminho que estão trilhando.

Aos amigos de laboratório Luciane, Fernanda, Amanda, Roberta e Natália pelos divertidos momentos que passamos juntos, pelos históricos almoços lá em casa e pela oportunidade que tive de aprender com vocês. Aos novos membros do laboratório, Mina, Bruno, Dudu e Juarez por darem continuidade àquilo que eu ajudei a começar.

À Cláudia Kobarg, por ter me ajudado nos primeiros passos da biologia molecular.

Aos amigos dos laboratórios de biologia molecular de plantas, genoma e genética animal (o laboratório da frente) pelo convívio.

Aos meus amigos da turma 98 diurno de Biologia da Unicamp, foi uma prazer estar com vocês neste "trem" durante estes 6 anos, mas acho que minha estação está chegando.

À Natália, Thaís, Sylvia e Mário, amigos com os quais pude compartilhar vários momentos deste mundo de genética de plantas. Nós nos devemos ainda um "ultimate".

Ao meu grande amigo Luís Peroni (Caxias) por todos este anos de companheirismo e por colocar todas as nossas idéias em planilhas auto-explicativas.

Aos amigos Juba, Cams, Marcão, Natasha, Christian e Silvia por seu apoio incondicional. Vocês são minha família aqui.

Aos meus novos amigos do curso de especialização de Petrópolis, por terem se dedicado tanto na resolução dos nossos problemas NJ, MP e MV.

À Eliete, por ter sido como uma mãe para mim aqui em Campinas.

Aos meus pais, Vânia e José Luiz, por seu exemplo de vida e dedicação à minha formação. Vocês são a melhor coisa da minha vida.

À Bruna, minha grande companhia.

Aos meus parentes, em especial meus avós, que mesmo estando longe sempre estiveram presentes nos momentos mais importantes da minha vida.

Aos meus amigos de Ipatinga, pelo fato de existirem em minha vida.

Aos Ronzi, pela oportunidade de conhecer pessoas especiais e ter contato com uma cultura tão viva. *Je vous addore*!

Aos Kunt, minha segunda família, por seu grande afeto e amizade que fizeram laços eternos. *Sizi seviyorum*!

À todos aqueles que, de uma forma ou de outra, contribuíram para a realização desta tese.

A CAPES, CNPq e FAPESP pelo auxílio financeiro.

Lista de Abreviações	xi
Lista de Figuras e Tabela	 vii
Resumo	viii
Abstract	xiv
1. Introdução	1
1.1 Regulação da expressão gênica em eucariotos	1
1.2 Os fatores de transcrição bZIP	3
1.3 Grupos de Ortólogos	6
1.4 Filogenia	. 11
1.5 Objetivos	13
2. Material e Métodos	14
2.1 Material Computacional	14
2.2 Obtenção das seqüências	14
2.3 Construção de um banco não redundante de fatores de transcrição do tipo bZIP o	de
Angiospermas.	16
2.4 Montagem das seqüências de EST em <i>clusters</i>	16
2.5 Estabelecimento da metodologia para identificar seqüências genômicas de arroz	е
clusters de ESTs que possivelmente codifiquem para fatores de transcrição do tipo bZIP	17
2.6 Análise de redundância	18
2.7 Análise filogenética	18
2.8 Identificação de Possíveis Grupos de Ortólogos (PoGO)	20
2.9 Análise de duplicações no genoma de Arabidopsis thaliana	20
2.10 Obtenção de fusões traducionais de AtbZIP76 e AtbZIP78 com o gene repórter RF	° 21
2.11 Transformação por biobalística	23
2.12 Microscopia de Fluorescência	24
0. Desultados o Discussão	05
3. Resultados e Discussão	25
3.1 Identificação de genes que codificam para fatores de transcrição do tipo b21F	em
	26
3.2 Identificação de DZIPS em Dancos de ESIS	28
3.3 Analise filogenetica das sequencias de bZIP de arroz e A. thaliana para	a
identificação de grupos de genes nomologos	30
3.4 Analises Filogeneticas entre Mono e Dicotiledôneas e formação de PoGOs	38
3.5 Possível modelo evolutivo de bZIPs em Angiospermas	46

Índice

	3.6 Utilização da classificação em PoGOs para análise funcional	. 49
4.	. Conclusões e Perspectivas	. 58
5.	. Referências Bibliográficas	60

Lista de Abreviações

Am	Antirrhinum majus
At	Arabidopsis thaliana
Bn	Brassica napus
bZIP	basic leucine zipper
Cc	Capsicum chinense
cDNA	DNA complementar
Cr	Catharanthus roseus
DNA	ácido desoxirribonucléico
dN	taxa de substituição não sinônima
dS	taxa de substituição sinônima
EM	Evolução Mínima
EST	Etiqueta de Següência Expressa
GFP	Gree Fluorescent Protein
Gm	Glycine max
Ha	Helianthus annuus
Ηα	mercúrio
Hv	Hordeum vulgare
le	Licopersicon esculentum
Li	Lotus corniculatus var japonicus
J Т	
μL	
	Máxima Dancincânia
	Maxima Parcimonia
IVIT	Medicago truncatula
IVIV	Maxima verossimilnança
NJ	Neighbor-joining
nm	nanometro
Nt	Nicotiana tabacum
Us	Oryza sativa
Ра	Phaseolus acutitolis
рр	pares de base
PC	Petroselinum crispum
PCR	reação de polimerase em cadeia
PK	Paulownia kawakamii
PoGO	Possível Grupo de Ortólogos
PoGP	Possível Grupo de Paralogos
Pv	Phaseolus vulgaris
RFP	Red Fluorescent Protein
RNA	ácido ribonucleico
RNAm	RNA mensageiro
rpm	rotações por minuto
Rs	Raphanus sativus
Sa	Sinapis alba
Sb	Sorghum bicolor
SI	Spinacia oleracea
So	Saccharum ssp.
St	Solanum tuberosum
la	Triticum aestivum
Vf	Vicia faba
Vv	Vitis vinifera
Zm	Zea mays

Lista de Figuras e Tabelas

Figura 1. Representação esquemática da estrutura tridimensional do domínio conserva bIP	ado 4
Figura 2. Possíveis destinos de genes duplicados segundo uma trajetória hierárquica Figura 3. Reclassificação das angiospermas	7 9
Figura 4. Relações evolutivas de homologia entre genes	9
Figura 5. Esquema do protocolo bioinformático para a identificação de bZIPs	em
seqüências genômicas e EST	.19
Figura 6: Esquemas dos cassetes de expressão dos plasmídeos Pol RFP:::AtbZIP76	6 e
Pol RFP:::AtbZIP78 utilizados nos ensaios de expressão transiente em epiderme	de
cebola	.22
Figura 7. Classificação das bZIPs de arroz e Arabidopsis	.31
Figura 8. Definição dos grupos de genes homólogos A, D e F	.33
Figura 9. Posições conservadas de introns na região do domínio básico dos fatores	de
transcrição do tipo bZIP de angiospermas	.34
Figura 10. Árvore filogenética de bZIPs de arroz e Arabidopsis	.37
Figura 11. Árvore filogenética de bZIPs de mono e eudicotileôneas do Grupo C	.40
Figura 12. Árvore filogenética de bZIPs de mono e eudicotileôneas dos Grupos B e H	.41
Figura 13. Árvore filogenética de bZIPs de mono e eudicotileôneas do Grupo G	.42
Figura 14. Árvore filogenética de bZIPs de mono e eudicotileôneas do Grupo I	.43
Figura 15. Árvore filogenética de bZIPs de mono e eudicotileôneas dos Grupos J e S	.45
Figura 16. Modelo evolutivo dos onze grupos de genes homólogos de bZIPs	em
angiospermas	.47
Figura 17. Distribuição do número de PoGO e de genes nos vários Grupos	de
homólogos.	.49
Figura 18. Árvore filogenética de bZIPs de mono e eudicotileôneas do Grupo D	.51
Figura 19. Árvore filogenética de bZIPs de mono e eudicotileôneas do Grupo A	.54
Figura 20. Alinhamento do domínio bZIP do Grupo E	.55
Figura 21. Ausência de NLS B bipartido nos bZIPs do PoGO E1	.56
Figura 22. Análise da localização celular de AtbZIP76 e AtbZIP78	.57
Tabela I. Porcentagem de genes que codificam para fatores de transcrição em vár	ios
organismos	2
Tabela II. Distribuição de bZIPs nos genomas de eucariotos sequenciados	3
Tabela III. Lista de programas utilizados	.14
Tabela IV. Oligonucleotideos utilizados na cionagem de AtbZIP76 e AtbZIP78	.22
Tabela VI. Distribuição das DZIPS entre as diferentes fontes de sequencias	.28
Tabela VII. NUMERO de proz que pão forom utilizados no análico filosonático.	.29
Tabela VIII. Deres de genes regultantes de duplicação de segmentos de genero.	.30 do
Arabidopsis	.36
Tabela IX. Genes únicos	.55

Resumo

O Crescimento e desenvolvimento de todos os organismos dependem de uma regulação adequada da expressão gênica. A expressão gênica diferencial resulta do controle da taxa de iniciação da transcrição por fatores de transcrição. Os fatores de transcrição do tipo bZIP foram descritos em todos os eucariotos. Seu domínio conservado é constituído de uma região rica em aminoácidos básicos que se ligam às bases do DNA, e o zíper de leucinas, responsável pela dimerização. Análises genéticas, moleculares e bioquímicas indicam que os bZIPs são importante reguladores de processos específicos de angiospermas. A identificação de possíveis grupos de genes ortólogos (PoGOs) a partir de diferentes linhagens evolutivas permite racionalizar estudos funcionais. Dentro deste contexto, o presente trabalho visa um melhor conhecimento sobre a evolução e função dos fatores de transcrição do tipo bZIP em angiospermas. Identificamos um conjunto não redundante de 77 bZIPs no genoma de Arabidopsis e 113 no genoma de Oryza sativa. Análises filogenéticas entre o conjunto de Oryza e Arabidopsis permitiram identificar 11 grupos de genes homólogos. Análises mais detalhadas levaram à identificação de 33 PoGOs, que possivelmente representam 33 funções ancestrais de angiospermas, além de um PoGO exclusivo a monocotiledôneas e um possível grupo de parálogos exclusivo a Arabidopsis. Baseado no conjunto de resultados apresentado, propomos uma classificação atualizada das bZIP de angiospermas e um modelo evolutivo das mesmas. Esta abordagem está abrindo novas perspectivas para o estudo funcional de bZIPs. Ensaios de expressão transiente de fusões traducionais de AtbZIP76 e AtbZIP78 com RFP em células de cebola mostram que estes bZIPs são direcionados ao núcleo, apesar de possuírem um domínio básico defectivo.

Abstract

Growth and development of all organisms depend on proper regulation of gene expression. Differential gene expression is the result of the control of transcription initiation rates by transcription regulation factors. bZIP transcription factors were described in all eukaryotes. Their conserved domain is constituted by a basic aminoacids rich region that bond to DNA bases, and a leucine zipper, responsible for dimerization. Genetic, molecular and biochemical analysis indicate that bZIP factors are important regulator of specific angiosperm processes. Identification of possible groups of orthologous genes (PoGO) from different evolutionary lineages allows rationalizing functional studies. Within this context, the present work aims a better knowledge of the evolution and function of bZIP transcription factors in angiosperms. We identified a non-redundant set of 77 bZIP factors in the Arabidopsis genome and 113 bZIP in Oryza sativa genome. Phylogenetic analysis of Oryza and Arabidopsis sets of bZIP allowed identifying 11 groups of homologue genes. More detailed analysis led to the identification of 33 PoGOs, that possibly correspond to 33 ancestral functions in angiosperms, a exclusive monocot PoGO and a possible groups of paralogous genes in Arabidopsis. Based on the results presented here, we propose an updated bZIP classification and a model of their evolution. This approach is revealing new perspectives on bZIP functional studies. Transient expression essays of traductional fusions of AtbZIP76 and AtbZIP78 with RFP in onion cells show that these bZIPs do target the nucleus, despite their defective basic domain.

1.Introdução

1.1. Regulação da expressão gênica em eucariotos

O crescimento e o desenvolvimento de todos organismos depende, em sua maior parte, da regulação apropriada da expressão gênica. A regulação da expressão gênica pode se dar em diferentes níveis: transcrição, processamento de RNAm, transporte do RNAm para o citoplasma, tradução, pós-tradução e multimerização. No entanto, o controle da taxa de iniciação da transcrição por fatores de regulação representa o meio principal utilizado pelas células para modular a expressão gênica (Meshi & Iwabuchi, 1995 Beckett, 2001; Warren, 2002; Wray *et al.*, 2003). Os fatores de regulação da transcrição são conhecidos como elementos *trans* e podem ser definidos como proteínas que se ligam a seqüências específicas de DNA (elementos *cis*) na região montante aos genes e são capazes de modular a taxa de iniciação da transcrição (Kuhlemeier, 1992; Blackwood & Kadonaga, 1998; Singh, 1998; Holstege & Young, 1999; Kornberg, 1999; Lee & Young, 2000; Riechmann & Ratcliffe, 2000; Beckett, 2001; Warren, 2002).

Os reguladores da transcrição podem ser agrupados em famílias de proteínas relacionadas que apresentam similaridades na estrutura primária e/ou tridimensional do seu domínio de ligação ao DNA e de multimerização (Riechmann et al., 2000; Wingender et al., 2000; Warren, 2002).

Algumas famílias de fatores de transcrição, como bHLH (*basic Helix Loop Helix*), MYB e bZIP (*basic leucine zipper*), são reguladores compartilhados entre todos os eucariotos, e, portanto, já estavam presentes antes da divergência das principais linhagens eucarióticas (plantas, animais e fungos) datada de aproximadamente 1,5 bilhões de anos atrás (Riechmann *et al.*, 2000; Wingender *et al.*, 2000). No entanto, cada uma destas

1

linhagens evolutivas desenvolveu novos tipos de proteínas que estão envolvidas na regulação da transcrição como: AP2 (APETALA2) e WRKY em plantas; NHR e Adf-1 em animais; e C6 e Swi4/Swi6 em fungos (Riechmann *et al.*, 2000).

Os genes que codificam para fatores de transcrição representam uma parte significativa do genoma de eucariotos, sendo que em *Arabidopsis* eles correspondem a 5,9% dos genes, uma proporção maior do que a encontrada nos genomas de drosófila (4.5%), *C. elegans* (3.5%) e levedura (3.5%) (Tabela I, Riechmann *et al.*, 2000).

			Genes que codificam para fatores de transcrição	
Organismo	Número total de genes	Total	Porcentagem	
A. thaliana	~26.000*	1533	5.9	
S. cerevisae	~6.000	209	3.5	
C. elegans	~19.000	669	3.5	
D. melanogaster	~14.000	635	4.5	

Tabela I. Porcentagem de genes que codificam para fatores de transcrição em vários organismos.

* Atualmente este valor está em aproximadamente 28.000 genes (Yamada *et al*, 2003) Tabela retirada de Riechmann *et al.*, 2000.

Ao contrário de outros organismos, o genoma de *Arabidopsis* não apresenta nenhuma família multigênica que tenha sido amplificada desproporcionalmente, como é o caso dos receptores nucleares de hormônio em *C. elegans* (~38% dos seus fatores de transcrição), o zinc finger do tipo C2H2 em drosófila (~46%) e C6 e C2H2 em leveduras (25% cada). As famílias de fatores de transcrição mais representativas em plantas são AP2, MYB e bHLH, cada uma correspondendo a 9% do total de fatores de transcrição

(Riechmann *et al.*, 2000). A amplificação preferencial de certos tipos de reguladores pode estar relacionada com a especificidade e adaptação de cada linhagem (Lespinet *et al.* 2002).

1.2. Os fatores de transcrição bZIP

Os fatores de regulação de transcrição do tipo *basic leucine zipper* (bZIP) foram descritos em todos os organismos eucariotos (Wingender *et al.*, 2000). Os possíveis conjuntos não redundantes de fatores bZIP para quase todos os organismos eucarióticos que possuem o genoma completo seqüenciado já foram identificados (Tabela II; Riechmann *et al.*, 2000; Tupler *et al.*, 2001; Fassler *et al.*, 2002; http://compbio.cs.princeton.edu/bzip/). bZIPs são caracterizadas por um domínio conservado de 40 a 80 aminoácidos, conhecido como domínio bZIP. Este domínio é constituído por dois motivos: uma região básica e um zíper de leucina (Hurst, 1995; Figura 1).

Tabela II. Distribuição de bZIPs nos genomas de eucariotos seqüenciados		
Genoma	Número de bZIPs	
Anopheles gambiae (mosquito)	23	
Arabidopsis thalania	77	
Caenorhabditis elegans (verme)	31	
Drosophila melanogaster (mosca)	27	
Danio rerio (paulistinha)	51	
Fugu rubripes (baiacu)	54	
Homo sapiens (humano)	65	
Saccharomyces cerevisiae (levedura) 17		
Tabela retirada de http://compbio.cs.princeton.edu/bzip/		

A região básica é constituída por cerca de 30 aminoácidos básicos e é responsável pela especificidade de ligação ao DNA e por seu endereçamento nuclear (Izawa *et al.*, 1993; Varagona & Raikhell, 1994). Esta região se encaixa no sulco maior da seqüência de DNA alvo, onde cadeias laterais de aminoácidos básicos estabelecem ligações hidrofóbicas ou pontes de hidrogênio com as bases do sítio alvo (Ellenberg *et al.*, 1992; König & Richmond, 1993; Hurst, 1995; Figura 1).



Figura 1. Representação esquemática da estrutura tridimensional do domínio conservado bZIP. O motivo básico é responsável pela especificidade de ligação ao DNA e endereçamento nuclear. O zíper de leucina é o motivo responsável pela dimerização do fator.

O zíper de leucinas é constituído por repetições de resíduos hidrofóbicos, principalmente de leucina, a cada sete aminoácidos em uma extensão de 30 a 60 aminoácidos, sendo que o número de repetições pode variar de três a nove (Hurst, 1995,

Jacoby *et al.*, 2002). O zíper de leucinas, devido a sua característica hidrofóbica, é responsável pela dimerização dos bZIPs, que podem formar homo ou heterodímeros funcionais (Figura 1). As diferentes combinações entre bZIPs podem gerar diferentes níveis de controle da expressão gênica, uma vez que diferentes combinações podem ter diferentes afinidades de ligação ao DNA, entre outras coisas (Chiu *et al.*, 1989; Sigh, 1998; Wray *et al.*, 2003). Estudos cristalográficos do fator bZIP GCN4 de levedura ligado a sua seqüência de DNA alvo revelaram que o domínio bZIP forma uma α -hélice e Os dois zíperes de leucinas interagem produzindo uma estrutura "coiled-coil" (Ellenberg *et al.*, 1992; Hurst, 1995; Luscombe *et al.*, 2000; Figura 1).

Análises genéticas, moleculares e bioquímicas indicam que os fatores bZIP são reguladores importantes de processos específicos de angiospermas como o desenvolvimento de órgãos (Walsh *et al.*, 1997; Chuang *et al.*, 1999); elongação celular (Yin *et al.*, 1997; Fukuzawa *et al.*, 2000); controle do balanço nitrogênio/carbono (Ciceri *et al.*, 1999); mecanismos de defesa (Niggeweg *et al.*, 2000; Zhang *et al.*, 1999; Despres *et al.*, 2000; Pontier *et al.*, 2001); via de sinalização de hormônios e da sacarose (Rook *et al.*, 1998; Choi *et al.*, 2000; Finkelstein & Lynch, 2000; Uno *et al.*, 2000; Niggeweg *et al.*, 2000); resposta à luz (Schindler *et al.*, 1992; Wellmer *et al.*, 1999; Osterlund *et al.*, 2000; Ulm, *et al.*, 2004); e controle osmótico (Satoh *et al.*, 2004).

Inicialmente, cerca de 50 bZIPs de angiospermas foram identificados em plantas e classificadas em cinco famílias, de acordo com sua similaridade na região bZIP (Vettore *et al.*, 1998). Com o final do seqüenciamento do genoma de *Arabidopsis thaliana*, observouse que ele possivelmente codifica para 81 fatores bZIP (*Arabidopsis* Genome Iniciative, 2000; Riechmann *et al.*, 2000). Após uma análise mais detalhada, foram identificados 77 fatores bZIP que provavelmente representam o conjunto completo e não redundante deste tipo de fator em *A. thaliana*. A relação evolutiva entre membros desta coleção foi avaliada e resultou numa organização em famílias de proteínas evolutivamente relacionadas (Jacoby *at el.*, 2002; Vincentz *et al.*, 2003).

1.3. Grupos de Ortólogos

Duplicação gênica é um processo essencial na diversificação e adaptação de organismo durante a evolução, uma vez que genes duplicados são fontes de material para gerar novas funções através de mutações e seleção natural (Alvarez-Buylla *et al.*, 2000; Lawton-Rauh, 2003; Vandepoele *et al.*, 2003; Kellis *et al.*, 2004). A obtenção de novas cópias gênicas por recombinação não homóloga e/ou por duplicação de segmentos genômicos e de genomas inteiros, são mecanismos que levam à formação de famílias multigênicas (Wendel, 2000; Bennetzen, 2002; Kellis *et al.*, 2004). Estes mecanismos vêm contribuindo significativamente, ao longo da evolução, para moldar o formato atual dos genomas eucariotos conhecidos (Lespinet *et al.*, 2002; Lynch, 2002). Este fato pode ser ilustrado em eucariotos com *Arabidopsis*, onde as famílias multigênicas correspondem a 65% do seu genoma (Wendel, 2000).

Inicialmente, um gene duplicado apresenta redundância funcional em relação ao gene do qual se originou. A partir deste ponto, quatro podem ser os destinos destas duas cópias. Primeiro, se ambas as cópias mantêm sua função original, de modo que uma pode substituir integralmente a outra, então as cópias são funcionalmente redundantes (redundância funcional, Figura 2). Segundo, se as cópias apresentam mutações de forma que a função desempenhada por um único gene ancestral passa a ser desempenhada pela

ação em conjunto de ambas as cópias, então se tem uma subfuncionalização (Figura 2). No caso de uma cópia manter a função original, a outra pode acumular diferentes mutações, fazendo com que esta não desempenhe mais a mesma função ancestral, ou não participe mais de um mesmo processo. Deste ponto, este gene pode ser aproveitado em um processo diferente do processo original e/ou em um novo processo (neofuncionalização, Figura 2). Alternativamente, as mutações acumuladas podem resultar na formação de um pseudogene (revisado por Lawton-Rauh, 2003; Figura 2). Pseudogenes são seqüências inativas de DNA genômico que possuem similaridade com genes funcionais, e devido a esta similaridade, eles são considerados como evolutivamente relacionados aos seus pares funcionais.



Figura 2. Possíveis destinos de genes duplicados segundo uma trajetória hierárquica. Após a duplicação gênica, as duas cópias geradas apresentam redundância funcional. Esta redundância pode se manter ao longo da evolução, ou pode haver subfuncionalização, neofuncionlização ou formação de pseudogene (Figura retirada e modificada de Lawton-Rauh, 2003)

Estudos recentes de comparação de genomas de duas espécies de levedura, *K. waltii* e *S. cerevisae*, demonstraram que para 95% dos 457 genes duplicados, apenas uma das cópias sofre uma evolução acelerada (gene derivado), enquanto a outra cópia mantém a informação original, permitindo haver uma distinção entre as funções ancestrais e derivadas (Kellis *et al.*, 2004). Este resultado sugere que a neofuncionalização foi o mecanismo mais privilegiado durante o processo evolutivo em leveduras (Moore & Purugganan, 2003).

O seqüenciamento de Arabidopsis thaliana (Arabidopsis Genome Iniciative, 2000), uma eudicotiledônea (Figura 3; Savolainen & Chase, 2003; uma revisão mais detalhada pode ser vista na página http://www.csdl.tamu.edu/FLORA/newgate/apg1ang.htm), abriu uma fronteira para o estudo de genomas em angiospermas, uma vez que este corresponde ao primeiro genoma completo seqüenciado de uma planta, o que permitiu obter o conjunto gênico completo relativo a uma angiosperma, formando um esqueleto de base para estudos comparativos com os genomas de outros organismos (Soltis & Soltis, 2003). Com o término do seqüenciamento dos genomas de arroz (Goff et al., 2002; Yu et al., 2002), uma monocotiledônea (Figura 3), a comparação do seu genoma com o de Arabidopsis permitiu explorar as bases genéticas da diversidade das duas grandes divisões de angiospermas: mono e eudicotiledônea. Um dos aspectos da abordagem comparativa envolve o estabelecimento de relações entre genes de diferentes genomas em um sistema de genes homólogos, que inclui genes ortólogos e parálogos (Bennetzen, 2002; Pennacchio, 2003; Vincentz et al., 2004). O termo ortólogo refere-se aos genes homólogos que divergiram a partir de um evento de especiação (são versões do mesmo gene em organismos diferentes), e genes parálogos são genes homólogos que resultaram de um evento de duplicação dentro de um genoma (Tatusov et al., 1997; Thornton & DeSalle, 2000; Fitch, 2000; Meyerowitz,

2002, Figura 4). Genes ortólogos normalmente conservam a mesma função enquanto genes parálogos podem desenvolver novas funções (neofuncionalização).



Figura 3. Reclassificação das angiospermas. Uma nova classificação de angiospermas com base em estudos de filogenia molecular resultou na divisão das dicotiledôneas (A) em 4 grupos distintos: eudicotiledôneas, dicotiledôneas primitivas, Nymphea e *Amborella* (B). A junção das dicotiledôneas primitivas com as monocotiledôneas reflete a presença de pólen com uma abertura apenas, ao contrário das eudicotiledôneas, que possuem 3 aberturas. Tanto Nymphaea quanto *Amborella* são linhagens basais, uma vez que apresentam elementos primitivos. Modificada de Savolainen & Chase (2003)



Figura 4. Relações evolutivas de homologia entre genes. Genes ortólogos são originados de eventos de especiação (E). Genes parálogos são originados de eventos de duplicação

(D). A1 e A2 são genes parálogos originados por duplicação do gene A na espécie ancestral, e dão origem a dois grupos de ortólogos, A1' e A2' nas espécies I e II. A3' é parálogo de A2' e pode assumir uma função diferente deste.

As relações de ortologia entre genes de família multigênicas podem ser avaliadas através da definição de grupos de genes ortólogos (Tatusov et al., 1997). Cada grupo de ortólogo é assumido como sendo o resultado da evolução de um gene ancestral através de eventos de especiação e duplicação. Considerando que todos os genes dentro de um mesmo grupo de ortólogo possuem uma mesma origem ancestral, a delimitação de um grupo deve facilitar a transferência de informações bioquímicas, estruturais e funcionais de uma proteína para outra dentro de um mesmo grupo (Tatusov et al., 1997). Esta abordagem otimizaria a análise funcional de proteínas cujas funções ainda são desconhecidas. Além disso, os grupos de ortólogos facilitam a interpretação de padrões evolutivos de genes homólogos, evidenciando taxas de evolução e de duplicação. O estudo evolutivo de famílias gênicas pode trazer informações relevantes ao entendimento da função de um gene, o que acaba gerando um novo paradigma na biologia, pois ao invés de se estudar um fenótipo e procurar um gene responsável por este fenótipo, é possível eleger genes candidatos e através de técnicas de genética reversa determinar a função associada a este gene (Henikoff et al., 1997).

Uma limitação desta abordagem está relacionada à confiabilidade da organização em grupos de ortólogos, que depende, por parte, da amostragem de genes dos organismos considerados. Para contornar esta situação é possível utilizar os dados de EST (*Expressed Sequence Tag*). Existe uma grande quantidade de EST disponíveis de diversas angiospermas nos bancos de dados (GenBank, National Center for Biotechnology Information, NCBI, http://www.ncbi.nlm.nih.gov/), que correspondem a mais de dois milhões de seqüências, e podem contribuir significativamente para melhorar a resolução dos grupos de ortólogos (Gai *et al.*, 2000; Quackenbush *et al.*, 2000; Quackenbush *et al.*, 2000; Vincentz *et al.*, 2001; Yuan *et al.*, 2001, Lijavetzky *et al.*, 2003). Para aumentar a quantidade de informação que se pode obter a partir de ESTs, elas passam por um processo de montagem, no qual são formadas seqüências maiores e mais informativas a partir da sobreposição de seqüências de EST, gerando possíveis transcritos únicos (Pertea *et al.*, 2003; Rudd, 2003). Este tipo de abordagem vem se mostrando eficiente para estudo de diversos genes, em um esforço de representar todas as seqüências expressas de um organismo.

1.4. Filogenia

Para estabelecer as relações de ortologia/paralogia entre os membros homólogos que compõe uma família multigênica é necessário fazer inferências filogenéticas, na forma de árvore, que reconstruam a árvore mais provável que reflita as relações evolutivas.

Dentre os métodos existentes para uma reconstrução filogenética destacam-se a três métodos diferentes: métodos de parcimônia, métodos de verossimilhança e métodos de distância. Aos métodos de reconstrução estão associados dois processos: a estimativa da topologia e a estimativa do valor do comprimento de cada ramo de uma árvore, a fim de gerar uma árvore que represente uma inferência estatística de reconstruir a árvore verdadeira (Nei & Kumar, 2000).

O método de Máxima Parcimônia (MP; Henning, 1966; Eck & Dayhoff, 1967) baseia-se na mudança do estado dos caracteres (Graur & Li, 2000) e suporta a hipótese de que a via menos complicada é a mais provável, minimizando o número de etapas evolucionárias necessárias para explicar um determinado evento. Este método procura todas as topologias de árvores possíveis, em busca da árvore mínima ótima. Apesar de fazer análises de diferentes topologias, MP é um método que requer um grande tempo de comparação, o tamanho dos ramos não são informativos e se utiliza apenas dos sítios informativos (Nei & Kumar, 2000).

O método de Máxima Verossimilhança (MV; Cavalli-Sforza & Edwards, 1967; Felsenstein, 1981) se baseia em testar hipóteses ou modelos evolutivos e observar se qual topologia obtida melhor adequou ao modelo utilizado. Este método apresenta menor variância do que outros métodos tende a ser robusto ao assumir o modelo evolutivo e tem uma boa base estatística, no entanto, é um método quase proibitivo quando se trata de análise com grandes bancos de dados, considerando seu grande consumo de memória (Nei & Kumar, 2000).

O método de junção de vizinhos mais próximos ou *Neighbor-Joining* (NJ; Saitou & Nei, 1987) é uma simplificação do método de Evolução Mínima (EM), o qual se baseia nas no cálculo das distâncias evolutivas (Graur & Li, 2000) para todos os pares e reconstrução de uma árvore que leva em consideração as relações entre todas as distâncias. No caso de NJ, a metodologia recupera apenas uma topologia, que é assumida como mais provável. NJ é um método rápido, apropriado para grandes conjuntos de dados, permite linhagens com diferentes tamanhos de ramos e substituições múltiplas, no entanto ele mostra apenas uma topologia possível (Nei & Kumar, 2000).

É importante ressaltar que tanto MP quanto MV são considerados métodos cladísticos, ou seja, levam em consideração o estado dos caracteres analisados e

12

estabelecem um relação ancestral-derivado, resultando em uma topologia de árvore denominada clado, na qual se observa um raiz. Por outro lado, NJ é um método fenético, que se baseia no grau de similaridade entre as seqüências. A fenética tem como pressuposto que todas as seqüências analisadas têm uma taxa de evolução igual, e, portanto, não visa estabelecer relações de ancestralidade (Graur & Li, 2000).

A utilização de métodos que requerem uma menor capacidade computacional pode levar a conclusões similares àquelas obtidas por métodos mais dispendiosos de tempo, como MP e MV (Nei *et al.*, 1998), o que levou à utilização de NJ como o método de inferência.

1.5. Objetivos

- Identificar fatores de transcrição do tipo bZIP em seqüências genômicas de arroz e de ESTs de angiospermas para utilizá-los em uma análise filogenética;

Definição de grupos de genes homólogos e grupos de genes ortólogos dos fatores bZIPs
de angiospermas e utilização deste dados para racionalizar estudos funcionais;

 Iniciar um estudo funcional relacionado ao direcionamento nuclear dos bZIPs AtbZIP76 e AtbZIP78.

2. Materiais e Métodos

2.1. Material Computacional

Todas as análises computacionais foram realizadas com um computador Pentium[®] 4, com processador de 1,8 GHz, disco rígido de 80 Gb e memória RAM de 1Gb. Os sistemas operacionais utilizados foram Linux e Windows XP, operando na mesma máquina, através de um sistema de *dual boot*. Todos os pacotes e programas utilizados, com exceção do Office[®] XP foram obtidos gratuitamente (Tabela III). Os *scripts* utilizados para as análises, desde a captura de seqüência até a geração de tabelas, foram desenvolvidos no próprio laboratório ou em colaboração com outros laboratórios (disponíveis em www.unicamp.br/~lgcorrea/scripts).

Tabela III. Lista de programas utilizados.			
Programa	Site	Referência	
Clustal X	www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html	Thompson et al., 1997	
Blast	www.ncbi.nih.gov/blast	Altschul <i>et al</i> ., 1990	
TGICL	www.tigr.org	Pertea <i>et al</i> ., 2003	
Phylip	http://evolution.genetics.washington.edu/phylip.html	Felsenstein J., 1993	
GeneRunner	http://www.generunner.com/	-	
MEME	http://meme.sdsc.edu/meme/website/intro.html	Bailey & Elkan, 1994	

2.2. Obtenção das seqüências

Os genomas de *Oryza sativa* (arroz) correspondentes a ssp *indica* (103044 seqüências, Yu *et al.*, 2002) e ssp *japonica* (42109 seqüências, Goff *et al.*, 2002) foram obtidos através do Instituto de Genômica de Pequim (http://btn.genomics.org.cn/rice/) e da Syngenta (www.syngenta.com), respectivamente. Os 63 cDNAs de arroz que codificam para bZIP foram selecionadas através do Consórcio de cDNA inteiro de arroz

(www.gramene.org; The Rice Full-Length cDNA Consortium, 2003). As seqüências dos 12 cromossomos de *Oryza sativa*. ssp *japonica* foram obtidas através do Consórcio Internacional de Seqüenciamento do Genoma do Arroz (IRGSP; www.gramen.org; Sasaki & Burr, 2000).

Os ESTs de monocotiledôneas Oryza sativa (37109 següências), Sorghum bicolor (84712 sequências), Zea mays (207420 sequências), Hordeum vulgare (372383 e Triticum aestivum (415728 seqüências) e dicotiledôneas Glvcine max seqüências) (308572 seqüências), Medicago truncatula (181440 seqüências), Lycopersicon esculentum (148650 seqüências), Solanum tuberosum (94422 seqüências) e Lotus corniculatus var. *japonicus* (36311 seqüências) foram obtidos através do banco de dados de EST (dbEST) do NCBI (http://www.ncbi.nih.gov/dbEST/index.html) em 22/04/2003, e no caso de arroz elas foram obtidas também através do Instituto de Genômica de Pequim. Os ESTs de Saccharum SSD. (237954 seqüências) foram obtidos através do SUCEST (http://sucest.lad.dcc.unicamp.br/en/). Os ESTs relativos à alga verde Chlamydomonas reinhardtii (27511 seqüências), à briófita Marchantia polimorfa (1415) e à gimnosperma Pinus taeda (110622 seqüências) forma obtidos através do NCBI em 26/01/2004. Os ESTs relativos à briófita *Physcomitrella patens* foram cedidos pela Universidade de Leeds (Inglaterra, por comunicação pessoal com Martin Lomas).

2.3. Construção de um banco não redundante de fatores de transcrição do tipo bZIP de Angiospermas.

A construção de um banco não redundante de fatores de transcrição do tipo bZIP de angiospermas (Angiotot) foi realizada através de buscas interativas no banco de GenBank (NCBI. www.ncbi.nih.gov) proteínas do e do **MAtDB** (www.mips.biochem.mpg.de/proj/thal/) utilizando seqüências de bZIPs conhecidas e evolutivamente distantes umas das outras como següências query para buscas nos bancos de dados com os programas blastp e tblastn, nos servidores do NCBI e MAtDB. Buscas adicionais por palavra chave também foram feitas no MAtDB e no NCBI. Edição de exons que codificam alguns domínios bZIP envolveu modificações nas junções intron/exon, que foram guiadas pelo alinhamento de seqüências de aminoácidos e pela presença de sítios canônicos de splicing (GT-AG) (Vincentz et al, 2003). A versão utilizada para as análises corresponde a uma atualização realizada em outubro/2003.

2.4. Montagem das seqüências de EST em clusters

As montagens dos ESTs em *clusters* foi feita através do programa TGICL (Pertea et al., 2003). *Cluster* é o resultado do agrupamento de seqüências feito por um programa de montagem, e é formado por *contig* (quando o *cluster* é formado por duas ou mais seqüências) e *singlet* ou *singleton* (quando o *cluster* é formado apenas por uma seqüência).

Para as montagens dos ESTs de cevada, soja e alfafa, foi necessário um passo anterior à montagem por TGICL, devido à formação de *clusters* numerosos e presença de seqüências de baixa complexidade. Para solucionar este problema, duas etapas anteriores à montagem foram feitas. A primeira consistiu em selecionar as ESTs que possuem similaridade com as bZIPs do banco Angiotot, através de Blastx e utilizando um valor de corte de 1e-4, resultando em um subconjunto de seqüências que possuem similaridade com bZIPs. A segunda corresponde a um novo ciclo de seleção, utilizando o subconjunto obtido na primeira etapa como semente para selecionar as ESTs que possuem similaridade com este subconjunto. Esta segunda etapa foi feita por Blastn com valor de corte 1e-4, e as ESTs selecionadas nesta etapas foram então montadas em *clusters*.

2.5. Estabelecimento da metodologia para identificar seqüências genômicas de arroz e *clusters* de ESTs que possivelmente codifiquem para fatores de transcrição do tipo bZIP

A identificação de fatores de transcrição do tipo bZIPs em seqüências genômicas de *Oryza sativa* (arroz) e em *clusters* de ESTs foi realizada utilizando o programa tBlastn e o conjunto de referência de bZIPs não redundantes em angiospermas (Angiotot), descrito em Vincentz *et al.* (2003). As seqüências que possuíam um *e-value* inferior a 1e-4 foram selecionadas, formando um subconjunto (SeqZIP, Figura 4).

Os possíveis falsos positivos, correspondentes principalmente a regiões de baixa complexidade, foram excluídos deste subconjunto, através de conferência visual dos alinhamentos.

Um análise mais detalhada das demais seqüências de SeqZIP foi feita através da comparação por Blastx 2a2 (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html) com o respectivo melhor alinhamento (*best hit*) em Angiotot, a fim de definir sua fase de leitura

17

aberta, e, no caso das seqüências genômicas de arroz, definir a estrutura do gene. A definição desta estrutura tem como base os alinhamentos obtidos por Blastx 2a2, a posição de introns conservados em bZIPs homólogas e a presença de sítios canônicos de *splicing* GT-AG. Para auxiliar a montagem das seqüências codificantes foi utilizado o programa GeneRunner. O Protocolo estabelecido está descrito na Figura 4.

2.6. Análise de redundância

A redundância existente entre os conjuntos de bZIP de *Oryza sativa* ssp *indica* e *Oryza sativa* ssp *japonica* foi avaliada usando Blastp entre os dois conjuntos de bZIP, fixando como critério de identidade 97% de aminoácidos idênticos sobre um alinhamento de no mínimo 40 aminoácidos em região conservada. Este critério se baseia na observação de 96.5% de identidade observada entre dois parálogos recentes em milho, OHP1 e OHP1b, e, portanto apenas os genes que possuem uma identidade menor que 97% são considerados novos genes (Vincentz *et al.*, 2003).

2.7. Análise filogenética

As seqüências de aminoácidos das bZIPs foram alinhadas com Clustal X (Thompson *et al.*, 1997), utilizando as condições padrão do programa. Análise das distâncias foi realizada por *neighbor-joining* (NJ) com o programa NEIGHBOR (PHYLIP). As distâncias foram obtidas usando uma matriz de distâncias do tipo PAM (Dayhoff *et al.*, 1978), através do programa PRODIST (PHYLIP). As re-amostragens do conjunto original de bZIPs foi feita por um *bootstrap* de 1000 repetições através do programa SEQBOOT (PHYLIP).



Figura 5. Esquema do protocolo bioinformático para a identificação de bZIPs em seqüências genômicas e EST. As seqüências de genomas de arroz (ssp. *indica* e ssp. *japonica*) ou *clusters* de ESTs são comparadas ao banco Angiotot por tBlastn, originando um conjunto de seqüências que possivelmente codifiquem bZIP (SeqZIP). Deste conjunto, uma análise visual permitiu retirar os falsos positivos, que correspondem, na maioria das vezes, às seqüências de baixa complexidade. As demais seqüências são analisadas por Blastx 2x2 com o melhor alinhamento obtido com um bZIP de Angiotot, a fim de identificar seu provável quadro aberto de leitura. Além disto, no caso das seqüências genômicas de arroz, foi definida a estrutura dos genes, baseada no alinhamento obtido pelo Blastx 2x2, e considerando a presença de seqüências canônicas GT-AG na junção intronexon e a posição de introns conservados nos seus prováveis pares homólogos.

2.8. Identificação de Possíveis Grupos de Ortólogos (PoGO)

A identificação de PoGOs de bZIPs em mono e dicotiledôneas foi realizada à partir da avaliação das relações evolutivas estabelecidas por NJ dos alinhamentos de seqüências em aminoácidos de motivos conservados de bZIPs em *Arabidopsis*, em arroz, no banco Angiotot e nas bZIPs identificadas a partir de ESTs. A identificação de um PoGO obedece a 3 premissas: (i) cada PoGO deve possuir um *bootstrap* maior que 50%; (ii) o PoGO deve possuir no mínimo um representante de *Arabidopsis* e um de arroz, assumindo que o possível conjunto completo de bZIPs para estes dois organismos foi identificado e que nenhuma perda seletiva de genes ocorreu; e (iii) a filogenia obtida deve refletir a filogenia das espécies (Vincentz *et al.*, 2003).

A identificação dos motivos conservados foi feita através de análises de motivos com o programa MEME e de alinhamentos por ClustalX.

2.9. Análise de duplicações no genoma de Arabidopsis thaliana

Para identificar as bZIPs parálogas originadas por duplicações de blocos do genoma de *Arabidopsis*, foi feita uma busca através do site *Paralogons in Arabidopsis thaliana* (http://wolfe.gen.tcd.ie/athal/dup).

2.10. Obtenção de fusões traducionais de AtbZIP76 e AtbZIP78 com o gene repórter RFP

Os cDNAs de AtBZIP76 (N.º acesso NCBI AV566454) e AtBZIP78 (N.º acesso NCBI AV440431) foram obtidos através do Kazuza DNA Research Institute (www.kazuza.or.jp) e do Riken Institute of Physical and Chemical Research (www.riken.go.jp), respectivamente. O cDNA de AtBZIP76 possui 1125 pb e codifica para uma proteína de 374 aminoácidos. Já o cDNA completo de AtBZIP78 possui 798 pb e codifica para uma proteína de 265 aminoácidos.

Procedimentos básicos foram usados para amplificações, digestões enzimáticas, e clonagens de DNA; transformação de *E. coli* e minipreparações de plasmídeos (Sambrook *et al.*, 1989). As purificações de fragmentos de DNA foram realizadas em gel de agarose 1% utilizando o kit Concert Rapid Gel Extraction System (GibcoBRL) seguindo o protocolo do fabricante. O seqüenciamento do DNA foi realizado segundo a metodologia Dye Terminator (Big Dye Kit, Perkin Elmer) e analisado com os programas Chromas e BLAST.

As seqüências completas codificantes para esses dois fatores de regulação da transcrição de Arabidopsis foram amplificadas por PCR com os oligonucleotídeos LGC3 (5')/LGC4 (3') para AtBZIP76 e LGC1 (5')/LGC2 (3') para AtBZIP78 (Tabela IV), que adicionam um sítio SpeI na extremidade N-terminal, antes do códon de iniciação e um sítio para KpnI antes do códon de terminação, de forma que os fragmentos resultantes pudessem ser clonados em fase com RFP na porção N-terminal.

Tabela IV. Oligonucleotídeos utilizados na clonagem de AtbZIP76 e AtbZIP78		
AtBZIP76	AtBZIP78	
Para 5' LGC3	Para 5' LGC1	
5'GA <u>ACTAGT</u> GTGACAATGGC 3'	5' TT <u>ACTAGT</u> AATATAGAGATGGC 3'	
Sítio SpeI	Sítio Spel	
Para 3' LGC4	Para 3' LGC2	
5'AGT <u>GGTACC</u> GACCCATCACAGATAC 3'	5' GAA <u>GGTACC</u> TTTCCAGTTCAGCTA 3'	
Sítio Kpn	Sítio Kpn	

Um fragmento de 798 pb foi obtido para AtBZIP78 e um de 1134 pb para AtBZIP76. Estes fragmentos foram clonados em pOL-RFP (Figura 5). O vetor pOL-RFP se origina de vetor pOL-GFP (Peetrers *et al.*, 2000), cuja seqüência de GFP foi excisada e trocada pela seqüência de RFP originada de pDsRed1-N1 (Clontech).



Figura 6: Esquemas dos cassetes de expressão dos plasmídeos Pol RFP:::AtbZIP76 e Pol RFP:::AtbZIP78 utilizados nos ensaios de expressão transiente em epiderme de cebola. A) Promotor 35S comandando a expressão da proteína AtbZIP76 fusionada a RFP. B) Promotor 35S comandando a expressão da proteína AtbZIP78 fusionada a RFP.
2.11. Transformação por biobalística

Tiras de epiderme de cebola de aproximadamente 2x2 cm, previamente dispostas em placas de Petri contendo sal Murasshige & Skoog (MS, 1962) 1X, 30g/L de sacarose, 2% de agar e pH 5,7, foram bombardeadas utilizando-se um canhão de hélio construído em série pelo CENARGEN/Embrapa de Brasília. Aproximadamente 5µg de DNA plasmidial de cada construção foram precipitados sobre micropartículas de ouro (250µg/µl; 50µl), previamente lavadas com etanol 70%, centrifugadas a 14.000 rpm por 5 min., lavadas com água milli-Q estéril e ressuspendidas em 60µl de glicerol 50%. Nessa preparação, a mistura de DNA e micropartícula foi homogeneizada em vortex por 5 min, adicionando-se simultaneamente, em seguida, 20 μ L de espermidina 0,1 M, e 50 μ L de cloreto de cálcio 2,5 M. Essa mistura continuou sendo homogeneizada em vortex por mais 3 min e permaneceu em repouso por mais 3 min. Em seguida, o material foi centrifugado por 5 s a 5.000 rpm, o sobrenadante foi descartado e o sedimento lavado uma vez com 150µL de etanol 70% e 3x com etanol absoluto. As micropartículas recobertas com o DNA de interesse foram ressuspendidas em 60µL de etanol absoluto e 7µL dessa preparação foi depositado sobre discos "kapiton", que ficaram em repouso, secando em fluxo laminar. No bombardeamento foram empregadas uma pressão de 1200 psi de gás hélio e vácuo de 28 mm de Hg. As tiras de cebola foram posicionadas dentro da câmera de vácuo do canhão a uma distância de 6 cm do disco de ruptura. Após o bombardeamento, células transformadas foram incubadas a 22°C por um período de no mínimo 24 horas.

Os controles pRJG23 e pRJG32, com seqüências fusionadas a GFP que direcionam para núcleo e citoplasma, respectivamente, foram gentilmente cedidos por Robert J. Grebenok (Departamento de Biologia, Canisius College, Nova Iorque, EUA).

2.12. Microscopia de Fluorescência

O material transformado foi observado em microscópio de fluorescência Nikon Eclipse E600, com câmara CCD: Cool SNAP-Pro digital system from Media Cybernetics (Silver Spring, Maryland, USA). Os comprimentos de onda dos filtros utilizados para excitação foram 330-380 nm para a visualização de DAPI/Hoechst, 465-495nm para GFP e 528-553 nm para RFP.

3. Resultados e Discussão

Projetos genomas de diferentes organismos visam elucidar a arquitetura dos diferentes genomas e obter informações a respeito de seus genes (Jacq, C. *et al.* 1997; The *C. elegans* Sequencing Consortium, 1998; Adams at al., 2000; *Arabidopsis* Genome Initiative, 2000; Goff at al., 2002; Yu at al., 2002). Apesar de o número de seqüências depositadas em bancos de dados crescer exponencialmente, os esforços para caracterizar os genes codificados por estes genomas não conseguem acompanhar o ritmo deste crescimento. Do genoma de *Arabidopsis*, apenas cerca de 10 a 20% dos seus 28 mil genes foram funcionalmente caracterizados, mas existem esforços internacionais que visam caracterizar todo o transcriptoma de *Arabidopsis* até 2010 (Riechmann & Ratcliffe, 2000; The Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project, 2003).

A comparação dos genes de diferentes genomas permite uma classificação em famílias de genes homólogos, as quais incluem ortólogos e parálogos (Tatusov *et al.*, 1997). Um Grupo de Genes Ortólogos assume que os genes deste grupo evoluíram de um ancestral único a partir de eventos de especiação e duplicação. Como genes ortólogos são considerados representantes de uma mesma função ancestral, a organização em grupos de ortólogos deve permitir racionalizar estudos funcionais de genes em diversas linhagens evolutivas (Tatusov *et al.*, 1997; Matinez-Castilla & Alvarez-Buylla, 2003; Nam *et al.*, 2004).

Os genomas de arroz (Goff at al., 2002; Yu at al., 2002) e *Arabidopsis (Arabidopsis* Genome Iniciative, 2000), representantes das duas sub-classes de angiospermas, mono e eudicotiledôneas respectivamente, representam uma oportunidade única para o estudo exaustivo da evolução de família multigênica e suas funções associadas em angiospermas.

25

Dentro deste contexto, iniciamos uma análise filogenética detalhada dos fatores de transcrição do tipo bZIP em angiospermas, a fim de classificá-los em possíveis grupos de genes ortólogos (PoGO). Um conjunto completo e não redundante de 77 bZIPs foi identificado no genoma de *Arabidopsis* (Jacoby *et al.*, 2002; Vincentz *et al.*, 2003). Entretanto, o conjunto completo de bZIPs de arroz ainda não havia sido identificado ao início deste projeto, portanto, o objetivo inicial deste trabalho foi identificar os fatores bZIP deste organismo.

3.1. Identificação de genes que codificam para fatores de transcrição do tipo bZIP em arroz

As seqüências genômicas das duas subespécies de arroz (ssp *indica* e ssp *japonica*), que possivelmente codificam para fatores de transcrição do tipo bZIP, foram selecionadas pelo protocolo computacional descrito na Figura 5. Este protocolo envolve um comparação das seqüências genômicas de arroz com o banco de referência de bZIP de angiospermas (Angiotot), incluindo o conjunto completo de 77 bZIPs de *Arabidopsis* (Vincentz *et al.*, 2003).

Verificamos que o protocolo bioinformático desenvolvido consegue recuperar os 15 genes de arroz que codificam para bZIP descritos na literatura, o que sugere que o protocolo é eficiente e confiável.

Este protocolo permitiu identificar 72 genes bZIPs no genoma de *O. sativa* ssp *indica* e 73, no genoma de *O. sativa* ssp *japonica*. A redundância entre bZIPs provenientes dos genomas de *Oryza sativa* ssp. *indica* e ssp. *japonica* foi avaliada por uma comparação

26

dos dados obtidos para cada subespécie usando o programa Blastp, considerando que duas seqüências como representando um mesmo *locus* gênico quando apresentar uma identidade igual ou superior a 97% sobre o alinhamento total (Vincentz *et al.*, 2003). Desta maneira, identificamos um conjunto único e não redundante de 86 fatores bZIP de arroz, dos quais 59 são comuns às duas subespécies, um conjunto de 13 bZIPs exclusivos de ssp. *indica* e 14 exclusivos de ssp. *japonica*.

As diferenças observadas entre as subespécies *indica* e *japonica* podem estar relacionadas a dois fatores: (i) perda ou ganho de genes em uma das linhagens, uma vez que as subespécies divergiram há aproximadamente um milhão de anos atrás (Khush, 1997; Bennetzen, 2000; Song et al., 2002); (ii) parcela de genoma não seqüenciado ou não montado em cada uma das subespécies (Lijavetzky et al., 2003).

Com o objetivo de descobrir possíveis bZIPs de arroz, ainda não representados nas seqüências genômicas e também aumentar a informação de seqüências de aminoácidos dos bZIPs identificados a partir das seqüências genômicas, foram analisados os *clusters* de ESTs de arroz (<u>item 2.3</u> de Materiais e Métodos), as seqüências cromossômicas de *O. sativa* ssp. *japonica* originadas do Consórcio Internacional (IRGSP) e as seqüências de cDNAs inteiros (The Rice Full-Length cDNA Consortium, 2003), através do protocolo estabelecido (Figura 5). Desta análise foram identificados sete novos bZIPs a partir de *clusters* de ESTs, dez a partir das seqüências cromossômicas e dez a partir dos cDNAs. Além disto, foi possível aumentar a informação de seqüência de aminoácido para 16 bZIPs de arroz.

Concluindo, o conjunto destas buscas resultou na identificação de 113 bZIPs em arroz, que possivelmente representam o conjunto completo e não redundante de genes que codificam para bZIPs neste organismo (Tabela V). Destas 113 bZIPs, 98 são novas, e 15 já descritas previamente em bancos de dados. Considerando número de bZIPs encontrados em arroz é aproximadamente 1,47 vezes o número de bZIPs em *Arabidopsis*, podemos fazer uma estimativa do número de genes em arroz como sendo 1,47 maior que o número de genes em *Arabidopsis*, ou seja, aproximadamente 41.000 genes. Este número encontrado é compatível com o número de genes estimados para arroz.

Tabela V. Distribuição das bZIPs entre as diferentes fontes de seqüências de arroz.						
Fonte de seqüências	DNA genômico (ssp <i>indica</i> e <i>japonica</i>)	EST	cDNAs inteiros	IRGSP*	Total	
n ^o de bZIPs 86 7 10 10 113						
*International Rice Genome Sequencing Project						

Finalmente, pode-se mencionar que seis seqüências, que possuíam similaridade com bZIPs, apresentaram códon de término de tradução em sua fase de leitura, constituindo em possíveis pseudogenes.

Além das seqüências de arroz e de *Arabidopsis*, uma análise em bancos de dados de ESTs foi feita a fim de identificar bZIPs em outras angiospermas, para aumentar a resolução das análises evolutivas.

3.2. Identificação de bZIPs em bancos de ESTs

Bancos de dados de ESTs constituem-se em uma importante fonte de dados para a prospecção de genes de interesse. Para identificar seqüências que possivelmente

codifiquem para bZIP em ESTs de várias monocotiledôneas (milho, cevada, trigo, sorgo e cana) e eudicotiledôneas (soja, alfafa, tomate, batata e lótus), ESTs obtidos foram montados em *clusters* (itens 2.2 e 2.3 de Matérias e Métodos), os quais foram analisados através do protocolo estabelecido para análise das seqüências genômicas (Figura 5). O banco Angiotot, acrescido dos bZIPs identificadas no genoma de arroz neste trabalho, foi utilizado nesta busca, o que resultou na identificação de 226 bZIPs de eudicotiledôneas e 367 de monocotiledôneas (Tabela VI).

Tabela VI. Número de bZIPs identificados a partir de <i>clusters</i> de EST					
Organismo	No. de bZIPs encontradas				
Eudicotiledôneas	Eudicotiledôneas				
Alfafa	53				
Batata	35				
Lótus	22				
Soja	69				
Tomate	47				
Total	226				
Monocotiledôneas					
Cana	102				
Cevada	78				
Milho	50				
Sorgo	58				
Trigo	79				
Total	367				

É importante salientar que o número de *clusters* de EST não é representativo do número real de bZIPs de um organismo, uma vez que: (i) *clusters* diferentes podem representar um mesmo gene, na medida em que codificam domínios/motivos conservados que não se sobrepõem; (ii) possuem domínios em comum com outros tipos de proteínas.

3.3. Análise filogenética das seqüências de bZIP de arroz e *A. thaliana* para a identificação de grupos de genes homólogos.

Para definir grupos de bZIPs homólogos, foi realizada uma análise filogenética do conjunto não redundante de 172 bZIPs, sendo 96 de arroz e 76 de *Arabidopsis*. A lista completa dos bZIPs de arroz e de *Arabidopsis* utilizados na análise filogenética pode ser encontrada na Figura 7. Dos 113 bZIPs de arroz, 17 não foram utilizadas para análise filogenética, uma vez que não apresentam o motivo bZIP mínimo (Tabela VII). AtbZIP78 não fez parte desta análise por possuir apenas 2 repetições de leucina, nem AtbZIP73, por ser pseudogene.

Tabela VII. bZIPs de arroz que não foram utilizados na análise filogenética.						
Grupo	А	В	D	G	Н	S
Gene	OsbZIP97	OsbZIP98	OsbZIP101	ObZIP105	OsbZIP107	OsbZIP108
		OsbZIP99	OsbZIP102	ObZIP106		OsbZIP109
		OsbZIP100	OsbZIP103			OsbZIP110
			OsbZIP104			OsbZIP111
						OsbZIP112
						OsbZIP113
Grupo refere-se à classificação em grupos de genes homólogos identificados a partir das						
análises filogenéticas (Figuras 7 e 9). Os bZIPs foram classificados nos vários Grupos com						
Dase em seu memor alimnamento por Diastp.						

Todos os fatores de transcrição do tipo bZIP apresentam similaridade apenas no motivo bZIP mínimo, que engloba o motivo básico de ligação ao DNA mais 3 repetições de leucina, perfazendo 43 aminoácidos (Vettore *et al.*,1998). Uma análise deste motivo mínimo do conjunto de 172 bZIPs por *neighbor-joining* permitiu encontrar um *bootstrap* significativo que identifica duas famílias de genes evolutivamente relacionadas. Estes dois agrupamentos foram definidos como Grupo F (98% de *bootstrap*) e Grupo D (97% de *bootstrap*) (Figura 8). A topologia da árvore obtida e o *bootstrap* de 56% indicam que os Grupos F e D são grupos irmãos, isto é, podem compartilhar um ancestral comum (Figura

bZIPno.	Codigo	Publicado	GenBank	
OsbZIP23				
OsbZIP24 OsbZIP26				
OsbZIP29				PoGO A1
AtbZIP12	At2g41070	DPBF4	AF334209	
OsbZIP66	At3g56850	AREB3	AB017162	
OsbZIP34				
OsbZIP37	444-25000		DN000001	PoGO A2
AtbZIP14 AtbZIP27	At2g17770		BN000021 BN000022	
OsbZIP25				
OsbZIP33	At5a44080	OSE2	AAF65459	PoGO A3
AtbZIP40	At1g03970	GBF4	U01823	
OsbZIP35				
OsbZIP38 Ath7IP39	At2a36270	ABI5	AE334206	Pogo A4
AtbZIP67	At3g44460	DPBF2	AJ419600	
OsbZIP27				
OsbZIP26 OsbZIP31				
OsbZIP32		TRAB1	BAA83740	
OsbZIP36				PoGO A5
AtbZIP 15	At1g49720	ABF1	AF093544	
AtbZIP36		ABF2	AF093545	
AtbZIP37	At4g34000	ABF3	AF093546	
	Al3919290	ADF4	AF093047	
OsbZIP78				
AtbZIP17	At2g40950		AV441374	PoGO B1
	At3g10800		AJ419850	
	Algoudu	REB	RAA11431	
OsbZIP63		RISBZ1	BAB39173	PoGO C1
AtbZIP63	At5g28770	BZO2H3		
AtbZIP62	At4q02640	BZO2H1		PoGO C2
AtbZIP25	At3g54620	BZO2H4		
OsbZIP58		RITA1	AAC37418	Decc 02
OsbZIP59 OsbZIP60		RISBZ4	BAB39174 BAB39175	POGO C3
Ath7ID0	At5a24800	DZODUD	AE210222	
	Alog24000	BZUZHZ	AF310223	
OsbZIP6	At0y24000	BZUZHZ	AF310223	D-00 D4
OsbZIP6 OsbZIP10 OsbZIP21	AU924000	BZUZHZ	AF310223	PoGO D1
OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65	At5g06839	BZUZHZ	AJ314787	PoGO D1
OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP8	At5g06839	BZUZHZ	AJ314787	PoGO D1
OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP8 OsbZIP11 OsbZIP16	At5g06839	BZUZHZ	AJ314787	PoGO D1
OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP8 OsbZIP11 OsbZIP16 OsbZIP17	At5g06839	BZUZHZ	AJ314787	PoGO D1
OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP8 OsbZIP11 OsbZIP16 OsbZIP17 AtbZIP21	At5g06839	520282	AJ314787 AJ314757	PoGO D1 PoGO D2
OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP8 OsbZIP11 OsbZIP16 OsbZIP17 AtbZIP21 OsbZIP7 OsbZIP75	At5g06839 At1g08320	620202	AJ314787 AJ314757	PoGO D1 PoGO D2
OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP8 OsbZIP11 OsbZIP16 OsbZIP17 AtbZIP21 OsbZIP7 OsbZIP70 OsbZIP20	At5g06839 At1g08320	520282	AJ314787 AJ314757	PoGO D1 PoGO D2
OsbZIP6 OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP8 OsbZIP10 OsbZIP17 OsbZIP17 OsbZIP7 OsbZIP20 AtbZIP22 AtbZIP22	At5g06839 At1g08320 At1g22070	TGA3	AJ314787 AJ314757 L10209	PoGO D1 PoGO D2 PoGO D3
OsbZIP6 OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP8 OsbZIP10 OsbZIP11 OsbZIP16 OsbZIP17 OsbZIP7 OsbZIP20 AtbZIP22 AtbZIP22 AtbZIP27 AtbZIP27	At5g06839 At1g08320 At1g22070 At5g65210 At1g7920	TGA3 TGA1	AJ314787 AJ314757 L10209 X68053 AJ315736	PoGO D1 PoGO D2 PoGO D3
OsbZIP6 OsbZIP10 OsbZIP10 OsbZIP11 AtbZIP65 OsbZIP11 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP15 OsbZIP20 AtbZIP22 AtbZIP27 AtbZIP50	At5g06839 At1g08320 At1g22070 At5g65210 At1g77920 At5g10030	TGA3 TGA1 OBF4	AJ314787 AJ314757 L10209 X68053 AJ315736 X69899	PoGO D1 PoGO D2 PoGO D3
OsbZIP6 OsbZIP10 OsbZIP10 OsbZIP10 OsbZIP11 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP20 AtbZIP22 AtbZIP20 AtbZIP20 AtbZIP57 OsbZIP13 AtbZIP57	Attg06839 Attg08320 Attg22070 Attg22070 Attg22070 Attg22070 Attg22070 Attg22070 Attg22070 Attg22070 Attg22003	TGA3 TGA1 OBF4 PAN	AJ314787 AJ314757 L10209 X68053 AJ315736 X69899 AF111711	PoGO D1 PoGO D2 PoGO D3 PoGO D4
Cobb21P6 Osb21P61 Osb21P10 Osb21P21 Atb21P65 Osb21P11 Osb21P17 Osb21P17 Osb21P15 Osb21P15 Osb21P15 Osb21P15 Osb21P15 Osb21P15 Osb21P10 Atb21P40 Atb21P40 Atb21P40 Osb21P15 Osb21P16 Osb21P20 Atb21P46 Osb21P16 Osb21P20	Attg06839 Attg08320 Attg22070 Attg65210 Attg77920 Attg10303 Attg68640	TGA3 TGA1 OBF4 PAN	AJ314787 AJ314757 L10209 X68053 AJ315736 X69899 AF111711	PoGO D1 PoGO D2 PoGO D3 PoGO D4
Abb/19 OsbZIPG OsbZIP10 OsbZIP21 AtbZIP65 OsbZIP10 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP15 OsbZIP15 OsbZIP16 AtbZIP22 AtbZIP47 AtbZIP57 OsbZIP13 AtbZIP46 OsbZIP13 OsbZIP19 OsbZIP19	Attg06839 Attg08320 Attg22070 At5g65210 Attg77920 At5g10030 Attg68640	TGA3 TGA1 OBF4 PAN	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711	PoGO D1 PoGO D2 PoGO D3 PoGO D4
Cob21P6 Osb21P61 Osb21P21 Ab2/P62 Osb21P21 Osb21P16 Osb21P16 Osb21P17 Osb21P17 Osb21P17 Osb21P12 Atb21P27 Atb21P27 Osb21P17 Osb21P13 Atb21P46 Osb21P14 Osb21P14	Attg06839 Attg08320 Attg22070 Attg65210 Attg77920 Attg10030 Attg68640	TGA3 TGA1 OBF4 PAN	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711	PoGO D1 PoGO D2 PoGO D3 PoGO D4
CobJ21P6 OsbZ1P6 OsbZ1P10 OsbZ1P21 AtbZIP65 OsbZ1P8 OsbZ1P16 OsbZ1P17 OsbZ1P17 OsbZ1P17 OsbZ1P17 OsbZ1P10 OsbZ1P19 OsbZ1P3 AtbZ1P46 OsbZ1P19 OsbZ1P14 OsbZ1P18 OsbZ1P18	Attg06839 Attg08320 Attg22070 Attg22070 Attg65210 Attg77920 Attg68640	TGA3 TGA1 OBF4 PAN	AJ314787 AJ314787 L10209 L10209 L10209 A68053 AJ315736 X68059 AF111711 BAB44077	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5
Abb.10 - S OsbZ1P6 OsbZ1P1 OsbZ1P21 AtbZ1P65 OsbZ1P8 OsbZ1P10 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P12 AtbZ1P21 OsbZ1P13 AtbZ1P24 AtbZ1P25 OsbZ1P10 OsbZ1P12 AtbZ1P57 OsbZ1P19 OsbZ1P12 OsbZ1P12 OsbZ1P19 OsbZ1P11 OsbZ1P12 OsbZ1P18 OsbZ1P19 OsbZ1P19 OsbZ1P18 OsbZ1P19 OsbZ1P19 OsbZ1P19 OsbZ1P19 OsbZ1P19 OsbZ1P12	Attg06839 Attg08320 Attg22070 Attg22070 Attg22070 Attg22070 Attg20030 Attg20030 Attg20030	TGA3 TGA1 OBF4 PAN	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711 BAB44077	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5
Aub.1 9 OsbZIP6 OsbZIP10 OsbZIP11 OsbZIP12 OsbZIP13 OsbZIP11 OsbZIP11 OsbZIP12 OsbZIP11 OsbZIP11 OsbZIP15 OsbZIP15 OsbZIP17 OsbZIP15 OsbZIP17 OsbZIP17 OsbZIP13 AtbZIP47 OsbZIP13 AtbZIP46 OsbZIP14 OsbZIP18 OsbZIP19 OsbZIP192 AtbZIP20	Attg06839 Attg06839 Attg08320 Attg22070 At5g65210 Attg77920 At5g10030 Attg68640 Att5g06950 At5g06950	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711 BAB44077 D10042 X66900	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5
Abble 3 OsbZ1P6 OsbZ1P61 OsbZ1P10 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P15 OsbZ1P15 OsbZ1P15 OsbZ1P15 OsbZ1P17 OsbZ1P15 OsbZ1P17 OsbZ1P17 OsbZ1P18 OsbZ1P17 OsbZ1P18 OsbZ1P19 OsbZ1P19 OsbZ1P18 OsbZ1P19 OsbZ1P19 OsbZ1P19 OsbZ1P192 AtbZ1P27 AtbZ1P37 OsbZ1P18 OsbZ1P19 OsbZ1P19 OsbZ1P12 AtbZ1P26 AtbZ1P26	Attg06839 Attg08320 Attg22070 At5g65210 At5g65210 At5g65210 Attg77920 At5g10030 Attg68640 At5g06950 At5g06950 At5g06950 At5g06950	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b	AJ314787 AJ314787 L10209 X68053 AJ315736 AS69909 AJ320540	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5
Abble 3 OsbZ1P6 OsbZ1P21 OsbZ1P21 OsbZ1P21 OsbZ1P10 OsbZ1P11 OsbZ1P10 OsbZ1P11 OsbZ1P11 OsbZ1P12 OsbZ1P15 OsbZ1P15 OsbZ1P10 OsbZ1P10 OsbZ1P12 AtbZ1P27 AtbZ1P29 OsbZ1P13 AtbZ1P47 OsbZ1P13 OsbZ1P14 OsbZ1P12 OsbZ1P13 OsbZ1P14 OsbZ1P19 OsbZ1P12 AtbZ1P26 AtbZ1P27 OsbZ1P12 OsbZ1P12 OsbZ1P12 OsbZ1P12 AtbZ1P26 AtbZ1P26 OsbZ1P22 AtbZ1P26 OsbZ1P20 OsbZ1P20 OsbZ1P20 OsbZ1P20 OsbZ1P20 OsbZ1P20 OsbZ1P20 OsbZ1P20	Attg06839 Attg08320 Attg22070 At5g65210 At5g65210 Attg77920 At5g10030 Attg68640 At5g06950 At5g06950 At5g06950 At5g06950	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA6	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711 BAB44077 D10042 X69900 AJ320540	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5
Abble 3 OsbZ1P6 OsbZ1P10 OsbZ1P21 Abb2P65 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P15 OsbZ1P16 OsbZ1P17 Abb2P47 OsbZ1P15 OsbZ1P17 Atb2IP47 Atb2IP47 OsbZ1P13 OsbZ1P14 OsbZ1P18 OsbZ1P18 OsbZ1P18 OsbZ1P18 OsbZ1P18 OsbZ1P18 OsbZ1P18 OsbZ1P18 OsbZ1P18 OsbZ1P19 OsbZ1P18 OsbZ1P28 Atb21P45 OsbZ1P82 OsbZ1P82 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 Os	Attg06839 Attg08320 Attg22070 At5g65210 Attg77920 At5g10030 Attg10030 Attg68640 Att5g06950 At5g06950 At5g06950 At5g06950 Attsg06950	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA2	AJ314787 AJ314787 L10209 X68053 AJ315736 X6899 AF111711 BAB44077 D10042 X69900 AJ320540 AC079604	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5
Abble 3 OsbZ1P6 OsbZ1P10 OsbZ1P21 Atb21P65 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P12 OsbZ1P14 OsbZ1P15 OsbZ1P14 OsbZ1P15 OsbZ1P16 OsbZ1P17 OsbZ1P18 OsbZ1P18 OsbZ1P19 OsbZ1P19 OsbZ1P18 OsbZ1P19 OsbZ1P22 Atb21P26 Atb21P36 OsbZ1P86 Atb21P37	Attg06839 Attg08320 Attg22070 At5g65210 At5g65210 At5g10030 Attg68640 At5g06950 At5g06950 At5g06950 At3g12250	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA2	AJ314787 AJ314787 L10209 L16053 AJ315736 X69899 AF111711 BAB44077 D10042 X69900 AJ320540 AC079604 AC023064	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5
Abble 3 OsbZIP6 OsbZIP6 OsbZIP10 OsbZIP10 OsbZIP11 OsbZIP12 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP12 OsbZIP13 AtbZIP20 OsbZIP13 AtbZIP20 OsbZIP13 AtbZIP40 OsbZIP13 OsbZIP13 AtbZIP40 OsbZIP13 OsbZIP14 OsbZIP15 OsbZIP16 OsbZIP170 OsbZIP183 OsbZIP83 OsbZIP83 OsbZIP83	Attg06839 Attg08320 Attg22070 At5g65210 Attg77920 At5g10030 Attg68640 Attg06950 Attg06950 Attg06950 Attg12250	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA2	AJ314787 AJ314787 L10209 X68053 AJ315736 X68059 AJ315736 X68999 AF111711 BAB44077 D10047 X69900 AJ320540 AC079604 AC079604	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5
Abble 3 OsbZIP6 OsbZIP10 OsbZIP10 OsbZIP11 OsbZIP12 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP12 OsbZIP13 AtbZIP27 OsbZIP13 AtbZIP37 OsbZIP13 AtbZIP46 OsbZIP13 AtbZIP47 OsbZIP13 AtbZIP46 OsbZIP14 OsbZIP18 OsbZIP19 OsbZIP19 OsbZIP18 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP20 AtbZIP26 AtbZIP26 AtbZIP27 OsbZIP19 OsbZIP19 OsbZIP20 OsbZIP20 AtbZIP27 OsbZIP30 OsbZIP30 OsbZIP30 OsbZIP30	Attg06839 Attg08320 Attg22070 At5g65210 Attg22070 At5g65210 Attg68640 Attg068640 Attg068650 Attg068650 Attg068650	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA6	AJ314787 AJ314787 L10209 X68053 AJ315736 X69809 AF111711 BAB44077 D10042 X69900 AJ320540 AC079604 AC023064	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5 PoGO E1 PoGO E1
Abble 3 OsbZ1P6 OsbZ1P10 OsbZ1P10 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P15 OsbZ1P16 OsbZ1P17 OsbZ1P15 OsbZ1P15 OsbZ1P17 OsbZ1P15 OsbZ1P17 OsbZ1P17 OsbZ1P18 OsbZ1P19 OsbZ1P84 OsbZ1P83 OsbZ1P83 OsbZ1P84 OsbZ1P84 OsbZ1P83 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84 OsbZ1P84	Attg06839 Attg06839 Attg08320 Attg22070 At5g65210 Attg265210 Attg77920 Attg10030 Attg68640 Att5g06950 Att5g06950 Att5g06950 Att5g06950 Attg262380	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA6	AJ314787 AJ314787 L10209 X68053 AX315736 X69899 AF111711 BAB44077 D10042 X699004 AC079604 AC079604 AC023064	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5 PoGO E1
Abble 3 OsbZ1P6 OsbZ1P21 OsbZ1P21 OsbZ1P31 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P15 OsbZ1P15 OsbZ1P15 OsbZ1P15 OsbZ1P17 AtbZ1P27 AtbZ1P50 OsbZ1P11 OsbZ1P12 OsbZ1P13 AtbZ1P46 OsbZ1P18 OsbZ1P19 OsbZ1P192 AtbZ1P46 OsbZ1P192 AtbZ1P45 OsbZ1P18 OsbZ1P192 AtbZ1P46 OsbZ1P192 OsbZ1P192 AtbZ1P46 OsbZ1P183 OsbZ1P184 OsbZ1P184 OsbZ1P184 OsbZ1P184 OsbZ1P184 OsbZ1P184 OsbZ1P184 OsbZ1P184 OsbZ1P184 OsbZ1P185 OsbZ1P361	Attg06839 Attg08320 Attg08320 Attg22070 At5g65210 At5g65210 Attg77920 At5g10030 Attg68640 Attg06950 At5g06	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA6	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711 BAB44077 D10042 X69900 AJ320540 AC079604 AC023064 AC023064 AF401299 AF401299	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5 PoGO E1 PoGO E2
Abble 3 OsbZ1P6 OsbZ1P10 OsbZ1P21 Abble 26 OsbZ1P31 OsbZ1P10 OsbZ1P11 OsbZ1P11 OsbZ1P11 OsbZ1P17 OsbZ1P17 OsbZ1P17 OsbZ1P17 OsbZ1P17 OsbZ1P17 OsbZ1P17 OsbZ1P17 OsbZ1P13 OsbZ1P13 OsbZ1P14 OsbZ1P13 OsbZ1P13 OsbZ1P13 OsbZ1P14 OsbZ1P15 OsbZ1P15 OsbZ1P16 OsbZ1P17 OsbZ1P18 OsbZ1P18 OsbZ1P18 OsbZ1P84 OsbZ1P84 OsbZ1P85 AtbZ1P47 OsbZ1P83 OsbZ1P84 OsbZ1P84 OsbZ1P85 AtbZ1P36 OsbZ1P81 OsbZ1P81 OsbZ1P81 OsbZ1P81	Attg06839 Attg08320 Attg22070 Attg22070 Attg55210 Attg57920 At5g65210 Attg7920 At5g10030 Attg068640 At5g06950 At5g06	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA2	AJ314787 AJ314787 L10209 X68053 AJ315736 X68999 AF111711 BAB44077 D10042 X69900 AJ320540 AC079604 AC079604 AC079604 AC079604 AC079604 AC079604	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5 PoGO E1 PoGO E2
Abble 3 OsbZIP6 OsbZIP6 OsbZIP10 OsbZIP10 OsbZIP11 OsbZIP12 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP15 OsbZIP15 OsbZIP17 AtbZIP22 AtbZIP23 AtbZIP17 OsbZIP13 AtbZIP26 OsbZIP13 AtbZIP46 OsbZIP19 OsbZIP192 AtbZIP26 AtbZIP26 OsbZIP83 OsbZIP83 OsbZIP84 OsbZIP3 OsbZIP3 OsbZIP3 OsbZIP3 OsbZIP3 OsbZI	Attg06839 Attg06839 Attg08320 Attg22070 At5g65210 Attg77920 At5g10030 Attg068640 At5g06950 Attg06950 Attg06950 Attg06950 Attg068640 Attg12250	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA2	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711 BAB44077 D10042 X69900 AJ320540 AC079604 AC023064 AF401299 AF401299	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5 PoGO E1 PoGO E1
Aubzi 9 OsbZIP6 OsbZIP10 OsbZIP10 OsbZIP11 OsbZIP12 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP15 OsbZIP15 OsbZIP15 OsbZIP15 OsbZIP17 OsbZIP15 OsbZIP17 OsbZIP13 AtbZIP47 OsbZIP13 AtbZIP46 OsbZIP18 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP19 OsbZIP8 OsbZIP8 OsbZIP8 OsbZIP8 OsbZIP8 OsbZIP8 OsbZIP8 OsbZIP81 OsbZIP81 OsbZIP81 OsbZIP81 OsbZIP81 OsbZIP81 OsbZIP3 OsbZIP3<	Attg06839 Attg06839 Attg08320 Attg22070 At5g65210 Attg65210 Attg77920 Attg10330 Attg68640 Attg10250 Attg06950 Attg06950 Attg12250 Attg242380 Attg242380 Attg58120	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA6	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711 BAB44077 D10042 X69900 AJ320540 AC079604 AC023064 AC023064 AF401299 AF401300	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5 PoGO E1 PoGO E2 PoGO F1
Abble 3 OsbZIP6 OsbZIP6 OsbZIP10 OsbZIP10 OsbZIP10 OsbZIP10 OsbZIP11 OsbZIP11 OsbZIP11 OsbZIP12 AtbZIP20 Abble21P21 OsbZIP15 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP12 OsbZIP13 AtbZIP47 OsbZIP13 AtbZIP47 OsbZIP14 OsbZIP15 OsbZIP16 OsbZIP17 OsbZIP182 OsbZIP182 OsbZIP182 OsbZIP183 OsbZIP184 OsbZIP85 OsbZIP84 OsbZIP85 OsbZIP84 OsbZIP81	Attg06839 Attg06839 Attg08320 Attg22070 At5g65210 Attg65210 Attg65210 Attg66520 Attg66950 Attg075 Attg06950 Attg075 At	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA2 HBP-1b	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711 BAB44077 D10042 X69900 AJ320540 AC079604 AC023064 AC023064 AF401299 AF401300	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5 PoGO E1 PoGO E2
Abble 3 OsbZIP6 OsbZIP61 OsbZIP10 OsbZIP17 OsbZIP11 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP15 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP17 OsbZIP12 OsbZIP13 OsbZIP14 OsbZIP120 AtbZIP26 AtbZIP26 OsbZIP120 AtbZIP26 OsbZIP120 AtbZIP26 OsbZIP120 OsbZIP120 OsbZIP130 OsbZIP14 OsbZIP13 OsbZIP14 OsbZIP12 OsbZIP12 OsbZIP12	Attg06839 Attg06839 Attg08320 Attg22070 At5g65210 Att5g65210 Attg77920 Attg10030 Attg068640 Attg06950 At5g06950 At5g06950 At5g06950 Attg242380 At3g51960 Attg251960	TGA3 TGA1 OBF4 PAN HBP-1b TGA2 HBP-1b TGA6	AJ314787 AJ314787 L10209 X68053 AJ315736 X69899 AF111711 BAB44077 D10042 X69300540 AJ3020540 AC079604 AC079607 AC079607 AC079607 AC079607 AC079607 AC079607 AC079607 AC079607 AC079607 AC079607 AC079607 AC079607 AC079607 AC079767 AC079767 AC079767 AC07977777777777777777777777777777777777	PoGO D1 PoGO D2 PoGO D3 PoGO D4 PoGO D5 PoGO E1 PoGO E2 PoGO F1 PoGO E2



(1) LRRTES

A

Os genes do Grupo A indicam que eles têm um papel na via de sinalização de ácido abscísico (ABA) e de estresse abiótico em sementes e tecidos vegetativos. Estes bZIPs são capazes de se ligar a seqüências de elementos de responsta a ABA (ARE) para subscription de terresponsta a ABA (ARE) subscription d regular a taxa de transcrição de genes do tipo LEA (*Late embryogenesis abundant*)



(1) MPMYPAMYPLPMPWMHPYPMRGSQVPLVPIPRLKPQ (2) KAVEEKAKKKTKTKKVASISLLGLL



O Grupo C possui bZIPs bem caracterizadas como Opaco-2 e CFRF2. Genes ortólogos a Opaco-2 atuam na regulação da expressão de genes de proteínas de reserva e no metabolismo de carbono e nitrogênio, enquanto os ortólogos a CPRF2 atuam em resposta a patógenos ou estresse



(1) GAMAFDMEYARWLEEHNRHINELRTAVNAHA (2) HYDEIFRMKGVAAKADVFHVMSGMWKTPAERC FMWIGGFRPSELLKVLVPHLEPLTE (3) QQIMGICNLQQSCQQAEDALSQGMEKLQQSLA ETLAGGSPP

D

Genes do Grupo D podem atuar em dois processos desenvolvimento e defesa contra patógenos. O gene LG2 do milho atua na formação da lígula e o gene PAN de no controle do número de órgãos florais (PAN). Os fatores TGA estão envolvidos na indução sistêmica de genes relacionados a patogênese (PR)





bZIPno.	Codigo	Publicado	GenBank	
OsbZIP65				
OsbZIP66				
OsbZIP72		OsZIP-2a	AAC49557	PoGO G1
	At1a10400	USZIP-20	AAC49556	
OsbZIP67	Aligio430			
OsbZIP74		OSBZ8	AAB40291	
OsbZIP75				PoGO G2
AtbZIP54	At4g01120	GBF2	AF053228	
AtbZIP55	At2g46270	GBF3	U51850	
OsbZIP70	444-20720	ODEA	VCODOA	D =00.00
OshZIP69	At4930730	GDF1	703094	F000 03
OsbZIP71		GBF-1a	T03241	
AtbZIP16				PoGO G4
AtbZIP68	At1g32150			
OsbZIP64				D. 00.05
USDZIP66		TH 13 (T		POGO Go
OsbZIP81	A+2a17600	THY5	BAB62558	
Osh7IP77	Alby17009	HTD-like	AF400477	FUGU HI
OsbZIP79				PoGO H2
AtbZIP56	At5g11260	HY5	AB005295	
OsbZIP80				
AtbZIP60	At1g42990		AY045964	PoGO H3
OsbZIP91				
OsbZIP94	414 - 40700	1004	4 500 5000	PoGO I1
AtDZIP51	At1g43700	VIP1	AF225983	
OsbZIP 00		RF2a	AAC49832	
AtbZIP59	At2g31370	PosF21	X61031	PoGO I2
AtbZIP69	At1g06070		AJ419854	
OsbZIP89				
OsbZIP90				
Ath7IP52	At1q06850		AT0744208	POGO IS
OsbZIP87	Arigotooo		704 00107	
OsbZIP92				
OsbZIP93				
OsbZIP96			15404007	PoGO I4
AtbZIP29	At4g38900		AF401297	
AtbZIP31	At2g21250 At2g13150		AF401230	
AtbZIP32	At2g12980		AV566578	
AtbZIP33	At2g12900			PoGP I1
AtbZIP71	At2g24340			
AtDZIP74	At2g21235			
OsbZIP51				
OsbZIP52 OsbZIP53				
OsbZIP54				
OsbZIP57				
AtbZIP3	At5g15830		AV549429	PoGO J1
AtbZIP8	At1g68880		AF400621	
AlDZIP42 Ath7IP43	A13930530 At5a38800		BABUTUZU	
AtbZIP48	At2q04038		AC007178	
AtbZIP58	At1g13600		AF332430	
OsbZIP39				
OsbZIP40				
OsbZIP41				
OsbZIF42 OsbZIP43				
OsbZIP44				
OsbZIP45				
OsbZIP46		=		
OsbZIP47		LIP19	CAA40596	
OsbZIP48 OsbZIP49				
OsbZIP50				
OsbZIP55				
OsbZIP56				
AtbZIP1	At5g49450	00000	AF400618	
AtbZIP2	At2g18160 At1g59530	GBF5	AF53939 AF400610	
AtbZIP5	At3q49760		71 400019	
AtbZIP6	At2g22850			
AtbZIP7	At4g37730			
AtbZIP11	AT434590	ATB2		
AtbZIP44	At1g75390		AV506155	
AtbZIP70	At5g60830		74 400020	
AtbZIP75	At5g08141			
AtbZIP72	At5g08141			



Figura 7. Classificação das bZIPs de arroz e *Arabidopsis*. Onze grupos de genes homólogos (A a J e S) foram definidos a partir de análises filogenéticas por NJ do conjunto de bZIPs de arroz e *Arabidopsis* (Figuras 8 e 10). A organização em PoGOs foi realizada com base em análises filogenéticas mais detalhadas, de cada Grupo de genes homólogos, usando NJ. Esta análise incluiu, além das proteínas de arroz e *Arabidopsis*, proteínas de outras mono e eudicotiledôneas, e foi realizada utilizando motivos conservados entre membros de cada Grupo. Os motivos conservados foram evidenciados por alinhamentos ou pelo programa MEME (http://meme.sdsc.edu/meme/website/intro.html). O esquema representativo de uma proteína bZIP de cada Grupo ilustra a posição do seu domínio bZIP e dos demais blocos conservados. A seqüência referente a um motivo conservado reflete um consenso em multinível obtido por MEME. As funções conhecidas de membros de cada família estão relatadas. As colunas Código, Publicado e GenBank se referem ao código no MAtDB, ao nome com o qual foram publicados e o número de acesso no GenBank, respectivamente. AtbZIP72 não foi incluída em nenhum grupo de genes ortólogos, uma vez que seu posicionamento não é claro. Posição conservada de intron na estrutura do gene AtbZIP72 sugere que ele está mais relacionado aos Grupos B e H. PoGP - Possível Grupo de Parárologos.



Figura 8. Definição dos grupos de genes homólogos A, D e F. A figura é uma representação parcial da árvore sem raiz inferida por NJ do conjunto de 172 bZIPs de arroz e *Arabidopsis* usando uma matriz de distância do tipo PAM e um *bootstrap* de 1000 repetições, que está indicado em porcentagem nos ramos. O alinhamento utilizado corresponde ao domínio bZIP mínimo de 43 aminoácidos. Os Grupos D e F são irmãos, suportados por um *bootstrap* de 56%. Em azul os bZIPs de *Arabidopsis*, e em vermelho, os de arroz.

Apesar de possuir um suporte fraco de *bootstrap* (45%), todos os genes do Grupo A, com exceção de dois genes de *Arabidopsis Gbf4* e *AtbZIP13*, possuem pelo menos uma posição de intron conservada em sua seqüência (Figura 9), sugerindo uma origem evolutiva comum e confirmando o agrupamento. Os genes *Gbf4* e *AtbZIP13* não possuem introns, enquanto seus possíveis ortólogos em arroz apresentam no mínimo um intron conservado deste Grupo (Figura 8).

Grupo	F	Q DQQENDHSDSS	SNKKRLCG <mark>N</mark> REAVF	RKYREKKI	KARTAY <mark>l</mark>	ede 1
Grupo	D	SSDRSKDKLDQ	TLRRLAQ <mark>N</mark> REAAF	RKSRLRKI	K A YVQQ <mark>L</mark>	ENS
Grupo	A	RVVDGPVEKVVI	ERRQRMIK <mark>N</mark> RESA <i>P</i>	ARSRARK	2 A YTVE <mark>L</mark>	EAE 2
Grupo	G	GVPQPWNEKEV	KREKRKQS <mark>N</mark> RESAF	RRSRLRK	AETEQ <mark>L</mark>	SVK
Grupo	С	DGETNMNPTNV	KRVKRMLSNRESAF	RRSRRRK	2ahlse <mark>l</mark>	ет <u>6</u>
Grupo	Н	KRGRTPAEKEN	KLRKR L LR <mark>N</mark> RVSAÇ)QARERKI	KAYLSE <mark>L</mark>	ENR
Grupo	E	GGSSGNRIHDP	KRVKR i la n rqsaç)RSRVRK	LQYISE <mark>l</mark>	ERS
Grupo	I	STKLKEVASDP	KEVRR <mark>I</mark> lk n qesa <i>f</i>	ARSKQKKI	LQYMIN <mark>L</mark>	ELK
Grupo	S	IFHNEGLAPEE	RARRMVSNRESAF	RRSRMRKI	KKQIEE <mark>l</mark>	222
Grupo Grupo	J B	EQQTNNNIINEE	RKQRRMIS n resaf Knvrlvr n resaf	RSRMRK	DRHLDEL Khyveel	WSQ 4 Edk
01 0P 0	-		Domínio Bás	ico	·····	

Figura 9. Posições conservadas de introns na região do domínio básico dos fatores de transcrição do tipo bZIP de angiospermas. A primeira leucina do zíper de leucinas está em verde e a asparagina conservada no domínio básico em vermelho. Os bZIPs utilizados nesta figura são AtbZIP24 (Grupo F), TGA6 (Grupo D), ABI5 (Grupo A), GBF2 (Grupo G), Bzo2h3 (Grupo C), HY5 (Grupo H), AtbZIP61 (Grupo E), AtbZIP31 (Grupo I), AtbZIP70 (Grupo S), AtbZIP42 (Grupo J), AtbZIP49 (Grupo B).

A explicação mais provável para a ausência de introns nestes genes envolve a síntese de um cDNA via transcriptase reversa e posterior integração no genoma. Este tipo de cópia duplicada por transcriptase reversa é conhecido como retrogene (Graur & Li, 2000; Sosinsky *et al.*, 2000). Em seguida, o gene ancestral com introns teria sido perdido, e a cópia sem intron teria sido duplicada, gerando os dois parálogos, GBF4 e AtbZIP13. A relação de paralogia entre estes genes foi confirmada pelo fato de estarem presentes em um segmento duplicado do genoma de *Arabidopsis* (Tabela VIII).

As seqüências dos Grupos A e D, por possuírem apenas 3 repetições de leucina, limitam a resolução da análise filogenética. A exclusão dos membros dos Grupos A e D gerou um subconjunto de 113 proteínas, para o alinhamento dos bZIPs foi aumentado em duas repetições de leucina em relação ao motivo mínimo, formando um novo motivo de 60 aminoácidos. Análise por NJ deste novo alinhamento permitiu melhorar significativamente a resolução das relações evolutiva dentro do subconjunto de 113 proteínas.

Esta análise resultou em uma árvore cuja topologia é próxima à obtida anteriormente por Vincentz *et al.* (2003). É possível identificar sete grupos (B, C, E, F, G, H e I) que apresentam um suporte de *bootstrap* superior a 50% (Figura 10). A presença de no mínimo um intron cuja posição é conservada e específica em genes de cada uma destes grupos, (Figura 9), confirma uma origem ancestral comum dos membros de cada um dos grupos. Apesar de possuírem um *bootstrap* de 66%, os bZIPs dos Grupos B e H foram classificados em grupos distintos de genes homólogos baseado em uma estrutura gênica diferente e na presença de motivos conservados (Figuras 7 e 9).

Tabela VIII.	Pares de genes resultantes				
de duplicação	de duplicação de segmentos do genoma				
de Arabidops	de Arabidonsis				
Grupo A					
•	DPBF4 e AREB3				
•	AtbZIP13 e GBF4				
•	ABF4 e ABF1				
•	ABF2 e ABF1				
•	AtbZIP14 e AtbZIP17				
Grupo B					
•	AtbZIP49 e AtbZIP17				
•	AtbZIP49 e AtbZIP28				
Grupo D					
•	TGA1 e OBF4				
•	AtbZIP50 e TGA3				
•	TGA3 e TGA1				
•	PAN e AtbZIP21				
•	TGA6 e TGA2				
Grupo E					
•	AtbZIP61 e AtbZIP34				
Grupo F					
•	AtbZIP23 e AtbZIP19				
Grupo G					
•	GBF2 e GBF3				
•	AtbZIP16 e AtbZIP68				
Grupo I					
•	AtbZIP18 e VP1				
•	AtbZIP18 e AtbZIP52				
•	POSF21 e AtbZIP69				
•	AtbZIP30 e AtbZIP29				
Grupo J					
•	AtbZIP58 e AtbZIP48				
•	AtbZIP42 e AtbZIP43				
Grupo S					
•	GBF5 e Atb2				
•	GBF5 e AtbZIP44				
•	AtbZIP6 e AtbZIP7				
Determinação de	parálogos foi obtida através do site				
Paralogons in Ar	abidopsis thaliana				



Figura 10. Árvore filogenética de bZIPs de arroz e *Arabidopsis*. A árvore sem raiz foi inferida por NJ do conjunto de 113 bZIPs usando uma matriz de distância do tipo PAM e um *bootstrap* de 1000 repetições, que está indicado em porcentagem nos ramos. O alinhamento utilizado corresponde ao domínio bZIP mínimo de 43 aminoácidos acrescido de duas repetições de leucinas, totalizando 60 aminoácidos. Os Grupos D e F, assim como I e E, são irmãos. Em azul os bZIPs de *Arabidopsis*, e em vermelho, os de arroz.

Para um conjunto de 17 bZIPs de *Arabidopsis* e 19 de arroz esta análise não permitiu obter nenhum agrupamento óbvio (Grupos J e S, Figura 10). Entretanto, ao contrário das demais bZIPs acima descritas, é importante ressaltar que nenhum destes 36 genes de bZIPs possuem introns em sua estrutura, sugerindo que estes bZIPs tenham uma origem comum. Baseados em análises mais detalhadas deste conjunto de 36 genes, que serão descritas no <u>item 3.4</u>, os membros deste conjunto foram posteriormente classificados em dois Grupos (Grupo J e Grupo S, Figura 15).

Os 11 Grupos de possíveis genes homólogos resultantes desta análise, correspondem a uma nova classificação de bZIPs de angiospermas (Figura 7), que considerou também os resultados obtidos em dois trabalhos anteriores (Jakoby *et al.*, 2002; Vincentz *et al.*, 2003).

Para obter grupos de ortólogos e evidenciar eventos chaves que contribuíram para a evolução da família multigênica de fatores bZIP em angiospermas, uma análise filogenética mais detalhada de cada Grupo foi realizada.

3.4. Análises Filogenéticas entre Mono e Dicotiledôneas e formação de PoGOs

Para a identificação de PoGOs de bZIPs de mono e eudicotiledôneas, foi realizada uma análise filogenética mais detalhada dos Grupos de genes homólogos A, B, C, D, E, F, G, H, I e S. Para tanto, procurou-se otimizar as resoluções das relações evolutivas dentro de cada Grupo pelo aumento da quantidade de informação utilizada nas análises, tanto no comprimento das seqüências, uma vez que bZIPs de um mesmo Grupo apresentam regiões conservadas específicas (Figura 7), quanto na inclusão de outras bZIPs de angiospermas, identificadas a partir de *clusters* de ESTs ou presentes no banco Angiotot. O uso dos *clusters* de ESTs requer uma distribuição preliminar dos mesmos, entre os vários Grupos, baseado na maior similaridade obtida com um bZIP do banco de referência Angiotot (<u>item</u> 2.4 de Materiais e Métodos e Figura 5).

O detalhamento das relações evolutivas entre mono e eudicotiledôneas por análises do tipo *neighbor-joining* possibilitou a identificação de 29 PoGOs nos Grupos de A a I (Figura 7). Vale ressaltar que para o Grupo C, a obtenção de três PoGOs só é possível quando levado em consideração a análise por Máxima Parcimônia realizada por Vincentz *et al.* (2003), um vez que as análises por *neighbor-joining* suportam a formação de apenas dois PoGOs por *bootstrap* (Figura 11). Este fato sugere que uma análise de cada Grupo por Máxima Parcimônia seria uma abordagem complementar ao trabalho aqui descrito, auxiliando em uma melhor determinação dos grupos de ortólogos onde a análise por *neighbor-joining* não foi suficientemente resolutiva, como no Grupo C. Também se pode mencionar que só é possível identificar o PoGO H2 quando o Grupo H é analisado em conjunto com o Grupo B (Figura 12).

Análises filogenéticas de bZIPs do Grupo G revelaram que além de quatro PoGOs entre mono e dicotiledôneas, este grupo possui um PoGO exclusivo a monocotiledôneas (PoGO G5, Figura 13). Duas hipóteses podem explicar este fato: (i) a linhagem das eudicotiledôneas perdeu o gene correspondente; (ii) a aquisição deste gene é resultado de uma duplicação específica de monocotiledôneas. No caso de a segunda hipótese estar correta, os bZIPs do PoGO G5 podem estar relacionados a processos específicos a monocotiledôneas. Dois dos bZIPs deste PoGO, EmBP de milho e de trigo, estão relacionados com a síntese de proteínas de reserva (prolaminas, Carlini *et al.*, 1999).



Figura 11. Árvore filogenética de bZIPs de mono e eudicotiledôneas do Grupo C. A árvore sem raiz foi inferida por uma análise NJ das distâncias obtidas a partir de uma matriz de distâncias do tipo PAM. Os valores de *bootstrap* correspondem a 1000 repetições e estão indicados em porcentagem nos ramos. O alinhamento em aminoácidos utilizado corresponde ao domínio conservado dentro deste Grupo (Figura 7). Em azul os bZIPs de *Arabidopsis*, em vermelho, os de arroz, em verde os das demais eudicotiledôneas e em negro, os das demais monocotiledôneas.



Figura 12. Árvore filogenética de bZIPs de mono e eudicotiledôneas dos Grupos B e H. A árvore sem raiz foi inferida por uma análise NJ das distâncias obtidas a partir de uma matriz de distâncias do tipo PAM. Os valores de *bootstrap* correspondem a 1000 repetições e estão indicados em porcentagem nos ramos. O alinhamento em aminoácidos utilizado corresponde ao domínio conservado dentro do Grupo B (Figura 7). A visualização dos PoGOs H1 e H2 só é possível quando o Grupo H é analisado em conjunto com o Grupo B. Em azul os bZIPs de *Arabidopsis*, em vermelho, os de arroz, em verde os das demais eudicotiledôneas e em negro, os das demais monocotiledôneas.



Figura 13. Árvore filogenética de bZIPs de mono e eudicotiledôneas do Grupo G. A árvore sem raiz foi inferida por uma análise NJ das distâncias obtidas a partir de uma matriz de distâncias do tipo PAM. Os valores de *bootstrap* correspondem a 1000 repetições e estão indicados em porcentagem nos ramos. O alinhamento em aminoácidos utilizado corresponde ao domínio conservado dentro deste Grupo (Figura 7). Em azul os bZIPs de *Arabidopsis*, em vermelho, os de arroz, em verde os das demais eudicotiledôneas e em negro, os das demais monocotiledôneas.

Buscas exaustivas demonstram que o possível grupo de genes parálogos (PoGP, Grupo I, Figura 7), formado pelos bZIPs AtbZIP74, AtbZIP71, AtbZIP33, AtbZIP32 e AtbZIP31 de *Arabidopsis*, é restrito a este organismo. A hipótese mais provável para explicar estas observações consiste na formação deste grupo de parálogos como resultante de um evento recente de duplicação restrita à linhagem da qual se originou *Arabidopsis* seguida de evolução rápida.



Figura 14. Árvore filogenética de bZIPs de mono e eudicotiledôneas do Grupo I. A árvore sem raiz foi inferida por uma análise NJ das distâncias obtidas a partir de uma matriz de distâncias do tipo PAM. Os valores de *bootstrap* correspondem a 1000 repetições e estão indicados em porcentagem nos ramos. O alinhamento em aminoácidos utilizado corresponde ao domínio conservado dentro deste Grupo (Figura 7). O PoGP I1 pode ser visualizado na árvore da Figura 10 com um ramo com 92% de *bootstrap*.Em azul os bZIPs de *Arabidopsis*, em vermelho, os de arroz, em verde os das demais eudicotiledôneas e em negro, os das demais monocotiledôneas.

Estes resultados sugerem que, tanto o PoGO G5, quanto o PoGP I1, por serem restritos a algumas linhagens evolutivas, podem estar desempenhando um conjunto de funções não essenciais para uma angiosperma.

Análises filogenéticas entre arroz e *Arabidopsis* indicam que os bZIPs dos Grupos J e S estão mais externos aos demais grupos na árvore, uma vez que não apresentam suporte de *bootstrap* nestas análises. Além disto, todos os genes destes Grupos não possuem introns, o que sugere uma origem ancestral comum.

As análises mais detalhadas entre mono e dicotiledôneas mostram que os genes do Grupo S não se agrupam de forma consistente, corroborando o resultado previamente obtido da análise entre arroz e *Arabidopsis* (Figura 10). Entretanto, é possível identificar o Grupo J (Figura 15), que apresenta um suporte de 84% de *bootstrap*, e corresponde a um único PoGO de mono e eudicotiledôneas. A dificuldade em organizar em PoGO as seqüências do Grupo S em relação aos demais grupos de homólogos sugere que seqüências deste Grupo divergiram rapidamente após a separação das mono e eudicotiledôneas. A possível rápida evolução pode estar relacionada a uma baixa seleção negativa ou a uma seleção positiva. Para identificar qual destas duas alternativas está correta testes da razão dN/dS entre pares de genes ortólogos podem ser realizados. A razão dN/dS (ω), quando inferior a 1 a presença de seleção purificadora ou negativa, e quando superior a 1, positiva (Kellogg, 2004; Nam *et al.*, 2004). No entanto, a comparação do valor de ω calculado a partir de onze pares de ortólogos de sorgo e cana de diferentes Grupos não é conclusivo a respeito de uma evolução mais acelerada no Grupo S (dados não apresentados).



Figura 15. Árvore filogenética de bZIPs de mono e eudicotiledôneas dos Grupos J e S. A árvore sem raiz foi inferida por uma análise NJ das distâncias obtidas a partir de uma matriz de distâncias do tipo PAM. Os valores de *bootstrap* correspondem a 1000 repetições e estão indicados em porcentagem nos ramos. O alinhamento em aminoácidos utilizado corresponde ao domínio conservado dentro destes Grupos (Figura 7). A identificação do Grupo J foi apenas possível nesta análise mais detalhada. A inclusão de bZIPs de outras angiospermas não aumentou a resolução da árvore. Em azul os bZIPs de *Arabidopsis*, em vermelho, os de arroz, em verde os das demais eudicotiledôneas e em negro, os das demais monocotiledôneas.

A existência de bZIPs de gimnosperma nestes Grupos (Dados não mostrados) exclui a possibilidade de os Grupos S e J serem uma aquisição recente restrita a angiospermas. Recentemente, Satoh *et al.* (2004) identificaram que os membros de um subgrupo do Grupo S (ATB2, AtbZIP44, GBF5 e AtbZIP53), em *Arabidopsis*, são aparentemente as únicas bZIPs capazes de reconhecer um elemento *cis* do gene da prolina desidrogenase (ProDH). O fato de que apenas estas bZIPs podem reconhecer este sítio seria uma evidência para agrupar estas bZIPs em um único grupo de genes homólogos. Os possíveis ortólogos em arroz deste subgrupo poderiam ser identificados por abordagens experimentais similares. Não é possível agrupar os demais genes do Grupo S, no entanto, como existem 3 grupos de parálogos distintos em *Arabidopsis* e 4 em arroz, podemos dizer que possivelmente existam 3 PoGOs.

Baseado no conjunto de resultados obtidos, é possível evidenciar 33 PoGOs distribuídos entre os Grupos de A a J e S, e um PoGO exclusivo de monocotiledôneas no Grupo G. Estes 33 PoGOs representam 33 possíveis funções ancestrais desempenhadas pelas bZIPs em angiospermas.

3.5. Possível modelo evolutivo de bZIPs em Angiospermas

Com base nos dados da análise filogenética e na estrutura dos genes (Figuras 8, 9 e 10), foi elaborado um possível modelo da história evolutiva das bZIPs em angiospermas (Figura 16).



Figura 16. Modelo evolutivo dos onze grupos de genes homólogos de bZIPs em angiospermas. O modelo foi construído baseado nas análises filogenéticas representadas Figuras 8 e 10 e na posição conservada de introns. Em parênteses, o número de PoGOs que foram identificados em cada Grupos. Os dois grandes conjuntos de genes estão indicados em azul e roxo. O padrão de ramificação dentro destes dois grandes conjuntos foi estabelecido baseado no *bootstrap* (indicados em porcentagem nos ramos) obtidos nas análises filogenéticas. A posição dos Grupos J e S não é bem definida dentro do modelo.

Este modelo consiste na separação de nove grupos possíveis de genes homólogos em dois grandes conjuntos, sendo que um ramo evolutivo envolve os Grupos I, H, E, C e B e o outro ramo inclui os Grupos D, A, G e F. O padrão de ramificação sugerido no modelo foi deduzido a partir do suporte *bootstrap* das análises globais (Figuras 8 e 10). A posição dos Grupos S e J dentro deste modelo não é clara. Estes grupos podem tanto ter uma origem ancestral quanto ter evoluído a partir de um dos dois conjuntos. Alternativamente, estes Grupos estariam sendo formados por um evento de recombinação entre os Grupos C e G (exon *shuffling*), e, portanto estariam posicionados na congruência dos dois grandes grupos. Esta última hipótese é baseada em análises preliminares de seqüências, indicando uma similaridade significativa tanto em exons do Grupo B quanto exons do Grupo G (dados não mostrados). A partir dos 11 grupos de genes homólogos foram evidenciadas 33 possíveis funções ancestrais de angiospermas representadas por PoGOs de mono e eudicotiledôneas. De maneira geral, a composição dos PoGOs reflete a distribuição de bZIPs entre arroz (113) e *Arabidopsis* (76), uma vez que para 29 dos 33 PoGOs os bZIPs de arroz estão representados em número igual ou maior que os de *Arabidopsis* (Figura 7).

A distribuição de PoGOs e genes nos diversos Grupos (Figura 17) indica que os Grupos A, D, G, I e S englobam 64% dos PoGOs e 70% do genes. É interessante notar que estes Grupos, com exceção de I e S, estão envolvidos com adaptação ao ambiente (Figura 7). Estes dados sugerem que genes associados a processos de resposta ao meio tendem a passar por processos de duplicação mais intensos (Shiu *et al.*, 2004). No caso dos Grupos I e S, não existe informação funcional suficiente para relacionar as duplicações gênicas com aspectos adaptativos precisos.

De modo interessante, quatro PoGOs (D4, G3, H1 e H3) apresentam relação de um gene de arroz para um de *Arabidopsis*. Esta relação 1:1 constitui-se em uma provável evidencia de que estes bZIPs estão envolvidos com funções essenciais relacionadas a aspectos do desenvolvimento comum a angiospermas (Shiu *et al.*, 2004).



Figura 17. Distribuição do número de PoGO e de genes nos vários Grupos de homólogos.

3.6. Utilização da classificação em PoGOs para análise funcional.

A quantidade de informações disponíveis nos bancos de dados requer a elaboração de novas metodologias a fim de racionalizar o estudo de genes.

A identificação de famílias multigênicas e sua organização em PoGOs representa uma abordagem recente que deve facilitar a transferência de informações bioquímicas, estruturais e funcionais de uma proteína para outra, uma vez que é possível admitir que existe uma grande probabilidade de que todos os outros genes de um PoGO desempenhem a mesma função (Tatusov *et al.*, 1997; Kellogg, 2004). A Figura 7 apresenta um resumo das possíveis funções atribuídas a cada agrupamento.

A validação da integração dos dados filogenéticos (PoGO) com estudos funcionais podem ser ilustradas com alguns exemplos de genes já caracterizados.

No Grupo D a presença dos bZIPs HBP-1b (Histone promoter-Binding Protein) de *Arabidopsis* e de trigo no mesmo PoGO (PoGO D5; Figuras 7 e 18), é de acordo com a observação de que estes dois bZIPs se ligam especificamente a um elemento *cis* das Histonas H3 e H4, e possivelmente regulam a transcrição destes genes (Kawata *et al.*, 1992; Mikami *et al.*, 1994). Além disto, no mesmo PoGO, foi mostrado que os parálogos TGA2, TGA6 e HBP-1b de *Arabidopsis*, atuam na defesa contra patógenos, mais especificamente na resistência sistêmica. Estudo de mutantes para estes três genes indica que eles têm funções redundantes (Zhang *et al.*, 2003). Este exemplo ilustra o grau de redundância que pode existir entre genes parálogos. Ortólogos destes genes em feijão, tabaco e batata, identificados pelas análises filogenéticas, apresentam, como esperado, funções relacionadas à resposta a patógenos (Feltkamp *et al.*, 1994; Niggeweg *et al.*, 2000; Tucker *et al.*, 2002).

No Grupo H, os genes HY5 e HY5-like de *Arabidopsis*, que apesar de estarem incluídos em PoGOs distintos (H1 e H2, Figura 12), apresentam redundância funcional. Estes dois genes estão envolvidos com a fotomorfogênese e podem atuar como homo ou heterodímeros (Holm *et al.*, 2002; Ulm *et al.*, 2004). Em arroz, existem dois possíveis ortólogos (Figura 12), e a caracterização destes pode ser um aspecto importante no entendimento da fotomorfogênese em monocotiledôneas.



Figura 18. Árvore filogenética de bZIPs de mono e eudicotiledôneas do Grupo D. A árvore sem raiz foi inferida por uma análise NJ das distâncias obtidas a partir de uma matriz de distâncias do tipo PAM. Os valores de *bootstrap* correspondem a 1000 repetições e estão indicados em porcentagem nos ramos. O alinhamento em aminoácidos utilizado corresponde ao domínio conservado dentro deste Grupo (Figura 7). Em azul os bZIPs de *Arabidopsis*, em vermelho, os de arroz, em verde os das demais eudicotiledôneas e em negro, os das demais monocotiledôneas.

No Grupo C, o PoGO C1 é formado por um par de parálogos em monocotiledôneas que inclui, entre outros, os genes Opaco-2 e o par OHP1 e OHP2, de milho (Figura 11). Opaco-2 de milho é um regulador chave do balanço nitrogênio/carbono durante o desenvolvimento da semente (Schmidt *et al.*, 1990). O grupo de ortólogos de Opaco-2 em outras monocotiledôneas tem sua expressão restrita ao endosperma (OsbZIP63, SPATa, BLZ2Hv, O2Sb, O2Zm e O2Clj; Onodera *et al.*, 2001; Albani *et al.*, 1997; Oñate *et al.*, 1999; Pirovano *et al.*, 1994; Schmidt *et al.*, 1990; Vettore *et al.*, 1998). Os demais bZIPs do Grupo C não apresentam um padrão de expressão restrito a endosperma. Análises filogenéticas indicam que o ortólogo mais provável de Opaco-2 em *Arabidopsis* é Bzo2h3. A possível função de Bzo2h3 pode ser investigada considerando sua eventual participação em processos de metabolismo do nitrogênio e carbono, embora a expressão deste gene não é restrita a endosperma.

A fim de elucidar a função dos genes do Grupo C, iniciamos um estudo funcional do gene *Bzo2h2*, que foi definido como gene único que representa uma função ancestral e representa um alvo para estudos de genética reversa. Apesar de ser um gene único, foi mostrado que alelos mutantes nulos por inserção de TDNA no gene *Bzo2h2* não alteraram o desenvolvimento padrão de *Arabidopsis*. Este fato sugere que possivelmente exista uma redundância funcional com os outros membros do Grupo C (Bzo2h1, Bzo2h3 e Bzo2h4, Gauer, 2004).

Um caso de funções antagonistas foi descrito no Grupo A. O bZIP ABI5 (PoGO A4, Figura 19) está relacionado a sensitividade a ABA durante a germinação e à regulação de expressão dos genes tipo LEA (*late embryogenesis-abundant*) (Finkelstein, 1993, 1994; Gaubier *et al.*, 1993). Mais recentemente estudos funcionais de DPBF4 (PoGO A1)

52

indicam que o mutante nulo para seu gene é responsável por um aumento na taxa de transcrição de alguns genes do tipo LEA. DPBF4 e ABI5 são capazes de se ligar ao promotor do gene AtEm1 (um tipo LEA), sendo que existe uma competição entre eles, na qual DPBF4 atua como repressor e ABI5 como ativador da transcrição (Bensmihen *et al.*, 2002).

A organização em PoGO permitiu identificar vários genes únicos de *Arabidopsis* ou de arroz, que possivelmente representam funções ancestrais de angiospermas (Tabela IX). Tais genes, pela possível ausência de redundância de função, constituem-se em alvos privilegiados para estudos funcionais envolvendo técnicas de genética reversa (ex.: obtenção de alelos nulos por inserção de TDNA, superexpressão entre outros).

Até o presente momento não existe qualquer informação funcional sobre as bZIPs do PoGO E1 . O alinhamento das seqüências de bZIP deste PoGO evidenciou a presença de um domínio básico 5 aminoácidos mais curto e a asparagina (N), altamente conservada dentro do motivo básico de todas as bZIPs de angiospermas, é substituída por lisina (K) ou arginina (R) (Figura 20). Estudos com Opaco-2 de milho mostraram que a região básica inclui uma seqüência de localização nuclear (NLS B), sendo necessária ao seu endereçamento nuclear correto (Figura 21, Varagona *et al.*, 1994). Além disto, a região básica das bZIPs é responsáveis pela especificidade de ligação das bZIPs (Varagona *et al.*, 1994; Hurst, 1995).



Figura 19. Árvore filogenética de bZIPs de mono e eudicotiledôneas do Grupo A. A árvore sem raiz foi inferida por uma análise NJ das distâncias obtidas a partir de uma matriz de distâncias do tipo PAM. Os valores de *bootstrap* correspondem a 1000 repetições e estão indicados em porcentagem nos ramos. O alinhamento em aminoácidos utilizado corresponde ao domínio conservado dentro deste Grupo (Figura 7). Em azul os bZIPs de *Arabidopsis*, em vermelho, os de arroz, em verde os das demais eudicotiledôneas e em negro, os das demais monocotiledôneas.

ingrospermus		
Gene	PoGO	
OsZIP62	PoGO C2	
AtbZIP65	PoGO D1	
AtbZIP21	PoGO D2	
PAN	PoGO D4	
AtbZIP24	PoGO F1	
OsbZIP2	PoGO F2	
AtbZIP62	PoGO G1	
GBF1	PoGO G4	
OsZIP81	PoGO H1	
VIP1	PoGO I1	

Tabela IX. Genes únicos representativos de funções ancestrais de angiospermas

	Região Básica		Zipper de Leucinas
BF051625Le	KRIKRILA <mark>N</mark> RQSAQRSRVRH	∣ LQY	ISE l ersvtt l qaevsv l sp
AtbZIP61	KRVKRILA <mark>N</mark> RQSAQRSRVRF	LQY	ISE l ersvts l qtevsv l sp
AtbZIP34	KRVKRILA <mark>N</mark> RQSAQRSRVRF	LQY	ISE l ersvts l qaevsv l sp
BG044358Gm	KRVKRILA <mark>N</mark> RQSAQRSRVRI	LQY	ISE l ersvts l qaevsv l sp
BJ279680Ta	KRVKRILA <mark>N</mark> RQSAQRSRVRI	LQY	ISE L ERCVTT L QNEVSV L SP
AAAA010016110s	KRVKRILA <mark>N</mark> RQSAQRSRVRH	LQY	ISE l ersvtt l qnevsv l sp
AAAA010120610s	KRVKRILA <mark>N</mark> RQSAQRSRVRH	LQY	ISE L ERSVTS L QTEVSA L SP
BF588212Sb	FRKTRI LA <mark>n</mark> rqsaqrsrvri	LQY	ISE l ersvts l qtevsa l sp
BE420598Hv	KRVKRILA <mark>N</mark> RQSAQRSRVRI	LQY	ISE L ERSVTG L QMEVSA L SP
AtbZIP77	KRI <mark>K</mark> HQNAHRARLRI	LEY	ISD l ertiqv l qvegce m ss
AtbZIP76	KRA <mark>K</mark> QQFAQRSRVRF	LQY	ISE l ernvqt l qaegsk <mark>v</mark> sa
AtbZIP78	KRA <mark>R</mark> QQFAQRSRVRI	IQY	IAELERNVQMLQV
AW218730Le	KRA <mark>K</mark> QQFAQRSRVRI	LQY	IAE L ERNVQA L QAEGSE V SA
BM893453Gm	KRA <mark>K</mark> QQFAQRSRVRI	LQY	IAE L ERNVQA L QAEGSE V SA
AAAA010069690s	KRA <mark>K</mark> QQYAQRSRVRI	LQY	IAE l erkvqa l qsegid v sa
BE599212Sb	KRA <mark>K</mark> QQYAQRSRVRF	LQY	IAE l egrvqa l qsegve <mark>v</mark> sa
AJ431819Hv	KRA <mark>K</mark> QQYAQRSRVRF	LQY	IAE l egkvqs l qsegie <mark>v</mark> sa
BE44449Ta	KRA <mark>K</mark> QQYAQRSRVRF	LQY	IAE l egkvqs l qsegie <mark>v</mark> sa
		-	

Figura 20. Alinhamento do domínio bZIP do Grupo E. Em vermelho estão destacadas as proteínas do PoGO E1 e em verde as proteínas relacionadas. A asparagina (N) conservada dentro da maioria das bZIPs de angiospermas está mostrada em verde e as leucinas do zipper de leucina estão apresentadas em vermelho. AtbZIP78 apresenta apenas duas leucinas. As abreviações das espécies são: GM, *Glycine max*; HV, *Hordeum vulgare*; Le, *Lycopersicon esculentum*; Os, *Oryza sativa*; Sb, *Sorghum bicolor*; Ta, *Triticum aestivum*; Zm, *Zea mays*.



Figura 21. Ausência de NLS B bipartido nos bZIPs do PoGO E1. A seqüência NLS B bipartida de Opaco-2 está destacada em vermelho. NLS B é formado por duas partes: uma composta por quatro aminoácidos básicos (1), e a segunda composta de 7 aminoácidos com predomínio dos básicos (2). Em laranja estão destacadas as bZIPs do PoGO E1 que possuem um domínio básico diferente dos demais bZIPs, no qual não é possível destacar o NLS B bipartido. Hv, *Hordeum vulgare*.

A possível ausência de uma NLS B bipartida completa (Figura 20) sugere que as proteínas bZIP do PoGO E1 não estejam sendo direcionadas para ao núcleo. Para testar esta hipótese, fusões traducionais de AtbZIP76 e AtbZIP 78 com o gene repórter RFP (*Red Fluorescent Protein*) foram realizadas e utilizadas em um sistema de expressão transiente em células de cebola (*Allium cepa*) para analisar a localização celular destas proteínas. Resultados preliminares indicam que estes bZIPs estão direcionados ao núcleo (Figura 22), o que sugere que o domínio bipartido NLS B pode não ser essencial para o endereçamento nuclear, ou a primeira parte deste domínio pode ter características diferentes.



Figura 22. Análise da localização celular de AtbZIP76 e AtbZIP78. Expressão transiente dos fatores de regulação da transcrição do tipo bZIP AtbZIP76 e AtbZIP78 fusionados ao gene marcador RFP, em células epidérmicas de cebola (*Allium cepa*). A-C, Controles. A)Controle positivo com Pol RFP, mostrando um padrão de expressão P difuso pelo citoplasma e núcleo da célula. B) Controle de endereçamento citoplasmático. C) Controle de endereçameto nuclear. D) Bombardeamento com AtbZIP78 fusionada a RFP, demonstrando a expressão dessa proteína no núcleo da célula E) Coloração núcleo-específica com DAPI, mostrando o núcleo das células corados em azul mais intenso. F-K, Co-bombardeamento entre AtbZIP78 fusionada a RFP e o controle citoplasmático F) e J) Excitação em 528-523nm, mostrando o núcleo das células fluorescendo no vermelho G) e H) Excitação em 465-495nm, mostrando a expressão de GFP no citoplasma da célula H) Coloração núcleo-específica com DAPI. I) Células iluminadas com luz branca. L-O, Co-bombardeamento entre AtbZIP78 e o controle nuclear. L) Excitação em 528-523nm, mostrando o núcleo das células fluorescendo no vermelho G) e no núcleo das células iluminadas com luz branca. L-O, Co-bombardeamento entre AtbZIP78 e o controle nuclear. L) Excitação em 528-523nm, mostrando o núcleo das células fluorescendo no vermelho G) e AtbZIP78 e o controle nuclear. C) Excitação em 528-523nm, mostrando o núcleo das células fluorescendo no vermelho M) Excitação em 465-495nm, mostrando o núcleo das células fluorescendo no vermelho M) Excitação em 465-495nm, mostrando GFP sedo expresso GFP no núcleo das células N) Coloração núcleo-específica com DAPI O) Células iluminadas com luz branca.

4. Conclusões e Perspectivas

A utilização de métodos filogenéticos para estudos de genômica comparativa vem se provando um meio de otimizar os trabalhos de caracterização gênica, na media em que os grupos de ortólogos assinalados representam uma função ancestral, e a caracterização funcional de um dos membros deste grupo de ortólogos pode ser transferida para os demais membros deste grupo. No sentido de racionalizar estudos envolvendo bZIPs de angiospermas e em uma tentativa de entender a história evolutiva destes fatores de transcrição neste grupo, análises filogenéticas identificaram 33 PoGOs, distribuídos em 11 grupos de genes homólogos, além de um PoGO específico de monocotiledôneas (PoGO G5) e um grupo de parálogos de *Arabidopsis* (PoGP I1). No entanto, para alguns grupos não foi possível a identificação de PoGOs com suporte de *bootstrap* significativo. Análises filogenéticas adicionais, por MP ou MV, devem ser aplicadas para tentar solucionar o viés obtido.

Dentre as 116 seqüências de bZIP de arroz identificadas, 17 não foram utilizadas em análises filogenéticas pelo fato de não apresentarem o motivo bZIP mínimo para as análises filogenéticas (Tabela VII). Para tentar recuperar a seqüência completa destes 17 bZIPs uma busca em novos bancos de dados pode ser realizada, bem como entrar em contato com grupos que geraram seqüências de arroz, a fim de validar estes 17 bZIPs. Uma alternativa é a utilização de oligos para amplificar diretamente de arroz estes bZIPs e seqüência-los.

A formação de COGs (*Cluster of Orthologous Genes*, Tatusov *et al.*, 1997), pelo programa COGnitor (NCBI), é uma outra abordagem fenética que vem sendo usada para identificar a relação evolutiva entre de diversas linhegens (referencia COG e KOG). Os
COGs têm a mesma definição de um PoGO, mas se diferem quanto a sua formação, uma vez que os COGs são montados com base no melhor alinhamento obtido através do programa Blast. A análise dos nossos pelo programa COGnitor servirá como fonte para a comparação dos resultados obtidos nas duas abordagens.

Uma das questões que ainda permanece não resolvida é a origem dos Grupos J e S. Uma das hipóteses levantadas é que estes grupos de genes evoluíram rapidamente. Entretanto, os testes de dN/dS não foram conclusivos. Testes adicionais de taxa de evolução usando várias linhagens ancestrais, como gimnospermas, é uma abordagem que pode ser mais informativa quanto a esta questão.

Análises filogenéticas mais detalhadas, envolvendo linhagens evolutivas mais antigas que angiospermas, e posteriormente incluindo informações de outros eucariotos (humano, drosófila e levedura) pode permitir a elaboração de modelos mais completos sobre a evolução das bZIPs em eucariotos.

Análises de expressão transiente das construções Pol RFP:::AtbZIP76 e Pol RFP:::AtbZIP78 mostram que tanto AtbZIP76, quanto AtbZIP78 são direcionados ao núcleo. No entanto, experimentos complementares são necessários para confirmar os resultados e identificar as seqüências necessárias ao direcionamento nuclear.

Finalmente, a caracterização funcional de genes únicos (Tabela IX) complementa esta abordagem filogenética. Os genes mais interessantes para caracterização seriam aqueles que pertencem a Grupos para os quais há pouca ou nenhuma informação.

59

Referências

- ADAMS MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, WoodageT, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC. (2000). The genome sequence of Drosophila melanogaster. Science 287(5461): 2185-2195.
- ALBANI D, Hammond-Kosack MC, Smith C, Conlan S, Colot V, Holdsworth M, Bevan MW. (1997). The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. Plant Cell. 9(2): 171-184.
- ALTSCHUL S. F., Gish W., Miller W., Myers E. W. and Lipman D. (1990). Basic Local Alignment Tool. J. Mol. Bio. 215, 403-410.
- ALVAREZ-BUYLLA, ER; Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martinez-Castilla L, Yanofsky MF. (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. PNAS 97(10):5328-5333.
- Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408(6814):796-815.
- BAILEY, T.L. & Elkna, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the second International Conference on Intelligence Systems fo Molecular Biology, AAAI Press, Menlo Park, CA, pp 28-36.
- BECKETT, D. (2001). Regulated assembly os transcription factors and control of transcription initiation. J. Mol. Biol. (314): 335-352.
- BENNETZEN, J. (2002). Opening the door to comparative plant biology. Science (296): 60- 63.
- BENSMIHEN S, Rippa S, Lambert G, Jublot D, Pautot V, Granier F, Giraudat J, Parcy F. (2002). The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. Plant Cell. 14(6):1391-403.
- BLACKWOOD, E.M. & Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science (281): 61-63.
- CARLINI LE, Ketudat M, Parsons RL, Prabhakar S, Schmidt RJ, Guiltinan MJ. (1999). The maize EmBP-1 orthologue differentially regulates opaque2-dependent gene expression in yeast and cultured maize endosperm cells. Plant Mol. Biol. 41(3): 339-349.
- CAVALLI-SFORZA, L.L. & Edwards (1967). Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Genet. (19): 233-257.
- CICERI, P., Locatelli, F., Genga, A., Viotti, A. & Schmidt, R. J. (1999). The activity of the maize Opaque-2 transcriptional activation is regulated diurnally.
- CHIU R, Angel P, Karin M. (1989). Jun-B differs in its biological properties from, and is a negative regulator of, c-Jun. Cell 59(6): 979-986.
- CHOI, H., Hong, J., Ha, J., Kang, J. & Kim, S.Y. (2000). ABFs a family of ABA-responsive element binding factors. J. Biol. Chem., 275(3):1723-1730.
- CHUANG, C.F., Running, M.P., Williams, R.W.& Meyerowitz, E. (1999). The *Perianthia* gene encodes a bZIP protein involved in the determination of floral organ number in *Arabidopsis thaliana*. Genes and Development, 13: 334-344.

- DAYHOFF, M.O., Schwartz, R.C.; Orcutt, B.C. (1978). A model of evolutionary change in proteins. In: Atlas of protein sequence and structure (M.O. Dayhoff, ed.), pp. 301-310. National Biomedical Research Fundation, Silver Spring, MD.
- DESPRES, C., DeLong, C., Glaze, S., Liu, E. & Fobert, P.R. (2000). The Arabidopsis NPR1/NIM1 protein enhances the DNA binding activity of subgroup of the TGA family of bZIP transcription factors. Plant Cell 12(2): 179-81.
- ECK, R.V. & Dayhoff, M.O. (1967). Atlas of protein sequence and structures. National Biomedical Research Foudation, Silver Springs, MD.
- ELLENBERG, T.E., Brandl, C.J., Stuhl, K. & Harrison, S.C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α-helices: crystal structure of the protein-DNA complex. Cell, 71: 1223-1237.
- FASSLER J, Landsman D, Acharya A, Moll JR, Bonovich M, Vinson C. (2002). B-ZIP proteins encoded by the Drosophila genome: evaluation of potential dimerization partners. Genome Res. 12(8): 1190-1200.
- FELSENSTEIN, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. (17): 368-376.
- FELSENSTEIN J. (1993) PHYLIP (Phylogeny Inference Package) v 3.5c. Distributed by the author. Department of genetics, University of Washington, Seattle, WA. http://evolution.genetics.washington.edu/phylip.html.
- FELTKAMP D, Masterson R, Starke J, Rosahl S. (1994). Analysis of the involvement of ocs-like bZip-binding elements in the differential strength of the bidirectional mas1'2' promoter. Plant Physiol 105(1): 259-268.
- FINKELSTEIN RR. (1993). Abscisic acid-insensitive mutations provide evidence for stage-specific signal pathways regulating expression of an Arabidopsis late embryogenesis-abundant (lea) gene. Mol Gen Genet. 238(3):401-408.
- FINKELSTEIN RR. (1994). Maternal Effects Govern Variable Dominance of Two Abscisic Acid Response Mutations in Arabidopsis thaliana. Plant Physiol. 105(4):1203-1208.
- FINKELSTEIN, R.R. & Lynch, T.J.(2000). The Arabidopsis abscisic acid response gene *AB15* encodes a basic leucine zipper transcription factor. Plant Cell, 12(4): 599-609.
- FITCH, W.M. (2000). Homology: a personal view on some of the problems. TIG 16(5): 227-231.
- FUKASAWA, J., Sakai, T., Ishida, S., Yamaguchi, I., Kamiya, Y. & Takahashi, Y. (2000). REPRESSION OF SHOOT GROWT, a bZIP transcriptional activator, regulates all elongation by controlling the level of gibberellins. Plant Cell, 12: 901-915.
- GAI, X., Lal, S., Xing, L., Brendel, V. & Walbot, V. (2000). Gene discovery using the maize genome database ZmDB. Nucleic Acids Res., 28(1): 94-96.
- GAUBIER P, Raynal M, Hull G, Huestis GM, Grellet F, Arenas C, Pages M, Delseny M. (1993). Two different Em-like genes are expressed in Arabidopsis thaliana seeds during maturation. Mol Gen Genet. 238(3):409-418.
- GAUER, L. Caracterização Funcional do gene Bzo2h2 de *Arabidopsis thaliana*: um regulador da transcrição homólogo a Opaco2 do milho. Campinas: UNICAMP, 2004, 100p. Dissertação (doutorado em Genética e Biologia Molecular). Instituto de Biologia. Universidade Estadual de Campinas. 2004.
- GOFF, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Prus, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Felhaus, J., Malcama, T., Oliphant, A. & Briggs, S. (2002). A draft of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science, 296: 92-100.
- GRAUR, D. & Li, W.H. (2000). Fundamentals of Molecular Evolution. Sinauer Associates, Inc., Publishers. MA. EUA. 2^a edição.
- GREEN P. (1994). Phrap (http://www.washington.edu/UWGC/analysistools/phrap.htm).
- HENIKOFF, S., Greene, E.A., Pietrokovski, S., Bork, P., Attwood, T.K., Hood, L. (1997). Gene families: the taxonomy of protein paralogs and chimeras. Science (278): 609-614.
- HENNING, W. (1966). Phylogenetic systematics. University of Illinois Press, Urbana.
- HOLM M, Ma LG, Qu LJ, Deng XW. (2002) Two interacting bZIP proteins are direct targets of COP1mediated control of light-dependent gene expression in Arabidopsis. Genes Dev. 16(10):1247-59.
- HOLSTEGE, F.C.P., Young, R.A. (1999). Trancriptional regulation: Contending with complexity. Proc. Natl. Acad. Sci. USA 96:2-4.
- HURST, H.C. (1995). Leucine Zippers Transcription Factors. Academic Press, San Diego, USA.
- IZAWA T, Foster R, Chua NH. (1993). Plant bZIP protein DNA binding specificity. J Mol Biol. 230(4):1131-1144.
- JACQ C, Alt-Morbe J, Andre B, Arnold W, Bahr A, Ballesta JP, Bargues M, Baron L, Becker A, Biteau N, Blocker H, Blugeon C, Boskovic J, Brandt P, Bruckner M, Buitrago MJ, Coster F, Delaveau T, del Rey F,

Dujon B, Eide LG, Garcia-Cantalejo JM, Goffeau A, Gomez-Peris A, Zaccaria P, et al. (1997). The nucleotide sequence of Saccharomyces cerevisiae chromosome IV. Nature 387(6632 Suppl):75-78.

- JAKOBY M, Weisshaar B, Droge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F. (2002). bZIP transcription factors in Arabidopsis. Trends Plant Sci. 7(3):106-111.
- KAWATA T, Minami M, Tamura T, Sumita K, Iwabuchi M. (1992). Isolation and characterization of a cDNA clone encoding the TATA box-binding protein (TFIID) from wheat. Plant Mol Biol. 19(5): 867-872
- KELLOGG EA. (2004). Evolution of developmental traits. Curr. Opin. Plant Biol. 7(1):92-98
- KHUSH, G.S. (1997). Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology (35): 25-34.
- KÖNIG, P. & Richmond, T.J. (1993). The X-ray structure of the CGN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexbility. J. Mol. Biol., 233: 139-154.
- KORNBERG, R.D. (1999). Eukaryotic transcriptional control. Trend Cell Biology, 12: 46-49.
- KUHLEMEIER, C. (1992). Transcriptional and post-transcriptional regulation of gene expression in plants. Plant Mol. Biol. (19): 1-14.
- KELLIS, M.; Birren, B.W.; Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisae*. Nature (428): 617-624.
- LAWTON-RAUH, A. (2003). Evolutionary dynamics of duplicated genes in plants. Mol Phylogenet Evol. 29(3): 396-409.
- LEE, T.I. & Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. (34): 77-137.
- LESPINET O, Wolf YI, Koonin EV, Aravind L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12(17):1048-1059.
- LIJAVETZKY, D., Carbonero, P., Carbajosa, J. (2003). Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. BMC Evolutionary Biology 3 (1): 17.
- LUSCOMBE NM, Austin SE, Berman HM, Thornton JM.(2000). An overview of the structures of protein-DNA complexes. Genome Biol 1(1): REVIEWS001.
- LYNCH M. (2002). Genomics. Gene duplication and evolution. Science 297(5583):945-947.
- MARTINEZ-CASTILLA LP, Alvarez-Buylla ER.(2003). Adaptive evolution in the Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny. PNAS 100(23):13407-13412
- MESHI, T. & Iwabuchi, M. (1995). Plant transcription factors. Plant Cell Physiol. 36(8): 1405-1420.
- Meyerowitz, E.M. (2002). Plants compared to animals: the broadest comparative study of development. Science (295): 1482 1485.
- MIKAMI K, Sakamoto A, Iwabuchi M.(1994). The HBP-1 family of wheat basic/leucine zipper proteins interacts with overlapping cis-acting hexamer motifs of plant histone genes. J. Biol. Chem 269(13): 9974-9985.
- MOORE, R.C. & Purugganan, M.D. (2003). The early stages of duplicate gene evolution. PNAS 100(26): 15682-15687.
- MURASHIGE, T. & Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol. Plant, 15: 473-507.
- NAM J, Kim J, Lee S, An G, Ma H, Nei M. (2004). Type I MADS-box genes have experienced faster birth-anddeath evolution than type II MADS-box genes in angiosperms. PNAS 101(7):1910-1915.
- NEI, M., Kumar, S., Takahashi, K. (1998). The optimization principle in phylogenetic analysis tend to give incorrect topologies when the number of nucleotides or aminoacids is small. PNAS (95): 12390-12397.
- NEI, M. & Kumar, S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press.
- NIGGEWED, R., Thurow, C., Kegler, C. & Gatz, C. (2000). Tabacco transcritpion factor TGA2.2 is the main component of as-1 binding factor ASF-1 and is involved in salicylic acid and auxin inducible expression of as-1-containing target promoters. J. Miol. Chem., 275: 19897-19905.
- OÑATE L, Vicente-Carbajosa J, Lara P, Diaz I, Carbonero P. (1999). Barley BLZ2, a seed-specific bZIP protein that interacts with BLZ1 in vivo and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm. J Biol Chem. 274(14): 9175-9182.
- ONODERA Y, Suzuki A, Wu CY, Washida H, Takaiwa F. (2001). A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. J Biol Chem. 276(17): 14139-14152.
- OSTERLUND, M.T., Wei, N. & Deng, X.W. (2000). The roles of photorreceptor system and the COP-1 targeted distabilization of HY5 in the light control of Arabidopsis seedling development. Plant Physiol., 124: 1520-1524.
- PEETERS NM, Chapron A, Giritch A, Grandjean O, Lancelin D, Lhomme T, Vivrel A, Small I. (2000). Duplication and quadruplication of Arabidopsis thaliana cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. J Mol Evol. 50(5):413-423.
- PENNACHIO, L.A (2003) Insights from human/mouse genome comparisons. Mammalian Genome. (14): 429-436.

- PERTEA G., Huang X., Liang F., Antonescu V., Sultana R., Karamycheva S., Lee Y., White J., Cheung F., Parvizi B., Tsai J., Quackenbush J. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19(5): 651-652.
- PIROVANO L, Lanzini S, Hartings H, Lazzaroni N, Rossi V, Joshi R, Thompson RD, Salamini F, Motto M. (1994). Structural and functional analysis of an Opaque-2-related gene from sorghum. Plant Mol Biol. 24(1): 515-523.
- PONTIER, D., Miao, Z-H & Lam, E. (2001). Trans-dominant suppression of plant TGA factors reveals their negative and positive roles in plant defense responses. The Plant Journal 27(6): 529-538.
- QUAKENBUSH, J., Liang, F., Holt, I., Pertea, G. & Upton, J. (2000). The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res., 28(1): 141-145.
- QUACKENBUSH, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamancheva, S., Parvizi, B., Pertece, G., Sultana, R. & White, J. (2001). The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res. 29(1): 159-164.
- RIECHMANN, J. L., Heard, J., Martin, J, Reuber, F., Jiang, C.Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghanderhari, D., Sherman, B.K. & Yu, G.L. (2000). Arabidopsis transcription factors: genome-wide camparative analysis among eukaryontes. Science, 290: 2105-2110.
- RIECHMANN, J.L. & Ratcliffe, O.J. (2000). A genomic perspective on plant transcription factors. Curr. Opn. Plant Biol. 3 (5): 423-434.
- ROOK, F., Weisbeck, P. & Smeekens, S. (1998). The light-regulated Arabidopsis bZIP transcription factor gene ATB2 encodes a protein with na unusally lon leucine zipper domain. Plant Mol. Biol., 37: 171-178.
- RUDD S. (2003). Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci. 8(7):321-9.
- SAITOU, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. (24): 184-204.
- SATOH R, Fujita Y, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K. (2004). A Novel Subgroup of bZIP Proteins Functions as Transcriptional Activators in Hypoosmolarity-Responsive Expression of the ProDH Gene in Arabidopsis. Plant Cell Physiol. 45(3):309-317.
- SAMBROOK, J., Fritsch, E.F., Maniats, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd edition. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY, USA.
- SASAKI, T. & Burr, B. (2000). International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr. Opn. Plant Biol. (3):138-141.
- SAVOLAINEN, V. & Chase, M.W. (2003). A decade of progress in plant molecular phylogenetics. Trends in Genetics 19(12): 717-724.
- SCHMIDT RJ, Burr FA, Aukerman MJ, Burr B. (1990). Maize regulatory gene opaque-2 encodes a protein with a "leucine-zipper" motif that binds to zein DNA. PNAS 87(1): 46-50.
- SCHINDLER, U., Menkens, A.E., Beckman, H., Ecker, J.R. & Cashmore, A.R. (1992). Heterodimerization between light-regulated and ubiquitously expressed Arabidopsis GBF bZIP proteins. EMBO J., 11: 1261-1273.
- SHIU SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH. (2004). Comparative analysis of the receptorlike kinase family in Arabidopsis and rice. Plant Cell 16(5):1220-1234.
- SIGH, K.B. (1998). Transcriptional Regulation in plants: the importance of combinatorial control. Plant Physiol., 118: 1111-1120.
- SOLTIS, D.E. & Soltis, P.S. (2003). The role of phylogenetics in comparative genetics. Plant Physiol. (132): 1790-1800.
- SONG, R., Llaca, V. & Messing, J. (2002). Mosaic Organization of Orthologous Sequences in Grass Genomes. Genome Research: GR-26-83R.
- SOSINSKY A, Glusman G, Lancet D. (2000). The genomic structure of human olfactory receptor genes. Genomics 70(1):49-61.
- TATUSOV, R.L., Koonin, E.V. & Lipman, D.J. (1997). A genomic perspective on proteins families. Science, 278: 631-637.
- The *C. elegans* sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282(5396):2012-2018.
- The Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project. (2003). Annual Report 2003. www.arabidopsis.org.
- The Rice Full-Length cDNA Consortium (2003). Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from *japonica* Rice . Science (18) : 376-379.
- THOMPSON JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25(24):4876-82.
- THORNTON, J. & DeSalle, R. (2000) Gene family evolution and homology: genomics meets phylogenetics. Annu. Rev. Genomics Hum. Genet., 1:41-73.

- TUCKER ML, Whitelaw CA, Lyssenko NN, Nath P. (2002). Functional analysis of regulatory elements in the gene promoter for an abscission-specific cellulase from bean and isolation, expression, and binding affinity of three TGA-type basic leucine zipper transcription factors. Plant Physiol. 130(3): 1487-1496.
- TUPLER R, Perini G, Green MR. (2001). Expressing the human genome. Nature 42(4): 210-214.
- ULM R, Baumann A, Oravecz A, Mate Z, Adam E, Oakeley EJ, Schafer E, Nagy F. (2004). Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of Arabidopsis. PNAS 101(5):1397-1402.
- UNO, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. & Yamagushi-Shinozaki, K. (2000). Arabidopsis basic leucine zipper transcription factors involved in as abscisic acid-dependent signal transduction pathway under drought and high salinity conditions. PNAS USA, 97: 11632-11637.
- VANDEPOELE, K., Simillion C., Van de Peer, Y. (2003). Evidence that rice and other cereals are ancient aneuploids. Plant Cell. 15(9):2192-2202.
- VARAGONA M.J., Raikhel N.V. (1994). The basic domain in the bZIP regulatory protein Opaque2 serves two independent functions: DNA binding and nuclear localization. Plant J. 26(9): 437-441.
- VETTORE, A.L., Yunes, J.A., Neto, G.C., da Silva, M.J., Arruda, P. & Leite, A. (1998). The molecular and functional characterization of na *Opaque2* homologue gene from *Coix* and a new classification of plant bZIP proteins. Plant Molec. Biol., 36: 249-263.
- VINCENTZ, M., Schlögl, P., Corrêa, L.G.G, Khüne, F., Leite, A. (2001). Phylogenetic relationships between Arabidopsis and sugarcane bZIP transcriptional regulatory factors. Genetics and Molecular Biology, 24: 55-60.
- VINCENTZ, M., Bandeira-Kobarg, C., Gauer, L., Schlögl, P. & Leite, A. (2003). Evolutionary pattern of Angioperm bZIP factors homologous to the maize Opaque2 regulatory protein. Journal of Molecular Evolution 55: 1–12.
- VINCENTZ M., Cara F.A., Okura V.K., da Silva F.R., Pedrosa G.L., Hemerly A.S., Capella A.N., Marins M., Ferreira P.C., Franca S.C., Grivet L., Vettore A.L., Kemper E.L., Burnquist W.L., Targon M.L., Siqueira W.J., Kuramae E.E., Marino C.L., Camargo L.E., Carrer H., Coutinho L.L., Furlan L.R., Lemos M.V., Nunes L.R., Gomes S.L., Santelli R.V., Goldman M.H., Bacci M. Jr., Giglioti E.A., Thiemann O.H., Silva F.H., Van Sluys M.A., Nobrega F.G., Arruda P., Menck C.F. (2004). Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. Plant Physiol. 134(3): 951-959.
- WALSH, J.W. & Feeling, M. (1997). The mayze gene liguleles 2 encodes a basic leucine zipper envolved in the stabilishment of the leaf blade-sheath boundary. Genes and development, 11: 208-218.
- WARREN, A.L. (2002). Eukaryotic transcription factors. Curr. Opn. In Struct. Biol. (2002): 107-114.
- WELLMER, F., Kircher, S., Rügner, A., Frohnmeyer, H., Schäfer, E. & Harter, K. (1999). Phosphorylation of the parsley bZIP transcription factor CPRF-2 is regulated by light. J. Biol. Chem., 274(41): 29476-29489.
- WENDEL J.F. (2000). Genome evolution in polyploids. Plant Mol Biol 42(1):225-249.
- WINGENDER, E., Chen, X, Hehl, R., Karas, H., Liebich, I. & Matys, V. (2000). TRANSFAC: na integrated system for gene expression regulation. Nucleic Acids Res., 28: 316-319.
- WRAY G.A., Hahn M.W., Abouheif E., Balhoff J.P., Pizer M., Rockman M.V., Romano L.A. (2003). The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol. 20(9): 1377 1419.
- YAMADA K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR. (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome. Science 302(5646): 842-846
- YIN, Y., Zhu, Q., Da, S., Lam, C. & Beachy, R. (1997). RF2a, a bZIP transcriptional activator of the phloemspecific rice tungro baciliform virus promoter, function in vascular development. EMBO J., 16: 5247-5259.
- YU J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H.(2002). A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296(5565):79-92
- YUAN, Q., Quackenbush, J., Sultana, R., Pertea, M., Slazberg, S.L. & Buell, C.R. (2000). Rice Bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species. Plant Physiol., 125:1166-1174.

- ZHANG, Y., Fan, W., Kinkeman, M., Li, X. & Dong, X. (1999). Interaction of NPR1 with basic leucine zipper protein transcription fator that bind sequences required for salicylic acid induction of the PR-1 gene. PNAS USA, 96: 6523-6528.
- ZHANG Y, Tessaro MJ, Lassner M, Li X. (2003). Knockout analysis of Arabidopsis transcription factors TGA2, TGA5, and TGA6 reveals their redundant and essential roles in systemic acquired resistance. Plant Cell 15(11):2647-2653.