

# SISTEMA DE INFORMAÇÃO CIS-ELEMENTOS

MARCELO GONÇALVES NARCISO<sup>1</sup>, ROSÂNGELA BEVITORI<sup>2</sup>

**INTRODUÇÃO:** A resposta da planta aos estresses bióticos e abióticos tem sido alvo de ações de pesquisa de interesse estratégico. O estudo do transcrito de uma determinada espécie vegetal, associado a estes estresses, encontra uma poderosa ferramenta nas novas tecnologias de sequenciamento que possibilitam varredura massal do genoma a uma velocidade, precisão e escala sem precedentes (MARDIS, 2008) gerando um enorme acúmulo de informações. Estas são analisadas como um todo para auxiliar a interpretação biológica dos dados e identificar possíveis genes candidatos associados aos estresses estudados (MAGALHÃES et al., 2010). Torna-se, portanto, essencial uma análise criteriosa dessas informações visando um estudo mais acurado dos genes candidatos a serem identificados. Neste sentido, o conhecimento de que cis-elementos são essenciais para a regulação transcricional do gene, controlando muitos processos biológicos e respostas a estresses (IBRAHEEM et al., 2010), pode ser utilizado para complementar a identificação de genes candidatos responsivos a determinadas condições. Portanto, o objetivo deste estudo foi desenvolver um sistema computacional capaz de identificar a presença de cis-elementos conhecidos, utilizando como modelo a família de genes *ERDs* (Early Responsive to Dehydration) expressos por *Arabidopsis thaliana* durante estresse hídrico (TAJI et al., 1999). Este sistema computacional também pode ser usado para outras culturas, como o feijão, arroz, trigo, etc.

**MATERIAL E MÉTODOS:** Como a maioria da porção regulatória dos genes de plantas está localizada primariamente a 1000 pb *upstream* do sítio de início da transcrição gene (DEAN; SCHMIDT, 1995), optou-se por utilizar 1500 nucleotídeos das seqüências de cada gene *ERD* em *A. thaliana*. De posse do número de acesso de cada gene, as seqüências da região codante (CDS) foram obtidas a partir da base de dados do National Center for Biotechnology Information (ver [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Individualmente, as seqüências foram confirmadas submetendo as mesmas contra o genoma de *A. thaliana* usando basic Local Alignment Tool (BLASTN, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Um programa em linguagem C foi criado para obter-se a região *upstream* de cada gene a partir do cromossoma correspondente. A identificação de prováveis cis-elementos foi conduzida usando o software PLACE (<http://www.dna.affrc.go.jp/PLACE/signalscan.html>). Um script foi criado para inserir os arquivos gerados pelo software Place numa base de dados criada para armazenar e permitir consulta dos dados. Este programa foi feito em Shell script Bash, executado em ambiente Unix/Linux. A base de dados foi modelada e implementada utilizando o sistema gerenciador de banco de dados MySQL (ver [www.mysql.org](http://www.mysql.org)), que é um software livre. A interface de visualização e consulta dos dados foi criada de modo a atender as necessidades do usuário. Para isto, páginas HTML foram implementadas para acesso a opções de consulta. Para visualização dos resultados foram feitos programas em JSP (Java Server Page) e Javascript de modo a realizar a busca da informação solicitada no banco de dados sobre cis-elementos. O banco de dados foi feito em MySQL e a linguagem de consulta aos dados é SQL (Sequel Query Language). A linguagem JSP é usada para construir programas que são executados no servidor web. Estes programas acessam a base de dados sobre cis-elementos para realizar consultas pedidas pelo usuário e montam o resultado para ser exibido em formato HTML ou CSV (comma separated values) no computador do usuário. O formato CSV é facilmente lido por uma planilha (Excel ou Calc) e assim pode ser usado para outros fins de análise. A linguagem Javascript é usada para declarar e implementar funções na página HTML na qual os dados são preenchidos em um

<sup>1</sup> Pesquisador da Embrapa Arroz e Feijão. Rodovia GO-462, km 12 Zona Rural C.P. 179, Santo Antônio de Goiás, GO. E-mail: narciso@cnpaf.embrapa.br

<sup>2</sup> Pesquisadora da Embrapa Arroz e Feijão. Rodovia GO-462, km 12 Zona Rural C.P. 179, Santo Antônio de Goiás, GO. E-mail: bevitori@cnpaf.embrapa.br

formulário. Após inseridos os dados, são executadas as funções na máquina do usuário para verificação do correto preenchimento do formulário de entrada de dados. Caso algum dado esteja errado, o programa Javascript envia um alerta ao usuário. Se os dados estiverem preenchidos corretamente, então estes são enviados ao servidor para servirem de entrada para os programas feitos em JSP. Assim, uma parte do processamento é feita no computador do usuário e outra parte do processamento é feita no servidor web.

**RESULTADOS E DISCUSSÃO:** Foram utilizados os genes ERDs de *A. thaliana* para simular a entrada de dados no sistema computacional desenvolvido. Quando as seqüências de 1500 nucleotídeos de cada gene foram submetidas ao PLACE, uma página HTML foi obtida contendo os cis-elementos encontrados em cada gene, com sua respectiva seqüência e um link cuja ativação era necessária para se ter conhecimento da predita função. Assim, para saber a função de cada cis-elemento tinha como função, era necessário acessar todos os links. Esta tarefa, além de laboriosa e tediosa, despendia muito tempo. Para contornar esta situação, um programa foi criado usando linguagem Shell script Bash de modo a permitir o carregamento de todos os arquivos referentes a cada link. Desta forma, o link que indicava função foi transformado em informação diretamente disponível para visualização e consulta na base de dados, tornando a informação disponível rapidamente e de forma independente do link. A base de dados é constituída de um depósito das seqüências dos genes ERDs com respectiva posição dos cis-elementos, funções preditas, links para função completa, etc.. Isto proporciona ao usuário a busca por palavras chaves, nome dos genes, seqüência do cis-elementos dentre outros. Ao usar o sistema foram identificados vários cis-elementos que estão associados a diversos estresses ou desenvolvimento de plantas em todos os genes. Para exemplificar, quando se realizou a busca pela palavra chave *drought* (seca) (Figura 1), 100% dos genes continham cis-elementos relacionados a esta característica. Este resultado nos indica que a análise de cis-elementos na sequencia de genes regulatórios expressos pode ser utilizado para auxiliar a identificar genes candidatos associados a determinados estresses.

Sequencia Cis	CisActing	Variedade	ERD	Descrição - Link
CATGTG	MYC	Arabidopsis Thaliana	ERD1	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413</a>
CATGTG	MYC	Arabidopsis Thaliana	ERD6	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413</a>
CATGTG	MYC	Arabidopsis Thaliana	ERD10	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413</a>
CATGTG	MYC	Arabidopsis Thaliana	ERD16	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413</a>
CATGTG	MYC	Arabidopsis Thaliana	ERD15	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413</a>
CATGTG	MYC	Arabidopsis Thaliana	ERD13	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413</a>
CATGTG	MYC	Arabidopsis Thaliana	ERD4	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000413</a>
CACATG	MYC	Arabidopsis Thaliana	ERD1	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174</a>
CACATG	MYC	Arabidopsis Thaliana	ERD6	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174</a>
CACATG	MYC	Arabidopsis Thaliana	ERD10	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174</a>
CACATG	MYC	Arabidopsis Thaliana	ERD16	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174</a>
CACATG	MYC	Arabidopsis Thaliana	ERD15	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174</a>
CACATG	MYC	Arabidopsis Thaliana	ERD13	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174</a>
CACATG	MYC	Arabidopsis Thaliana	ERD4	<a href="http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174">http://www.dna.affrc.go.jp/sigscan/disp.cgi?S000174</a>

**Figura 1.** Resultado de uma busca por palavras (*drought*).

A busca ilustrada na Figura 1 mostra uma série de ERDs e também as variedades e dados sobre cis-elementos (sequência e cis Acting). Para fazer uma busca dessa sem ter um sistema, seria necessário obter todos os genes relativos a ERD, obter a sequência upstream e com esta gerar um resultado que conteria todos os links. Com todos estes links, seriam feitas consultas em todas as páginas para então

saber qual tem alguma característica ligada à seca (*drought*). Outros resultados que o usuário pode ter é quanto a consultas por cis-elementos conhecidos e funções. Um exemplo de saída, em HTML, seria tal como ilustra a Figura 2. Outras consultas são com respeito a busca por cis-elementos comuns entre ERDs (o que os ERDs cadastrados no sistema têm em comum) e busca por cis-elementos por ERDs (o que cada ERDs tem com respeito a cis-elementos). Com estes resultados fica mais fácil visualizar a participação de cada cis-elementos em um dado conjunto de ERDs. A interface web e o banco de dados vazios podem ser disponibilizados ao público, bem com todo o procedimento de captura de dados e sua inserção no banco de dados para consulta pelo sistema.

Nome do ERDS	CisActing	Espécie	Sequência Cis	Site com a descrição completa	Função
ERD1	-10PEHVPSBD	Arabidopsis Thaliana	TATTCT	<a href="#">S000392</a>	``-10 promoter element`` found in the barley (H.v.) chloroplast psbD gene promoter  Involved in the expression of the plastid gene psbD which encodes a photosystem II reaction center chlorophyll-bind
ERD1	14BPATERD1	Arabidopsis Thaliana	CACTAAATTGTCCAC	<a href="#">S000412</a>	``14 bp region`` (from -599 to -566) necessary for expression of erd1 (early responsive to dehydration) in dehydrated Arabidopsis
ERD1	2SSEEDPROTBANAPA	Arabidopsis Thaliana	CAAACAC	<a href="#">S000143</a>	Conserved in many storage-protein gene promoters  May be important for high activity of the napA promoter

**Figura 2.** Trecho de um resultado de uma consulta por Cis-Elementos e sua função.

**CONCLUSÕES:** Com a interface web para acesso aos dados diversos sobre cis-elementos, é possível obter uma série de conclusões de forma rápida e ainda obter estes resultados para outras análises. Para isso, um trabalho computacional foi feito, de forma que qualquer dado sobre ERDs e cis-elementos de qualquer cultura possa ser inserido na base e consultado facilmente. Embora os exemplos acima se referem à *Arabidopsis Thaliana*, este sistema vale para qualquer cultura. Assim, basta inserir os dados relativos à cultura de feijão que o sistema irá fornecer todos os resultados sobre cis-elementos tal como foi demonstrado para a *Arabidopsis Thaliana*.

## REFERÊNCIAS

DEAN, C., SCHIMIDT, R. Plant genomes: a current molecular description. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v.46, n.xxx, p. 395-418, 1995.

IBRAHEEM, O.; BOTHA, C. E. J.; BRADLEY, G. In silico analysis of cis-acting regulatory elements in 5<sub>′</sub> regulatory regions of sucrose transporter gene families in rice (*Oryza sativa* Japonica) and *Arabidopsis thaliana*. **Computational Biology and Chemistry**, Oxford, v. 34, n. 5/6, p. 268–283, 2010.

MAGALHÃES, J. P. de; FINCH, C. E.; JANSSENS, G. Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. **Ageing Research Reviews**, Kidlington, v. 9, n. 3, p. 315–323, 2010.

MARDIS, E. R. The impact of next-generation sequencing technology on genetics. **Trends in Genetics**, Amsterdam, v. 24, n. 3, p. 133-141, 2008.

TAJI, M.; SEKI, M.; YAMAGUCHI-SHINOZAKI, K.; KAMADA, H.; GIRAUDAT, J.; SHINOZAKI, K. **Mapping of 25 drought-inducible genes, RD and ERD, in *Arabidopsis thaliana*.** *Plant and Cell Physiology*, Kyoto, v. 40, n. 1, p. 119-123, 1999