

IDENTIFICAÇÃO *IN SILICO* DE GENES DA FAMÍLIA *IDD* EM CANA DE AÇÚCAR

CARLA PRISCILA COELHO¹, ANTÔNIO CHALFUN JUNIOR²; JOSEPH COLASANTI³,
SOLANGE APARECIDA SÁGIO⁴

RESUMO

O gene *ID1* é um gene do florescimento caracterizado em milho, pertencente à família do domínio ID. O objetivo deste trabalho foi identificar os genes *ID* em cana de açúcar pelo banco de dados SUCEST. Primeiramente, uma busca por *reads* foi realizada na plataforma do *Gene Project* e posteriormente, estes *reads* foram montados em EST-contigs, e então, analisados como possíveis sequências ID de cana de açúcar. Após uma anotação manual, as sequências foram analisadas para verificar a presença do domínio conservado da família ID, envolvendo a região NLS e os domínios *zinc finger*. Cinco contigs e quatro singlets foram selecionados para os estudos posteriores, considerados como possíveis sequências *ID* de cana de açúcar. O alinhamento foi realizado comparando-se as possíveis sequências *ID* de cana de açúcar, gene *ID1* de milho, e *ID1-like* de arroz e sorgo. A análise filogenética mostrou que o Singlet 1 é a sequência mais próxima do gene *ID1* de milho, arroz e sorgo, contudo ela não apresenta a região envolvendo o NLS. Portanto, pode-se concluir que todas as sequências encontradas podem constituir a família gênica *IDD* em cana de açúcar, contudo as sequências inteiras necessitam ser identificadas.

Palavras-chaves: bioinformática, genes *IDD*, análise filogenética, metabolismo da cana de açúcar

INTRODUÇÃO

Uma grande escala de ESTs (Expressed Sequence Tags) tem sido utilizada para o acesso rápido a informações de sequências. ESTs de mais de 1000 espécies compreendem o banco de dados de (dbEST) no *National Center for Biotechnology Information* (NCBI). As plantas representam 18% de todos os ESTs no dbEST, sendo que metade das sequências são de arroz, trigo, milho, cana de açúcar, sorgo e cevada. Essa quantidade de informação de transcritos em cereais, especialmente de cana de açúcar, sorgo e milho, proporciona uma ótima oportunidade para comparar os perfis de expressão entre essas espécies. Características associadas a metabolismo, desenvolvimento, e respostas a estresses bióticos e/ou abióticos podem ser acessadas por análises comparativas de expressão gênica entre espécies relacionadas (VETTORE et al., 2001). O projeto de ESTs de cana de açúcar (FIGUEIREDO et al., 2001), SUCEST, envolve uma coleção de 240.000 de ESTs geradas de 26 bibliotecas de cDNA construídas de diferentes órgãos e/ou tecidos em diferentes estádios de desenvolvimento (meristema apical, flores, folhas imaturas, sementes, entrenós, casca do caule, folhas estioladas, botões laterais, etc.) (VETTORE et al., 2001). Essas ESTs foram organizadas em contigs, nomeadas como *sugarcane assembled sequences* (SAS), o que permitiu encontrar mais de 30.000 genes de cana de açúcar (VETTORE et al., 2003). Uma comparação do transcriptoma de cana de açúcar com outras monocotiledôneas e dicotiledôneas mostrou que os genes existentes em ambos os grupos são altamente conservados, apesar do longo período de evolução independente (VINCENTZ et al., 2004). Análises no banco de dados do SUCEST revelaram a presença de genes codificando uma ampla gama de funções associadas a características agrônômicas importantes (VETTORE et al.,

¹ Doutoranda em Fisiologia Vegetal, DBI/UFLA, carllapc@posgrad.ufla.br

² Professor Adjunto, DBI/UFLA, chalfunjunior@dbi.ufla.br

³ Professor Assistente, Department of Molecular and Cellular Biology/University of Guelph, jcolasan@uoguelph.ca

⁴ Doutoranda em Fisiologia Vegetal, DBI/UFLA, solsagio@gmail.com.br

2003). Para identificar a família gênica IDD em cana de açúcar, neste estudo foi realizada uma análise *in silico* comparando as sequências encontradas no SUCEST com sequências já conhecidas do gene *IDI* de milho, arroz e sorgo.

MATERIAL E MÉTODOS

Busca por EST-contigs no Genbank

Uma busca por sequências IDD em espécies relacionadas foi feita por palavra-chave no banco de dados do NCBI e as sequências obtidas foram anotadas. Adicionalmente, uma busca utilizando-se o BLAST (*Basic Local Alignment Search Tool*) no mesmo banco de dados foi realizada, utilizando-se a sequência do gene *IDI* de milho como *query*. As sequências que apresentaram um alinhamento significativo ($e\text{-value} < 10^{-5}$) foram anotadas e utilizadas para comparar as sequências depositadas no SUCEST.

Comparação de sequências encontradas no Genbank com sequências depositadas na base de dados do SUCEST

As sequências encontradas no NCBI foram confrontadas utilizando-se o *blastn* no banco de dados do *The Gene Index Project*, onde os ESTs de cana de açúcar estão depositados. As sequências foram agrupadas em EST-contigs e anotadas utilizando-se o software *CLC Main WorkBench 5.1* (CLC bio®). Após a montagem das sequências, elas foram analisadas para a seleção daquelas contendo domínio ID (COLASANTI et al., 2006). Apenas as sequências contendo o domínio conservado foram selecionadas para as análises subsequentes.

Análise filogenética de possíveis sequências IDD com espécies relacionadas

O alinhamento dos EST-contigs foi feito com o software *CLC Main Workbench 5.6.1* seguindo os parâmetros default, utilizando-se as sequências traduzidas. Uma árvore filogenética foi construída pelo mesmo software. A distância filogenética dos *clusters* foi medida por *bootstraps*.

RESULTADOS E DISCUSSÃO

Sequências apresentando o domínio ID (sequências *IDD*) foram identificadas pelo banco de dados do SUCEST. Vinte *reads* foram encontrados, utilizando o gene *IDI* como *query*. A clusterização gerou nove EST-contigs. A sequência protéica de cada gene IDD putativo foi analisada para encontrar o domínio conservado e motivos. Apesar de algumas sequências selecionadas não apresentarem todos os quatro *zinc fingers*, as sequências que apresentavam, pelo menos, um dos *zinc fingers*, foi mantido. Um processo de saturação foi realizado com o objetivo de saturar a busca e possivelmente conseguir o gene inteiro, porém, a maioria das sequências ainda estão incompletas. Isto pode estar relacionado ao fato de, apesar do SUCEST compreender uma grande quantidade de ESTs depositados, os clone de cDNA utilizados para criar as bibliotecas foram parcialmente sequenciados (VETTORE et al., 2003).

XIX CONGRESSO DE PÓS-GRADUAÇÃO DA UFPA
 27 de setembro a 01 de outubro de 2010

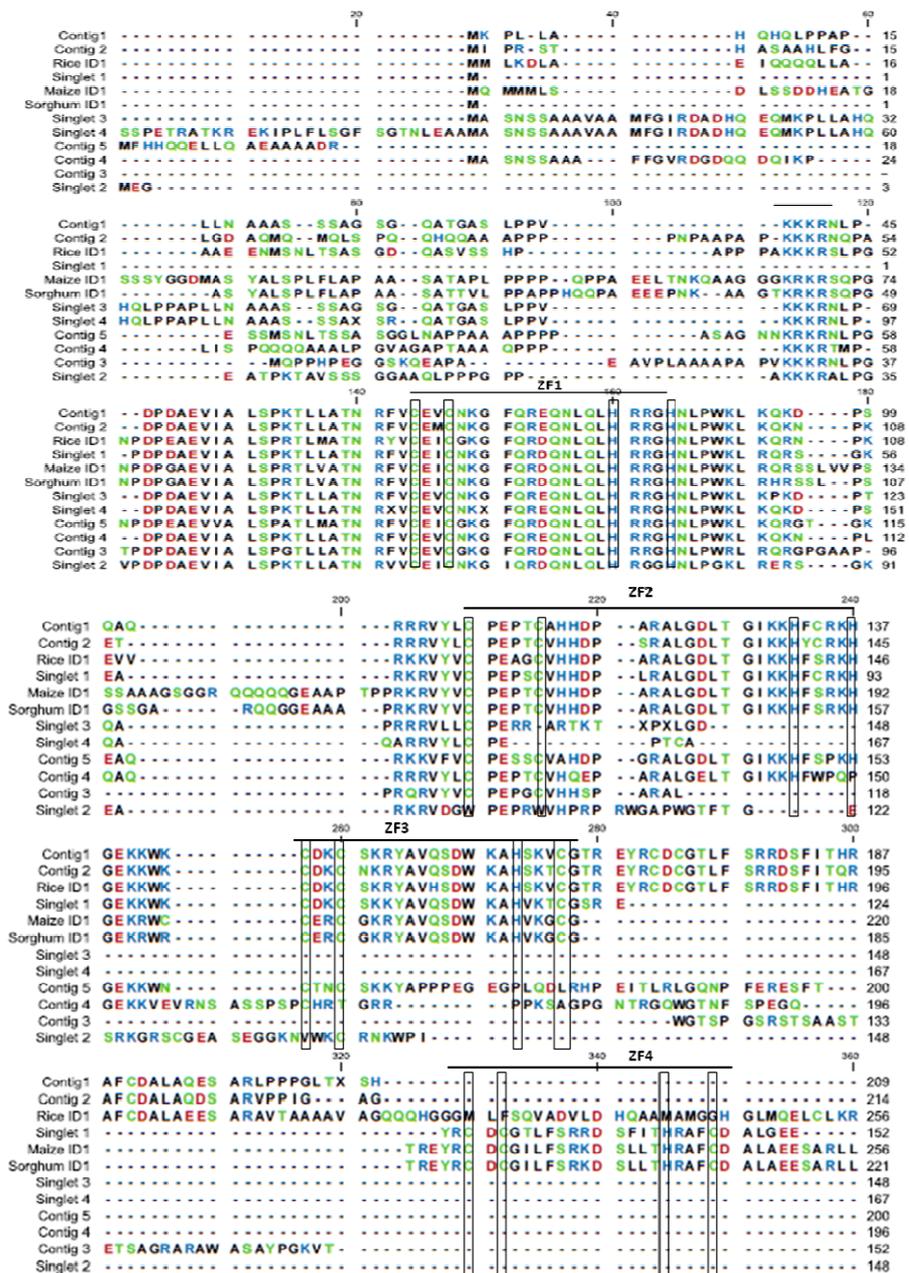


Figura 1- Alinhamento de EST-contigs de cana de açúcar com o gene *ID1* de milho, arroz e sorgo encontradas no banco de dados do NCBI. Aminoácidos compartilhados por pelo menos duas seqüências são mostrados na mesma cor; a região putativa NLS está destacada por uma barra preta. Cada *zinc finger* está indicado como ZF1, ZF2, ZF3 e ZF4, e as regiões dentro das caixas mostram os resíduos de C e H.

Uma comparação filogenética do gene *ID1* de milho, sorgo e arroz com seqüências *ID-like* de cana de açúcar (Figure 2) mostrou que o gene *ID1* das três espécies são relacionadas entre si, corroborando com os resultados previamente publicados por Colasanti et al. (2006).

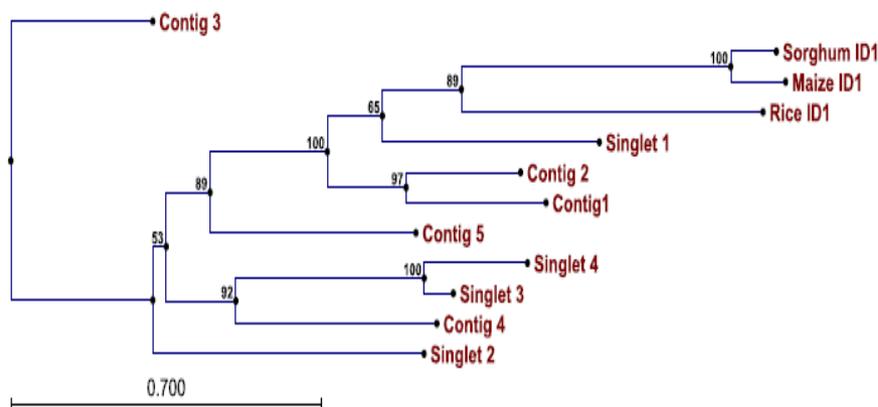


Figura 2 – Comparação filogenética de possíveis *ScID* com *ID1*, *SbID* e *OsID* (Número de acesso para *ID1*, *SbID* e *OsID* são 162461280, 242039639 e 55773665, respectivamente).

O Contig 1 e 2 são as maiores sequências encontradas, envolvendo todos os motivos IDD. A alta similaridade de genes em diferentes espécies, principalmente aquelas envolvendo domínios conservados, indica que a função das sequências pode ser conservada, sendo estes genes chamados de ortólogos. De acordo com isto, as sequências IDD encontradas podem ter as funções relacionadas aos genes *ZmIDD*, *AtIDD* e *OsIDD*.

CONCLUSÃO

Após as análises *in silico* no banco de dados SUCEST foi possível identificar nove candidatas da família gênica IDD. Tal fato permitiu a identificação de motivos e domínios conservados de possíveis sequências IDD encontradas no SUCEST, indicando a possibilidade de serem membros da família IDD de cana de açúcar.

REFERENCIAL BIBLIOGRÁFICO

COLASANTI, J. et al. The maize *Indeterminate1* flowering time regulator defines a highly conserved zinc finger protein family in higher plants. **BMC Genomics**, London, v. 7, p. 158-175, June 2006.

FIGUEIREDO, R. C. et al. Dissecting the sugarcane expressed sequence tag (SUCEST) database: unravelling flower-specific genes. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, n. 1/4, p. 77-84, 2001.

VETTORE, A. L. et al. The libraries that made Sucest. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, n. 1/4, p. 1-7, 2001.

VETTORE, A. L. et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, Cold Spring Harbor, v. 13, p. 2725-2735, Nov. 2003.

VINCENTZ, M. et al. Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. **Plant Physiology**, Rockville, v. 134, p. 951-959, Mar. 2004.