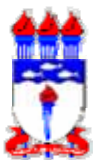


JANICE SILVA SALES

ANÁLISE *IN SILICO* EM ESTs DE CANA-DE-AÇÚCAR (*Saccharum officinarum*) RELACIONADOS À RESISTÊNCIA AO ESTRESSE HÍDRICO



UFAL

**UNIVERSIDADE FEDERAL DE ALAGOAS
CENTRO DE CIÊNCIAS AGRÁRIAS
CURSO DE MESTRADO EM AGRONOMIA
RIO LARGO, ESTADO DE ALAGOAS**



JULHO DE 2006



UFAL

**UNIVERSIDADE FEDERAL DE ALAGOAS
CENTRO DE CIÊNCIAS AGRÁRIAS
CURSO DE MESTRADO EM AGRONOMIA
RIO LARGO, ESTADO DE ALAGOAS
JULHO DE 2006**



JANICE SILVA SALES

ANÁLISE *IN SILICO* EM ESTs DE CANA-DE-AÇÚCAR (*Saccharum officinarum*) RELACIONADOS À RESISTÊNCIA AO ESTRESSE HÍDRICO

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre, pelo Curso de Mestrado em Agronomia, Área de Concentração “Produção Vegetal”, do Centro de Ciências Agrárias da Universidade Federal de Alagoas.

Orientação: Prof. Dr. Cícero Eduardo Ramalho Neto

Catlogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária Responsável: Helena Cristina Pimentel do Vale

S163a Sales, Janice Silva.
Análise in silico em ESTs de cana-de-açúcar (*Saccharum officinarum*) relacionados à resistência ao estresse hídrico / Janice Silva Sales. – Rio Largo, 2006.
xix, 89f. : il. tabs., grafs.

Orientador: Cícero Eduardo Ramalho Neto.
Dissertação (mestrado em Agronomia : Produção Vegetal) – Universidade Federal de Alagoas. Centro de Ciências Agrárias. Rio Largo, 2006.

Bibliografia: f. [58]-63.

1. Cana-de-açúcar. 2. Genética vegetal. 3. Estresse hídrico. 4. ESTs. I. Título.

CDU: 633.61



Universidade Federal de Alagoas



PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
CENTRO DE CIÊNCIAS AGRÁRIAS
MESTRADO EM AGRONOMIA - CONCENTRAÇÃO EM PRODUÇÃO VEGETAL
CÓDIGO-CAPEs - 2601012010M -9

UFAL

CECA

Aos vinte e seis dias do mês de julho de dois mil e seis, às 14:00: horas, na sala de reunião do Programa de Melhoramento de Cana do Centro de Ciências Agrárias da UFAL, sob a presidência do Prof.Dr. CICERO EDUARDO RAMALHO NETO- (CECA/UFAL), reuniu-se a Banca Examinadora para Defesa da Dissertação de Mestrado da Bióloga JANICE SILVA SALES aluna do Curso de Mestrado em Agronomia, "Produção Vegetal e Proteção de Plantas" da UFAL, sob o título: Análise In Silício em ESTs de Cana de Açúcar (Saccharum officinarum, L.) Relacionados à Resistência ao Stress Hídrico". A Banca Examinadora ficou assim constituída: Prof. Dr.Cícero Eduardo Ramalho Neto-Orientador - CECA/UFAL; Prof.ª Dr.ª. Denise Wanderlei Silva/CECA/UFAL; - Membro Titular - Dr.Onaldo Souza/Pesquisador-EMBRAPA- Membro Titular; Profª Drª Edna Peixoto da Rocha Amorim - CECA/UFAL-Membro Titular e Prof. Dr. Eurico Eduardo Pinto de Lemos/CECA/UFAL- Suplente .

Ocorrências: Abertura pelo presidente da banca, Prof. Dr.Cícero Eduardo Ramalho Neto demais membros componentes da banca, agradecendo suas valiosas presenças, manifestando sua satisfação pela defesa de mais uma Dissertação de Mestrado do Curso de Agronomia do CECA, desta feita sob sua orientação. A seguir, parabenizou a mestranda JANICE SILVA SALES pelo trabalho apresentado. O presidente da Banca Examinadora iniciou os trabalhos, passando a palavra ao Dr.Onaldo Souza e logo após foram ouvidos os comentários e análises dos demais componentes da banca.

Terminada a defesa, procedeu-se o julgamento, pelos membros examinadores, sendo o candidato APROVADO, Fazendo jus, portanto, ao título de MESTRE EM AGRONOMIA, área de Concentração em Produção Vegetal. E que, para constar, lavrou-se a presente ata, que vai assinada pelos senhores membros da Banca Examinadora e por mim, Geraldo de Lima, Secretário. Rio Largo (AL), 26 de julho de 2006.

Geraldo de Lima
GERALDO DE LIMA

Prof. Dr. Cícero Eduardo Ramalho Neto
Presidente

Prof.ª Dr.ª. Denise Wanderlei Silva
Membro

Profª Drª Edna Peixoto da Rocha Amorim
Membro

Dr.Onaldo Souza
Membro

*Esta, bem como todas as minhas demais conquistas, dedico aos meus amados pais, **Edvalda Davi e José Sales**, sinônimos de força, dedicação e amor incondicional. Foram os meus primeiros educadores, propiciaram a mim e meus irmãos um lar perfeito, nos mostrou a importância do estudo constante, nos incentivam e nos orientam nas decisões mais importantes e por serem as pessoas que mais amo, admiro e tenho orgulho nesta vida,*

DEDICO

*Aos meus irmãos que amo, **Jane Clay, Jaqueline, Jeanne e José Sales Jr.**, ao meu cunhado **Augusto César**, ao meu namorado **Francisco Chichorro** e minha mais que amiga **Muciana Cunha**, pelo amor e incentivo a mim sempre dados,*

OFEREÇO

Feliz aquele que transfere o que sabe e aprende o que ensina.

(Cora Coralina)

AGRADECIMENTOS

A **Deus**, por minha existência e condições saudáveis para desempenhar minhas atividades. Por ter-me concedido a oportunidade de crescer, ser educada e amada por uma família que me instruiu com valores morais e disciplina.

Aos meus **pais e irmãos**, por todo carinho, dedicação, apoio e amor incondicional. A **papai e mamãe** especialmente, por todos os esforços destinados a minha educação, pelos ensinamentos básicos e primordiais para meu desenvolvimento como pessoa e profissional e por serem meus melhores exemplos de determinação, força, coragem e perseverança.

Ao professor **Dr. Eduardo Ramalho Neto** (prof, Edu, Dudu), sinônimo de competência e admiração, meu estimado orientador e amigo, meus sinceros agradecimentos por seus ensinamentos, conselhos, pela oportunidade de aprendizado instruindo-me em diversas técnicas biomoleculares que possibilitaram maior presteza na realização desse trabalho, por proporcionar um ambiente harmonioso de convívio respeitoso, pela confiança, pelo apoio, estímulo, consideração, enfim, pela alegria de trabalharmos juntos.

A **Dr^a Denise Wanderlei**, pelos ensinamentos ao longo desses anos, pelo respeito, atenção, paciência e dedicação, exemplo de competência e profissionalismo.

Ao meu adorável namorado **Francisco Chichorro**, por seu amor, incentivo, compreensão, pelas palavras certas nas horas exatas e por todo apoio mesmo estando além mar.

A minha inestimável amiga de todas as horas **MuciMana (Muciana Cunha)** que em sua imensa generosidade presenteou-me com seu carinho constante, apoio, dedicação, com sua energia positiva de caráter prático, com sugestões sempre muito úteis, com companheirismo, reciprocidade, com incentivo e também atenuando as atividades durante a execução deste trabalho. Companheira ímpar e memorável.

A **Marusa, Marlécio, Ana Cunha e Adriana** pelo incentivo e carinho.

Às minhas estimadas amigas **Alana Sarmiento, Deyse Rocha, Eurídice Godoy, Karla Guimarães, Luciana Viana, Sarah Cavalcanti, Rosivânia Nascimento, Rosa Gonzaga** pela amizade, pelo apoio, pela força, pelos maravilhosos momentos de descontração e prontidão em auxiliar-me em qualquer circunstância.

*Aos amigos **André Galvão** e **Daniela Furtado**, que com paciência, dedicação, hospitalidade ajudaram-me estudar estatística para a seleção de mestrado, incentivando-me com amizade, contentamento e confiança.*

*A minha amiga, companheira, “admiradora”, fiel e inesquecível amiga **Rislene Moura** e meu querido amigo **Benedito** pessoas maravilhosas que ocupam lugar especial no meu coração.*

*As adoráveis **Giselle Amorim** e **Sabrina Gomes**, companheiras que tanto contribuíram para a execução deste trabalho, com carinho e dedicação.*

*A minha especial amiga e conselheira **D. Lourdes**.*

*Aos meus amigos **Thelmo** e **Gilnisom** pelo carinho.*

*A **Hamilton Oiticica** por toda ajuda nos meus primeiros passos de aprendizado em biotecnologia e por seu carinho e amizade.*

*Aos colegas de laboratório **Rívia**, **Paulo Pedro**, **Carlos Jorge**, **Alba** e **Francisco Freire** e **Betânia** pela convivência harmoniosa e participativa.*

*Aos funcionários e colegas do Laboratório de Biotecnologia Vegetal-BIOVEG, **prof Eurico Lemos**, **profª Leila**, **Hélio** e **Gilvânia**, **Jaqueline** e **Fernanda** pela amizade, dedicação e atenção.*

*A **profª Edna Amorim** pela especial compreensão e colaboração que foram essenciais à etapa de conclusão deste trabalho.*

*Ao secretário **Geraldo de Lima**, pela nobreza no desempenho de sua função, sempre pronto a atendermos com profissionalismo e dedicação.*

*Aos meus colegas de turma, em especial **Emília**, **Dulce**, **Guadalupe**, **Luciano**, **Gêvoness**, **Celene**, **Jaqueline** e **Fernanda**.*

*A todos os **professores** que fizeram parte de minha formação.*

A todos aqueles que direta ou indiretamente participaram de mais esta etapa de minha vida, mas que não foram citados.

*À **Universidade Federal de Alagoas**, Centro de Ciências Agrárias, pela oportunidade de realização do curso de Mestrado em Agronomia.*

*Ao **GEMPRO**, laboratório de execução deste trabalho.*

*À Fundação de Amparo à Pesquisa do Estado de Alagoas, **FAPEAL**, pela concessão da bolsa de estudos.*

SUMÁRIO

LISTA DE TABELAS.....	xi
LISTA DE FIGURAS.....	xii
LISTA DE ABREVIATURAS E SÍMBOLOS.....	xvi
RESUMO.....	xviii
ABSTRACT.....	xix
INTRODUÇÃO.....	1
REVISÃO DE LITERATURA.....	5
1. Cana-de-açúcar.....	5
1.1 - Características gerais.....	5
1.2 - Origem e dispersão.....	6
1.3 – Importância sócio - econômica.....	7
1.4 – Relações hídricas	9
2. Genoma e mineração de dados (<i>data mining</i>)	10
METODOLOGIA.....	14
1. Local de execução.....	14
2. Material biológico.....	14
3. Crescimento bacteriano.....	15
4. Extração de DNA plasmidial (Mini-prep).....	16
5. Análise do DNA plasmidial.....	18
6. Reação de PCR (<i>polymerase chain reaction</i>)	18
7. Precipitação e purificação.....	19
8. Desnaturação.....	19
9. Seqüenciamento automático.....	19

10. Análise do gel.....	22
11. Submissão das seqüências ao banco de dados SUCEST – UNICAMP.....	22
12. Mineração de dados (<i>data mining</i>)	23
RESULTADOS E DISCUSSÃO.....	27
1. Extração de DNA plasmidial (mini-prep)	28
2. Seqüenciamento de EST da cana-de-açúcar.....	29
3. Mineração de dados (<i>data mining</i>)	30
4. Identificação eletrônica das seqüências.....	34
5. Identificação de proteínas <i>in silico</i>	39
5.1 - <i>WSI76 protein induced by water stress</i>	39
5.2 - 41,9 KDa ou gene T25K17,60.....	47
5.3 - <i>Alpha amylase [precursor] (fragment)</i>	48
5.4 - <i>Galactinol synthase</i>	51
5.5 - <i>Cytosolic pyruvate orthophosphate dikinase</i>	53
5.6 - <i>Lactoylglutathione lyase</i>	54
CONCLUSÕES.....	57
REFERÊNCIAS.....	58
APÊNDICES.....	64

LISTA DE TABELAS

Tabela 1. Produção de cana-de-açúcar de 1998 a 2006.....	8
Tabela 2. Números de identificação das bibliotecas genômicas utilizadas para seqüenciamento.....	15
Tabela 3. Parâmetros eletroforéticos de seqüenciamento no ABI PRISM 377.....	21
Tabela 4. Descrição das ferramentas da família BLAST.....	24
Tabela 5 <i>Clusters</i> homólogos relativos às seqüências análogas das proteínas relacionadas à resistência ao estresse hídrico.....	33

LISTA DE FIGURAS

Figura 1.	Transferência de clones com replicador <i>Boekel</i>	16
Figura 2.	Culturas de bactérias em placas de crescimento sob agitação em <i>Shaker</i>	16
Figura 3.	ABI PRISM 377 DNA SEQUENCER (<i>Applied Biosystems</i>).....	20
Figura 4.	A) ABI PRISM 377 pronto para seqüenciamento B) Detalhe da etapa de carregamento das amostras no gel de eletroforese no seqüenciador automático com auxílio aplicador multicanal.....	21
Figura 5.	Alinhamento do gel em programa <i>Sequencing Analysis</i>	22
Figura 6.	A) Página inicial do NCBI; B) página inicial do SUCEST.....	24
Figura 7.	Busca inicial no NCBI - <i>keyword: water stress</i> (em destaque).....	25
Figura 8.	Seqüência no modo FASTA.....	25
Figura 9.	Transferência da seqüência FASTA para submissão ao BLAST no SUCEST.....	26
SALES, J. S. 2006. <i>Análise in silico em ESTs de cana-de-açúcar (Saccharum officinarum) relacionados à resistência</i>		xii

Figura 10. Gel de agarose 1,2% corado com brometo de etídio da biblioteca plasmidial NR1.051 RII.	28
Figura 11. Imagem do gel de eletroforese após seqüenciamento em ABI PRISM 377.	29
Figura 12. Gráfico de homologia entre seqüências após submissão a ferramenta BLAST.	30
Figura 13. Página inicial do NCBI apresentando a indicação do tipo de busca (nucleotídeos) e palavra chave (<i>water stress</i>) utilizada na mineração de dados (setas).	35
Figura 14. Resultados da busca no banco de nucleotídeos do NCBI utilizando a palavra chave <i>Water stress</i> no campo de pesquisa (<i>search</i>) (Seta).....	35
Figura 15. Item selecionado para adaptação da seqüência ao modo FASTA (setas).	36
Figura 16. Resultado da utilização da ferramenta FASTA. A seqüência relacionada à resistência ao estresse hídrico apresenta um sinal de > seguida da identificação.	36
Figura 17. Página inicial da ferramenta BLAST do SUCEST (acesso restrito) para análise de homologia (BLASTn). A seqüência submetida para análise (<i>Query</i>) foi comparada com seqüências de cana-de-açúcar (<i>Subject</i>).....	37
Figura 18. Resultado da análise de similaridade. Histograma gerado pela ferramenta BLASTn, onde se observa alta homologia entre as seqüências confrontadas. Linhas vermelhas indicam <i>scores</i> > 200.....	38
Figura 19. Valores de <i>scores</i> e <i>e-values</i> comprovando alta homologia entre as seqüências (Destaque).....	38
Figura 20. Submissão da proteína WSI76 como palavra chave no banco de dados SWISS-PROT/TrEMBL (Seta).....	40

Figura 21.	Resultado da busca com informações e <i>links</i> que dão acesso ao InterPro, Pfam entre outros (Setas).....	40
Figura 22.	Página de resultados com informações sobre a proteína no banco de dados do EBI acessado através de link no Swiss-Prot.....	41
Figura 23.	Resultado ilustrando domínio conservado da proteína.....	41
Figura 24.	Visualização da estrutura 2D da cadeia “A” na página do Pfam acessada através de <i>link</i> do Swiss -Prot.....	42
Figura 25.	Página do PDBsum, onde foi visualizada a estrutura cristalizada da proteína com <i>links</i> de acesso para outras informações.....	42
Figura 26.	Página do PDBsum contendo informações e <i>link</i> de acesso que leva ao KEGG.	43
Figura 27.	Rota metabólica do metabolismo da pirimidina a qual a enzima está inserida.	44
Figura 28.	Representação esquemática da cadeia “A” (estrutura secundária).....	45
Figura 29.	Acesso a visualização da estrutura 3D da proteína na página do PDBsum.	45
Figura 30.	Visualização da estrutura 3D da cadeia A da proteína WS76 no PDBsum através do programa RASMOL. Neste programa é possível girar, movimentar a estrutura em vários ângulos, mudar apresentação, entre outras funções.....	46
Figura 31.	Rota metabólica da pirimidina, a qual a proteína 41,9 KDa está inserida.	47
Figura 32.	Visualização da estrutura 3D da cadeia A da hipotética 41,9 KDa (gene T25K17,60) através do programa VRML do PDBsum.....	48

Figura 33.	Rota metabólica da sucrose onde a <i>Alpha amylase</i> atua.....	49
Figura 34.	Estrutura 3D da cadeia A da proteína Alpha-amilase (Gen.CUS AMY2).....	50
Figura 35.	Rota metabólica do nucleotídeo – açúcar onde atua a Galactinol synthase.	51
Figura 36.	Visualização da estrutura 3D cadeia A da <i>Galactinol synthase</i> através do programa VRML do PDB.....	52
Figura 37.	Visualização da estrutura 3D da cadeia A da proteína Cytosolic pyruvate orthophosphate dikinase através do programa VRML do PDB.	53
Figura 38.	Rota metabólica mostrando a participação da <i>lactoilglutaciona liase</i> no metabolismo da tirosina.....	55
Figura 39.	Visualização da estrutura 3D da cadeia A da proteína <i>lactoilglutaciona liase</i> através do programa VRML do PDB.....	55

LISTA DE ABREVIATURAS E SÍMBOLOS

% - por cento

∞ - infinito

µg – microgramas

µg/mL – micrograma por mililitro

µL –microlitro (s)

10x - dez vezes

1x - uma vez

aa - aminoácido

ADE - água destilada esterilizada

PMGCA - Programa de Melhoramento Genético da Cana de Açúcar

AL – Alagoas

BLAST - *Basic Local Alignment Search Tool*

BIOINFORMAL - Laboratório de Bioinformática da Universidade Federal de Alagoas

CBMEG - Centro de Biologia Molecular e Engenharia Genética

CECA – Centro de Ciências Agrárias

CIA – clorofórmio-álcool-isoamílico

DNA - ácido desoxirribonucléico

EDTA – ácido etileno diamino tetra-acético

F – *forward*

FAPEAL – Fundação de Amparo à Pesquisa de Alagoas

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

g – gramas

GEMPRO – Laboratório de Genética Molecular, Genômica e Proteômica

GET - (20% Glicose, 0,5M EDTA, 1M Tris-HCl, pH 7,4

h- horas

HCl –ácido clorídrico

kb – quilobase (mil pares de bases)

KOAc – acetato de potássio

L – litro

M – molar
mg – miligrama
MgCl₂ – cloreto de magnésio
Milli-Q – água destilada deionizada
min – minuto (s)
mL – mililitro
mM – milimolar
NaCl – cloreto de sódio
NaOH – hidróxido de sódio
NCBI - National Center for Biotechnology Information
ng – nanogramas
°C - graus Celsius
pb – pares de base
PCR – reação em cadeia da polimerase (*polymerase chain reaction*)
pH – potencial de hidrogênio
R – *reverse*
RJ – Rio de Janeiro
RNAse - ribonuclease A
rpm – rotações por minuto
SDS - dodecil sulfato de sódio
seg – segundos
SUCEST - Sugar Cane EST Genome Project
SP – São Paulo
Taq polimerase – DNA polimerase da bactéria *Thermus aquaticus*
TE – Tris-HCl 10mM, EDTA 1mM
Tris – trishidroximetil aminometano
U – unidade (s)
UFAL – Universidade Federal de Alagoas
UNICAMP - Universidade Estadual de Campinas
UV – ultravioleta
V – volts

RESUMO

Etiquetas de seqüências expressas geradas pelo Projeto ESTs de cana-de-açúcar (FAPESP) forneceram um banco de dados útil para a investigação das seqüências de DNA relacionadas a muitas características agroindustriais da cana-de-açúcar. Quando comparadas às seqüências de DNA de diversas espécies de plantas disponíveis no banco de dados do Centro Nacional de Informação Biotecnológica (NCBI), foi possível avaliar várias proteínas produzidas por esta planta, incluindo aquelas relacionadas à resistência ao estresse hídrico. O SUCEST gerou aproximadamente 43.150 *clusters* de onde foram obtidas 52 proteínas putativas relacionadas à resistência ao estresse hídrico em 2.998 *clusters* analisados. A identificação de proteínas *in silico* teve como suporte um protocolo modificado desenvolvido pelo SUCEST, que usou ferramentas de bioinformática específicas, disponíveis no NCBI, como por exemplo BLAST e FASTA..

Termos para indexação: ESTs, cana-de-açúcar, estresse hídrico.

ABSTRACT

The expressed sequenced tags (ESTs) generated by Sugarcane ESTs Project (FAPESP) provided a useful databank for the investigation of DNA sequences related to many different sugarcane agroindustrial features. When compared to DNA sequences from several plant species available at The National Center for Biotechnology Information Databank it was possible to perform a survey on various proteins produced by this plant including those related to water stress resistance. SUCEST generated approximately 43,150 clusters from where 52 best hits were obtained in this survey for putative proteins related to water stress resistance on 2,998 analyzed clusters. In silico protein identification was achieved following a modified protocol developed by SUCEST which used specific bioinformatics tools available at NCBI such as BLAST and FASTA.

Index terms: ESTs, sugarcane, water stress.

INTRODUÇÃO

A cana-de-açúcar (*Saccharum officinarum* L.), cultivada em áreas tropicais e subtropicais em mais de 80 países, constitui-se numa das mais importantes espécies vegetais da atualidade, particularmente em função de seu papel como fonte de energia (açúcar, etanol e eletricidade). Em 1995, a produção atingiu valores em torno de 1.2×10^9 t, resultante do cultivo em 18 milhões de hectares distribuídos ao redor do globo. Esta herbácea da família das gramíneas. A cana é matéria-prima de grande flexibilidade, com ela obtém-se o açúcar, álcool de vários tipos; fabrica-se bebidas como cachaça, rum e vodka e gera eletricidade a partir do bagaço via alcoolquímica. Também pode produzir papel, plásticos e produtos químicos. Da cana se aproveita absolutamente tudo: bagaço, méis, torta e resíduos de colheita, devido à sacarose contida em seu caule, formado por numerosos nós. O potencial de produção de bens de consumo gerados pela cana-de-açúcar gira em torno de 100 diferentes produtos comerciais, apesar de ser considerada exclusivamente como produtora de açúcar e álcool.

A atividade canavieira do Brasil tem exercido papel preponderante na formação e desenvolvimento das sociedades, em função de sua contribuição para a alimentação e produção de energia. É responsável por cerca de um milhão de empregos diretos, 511 mil

apenas na produção de cana-de-açúcar. O restante está distribuído na agroindústria de açúcar e álcool.

Os problemas relacionados à produção da cana-de-açúcar são praticamente os mesmos de outras culturas: necessidade de água e fertilizantes; compactação dos solos; concorrência de outras plantas (chamadas "daninhas" ou "invasoras"); assédio de fungos, bactérias, insetos e nematóides, que debilitam a cultura. Hoje, tais problemas são parcialmente resolvidos com a utilização de defensivos agrícolas, ou agrotóxicos. Entretanto, no Nordeste brasileiro, o principal fator limitante da produção de açúcar, álcool e derivados são os prolongados períodos de estiagem verificados anualmente. As chuvas irregulares têm contribuído para a redução da produtividade da cana-de-açúcar, além de determinar investimentos elevados nas chamadas "irrigações salvadoras", na tentativa de manter os níveis de produtividade atuais.

O estresse hídrico tem sido o principal fator limitante dos acréscimos na produtividade de cana-de-açúcar no Nordeste brasileiro. Diversos materiais genéticos de excelente desempenho construídos pelo PMGCA têm no déficit hídrico no solo, durante a maior parte do ano, seu maior adversário nas áreas produtoras dessa gramínea.

A deficiência hídrica provoca alterações, cuja irreversibilidade no comportamento vegetal vai depender do genótipo, duração e severidade do estresse, estágio de desenvolvimento e história prévia do estresse. As reações das plantas à falta de água no solo são manifestadas por um quadro sintomatológico bastante complexo, sendo que as alterações nas respostas fisiológicas, ocorrem bem antes que os sintomas possam ser percebidos visualmente. O estado de água na planta representa uma integração da demanda atmosférica, potencial de água no solo, densidade e distribuição de raízes, bem como de outras características da planta (CLARK; HILER, 1973).

Em decorrência da posição de destaque que ocupa na economia mundial, a cultura da cana está freqüentemente inserida em programas de melhoramento de espécies cultivadas, visando à introdução de características de interesse agrônomo, como resistência a pragas e patógenos, tolerância a herbicidas, e aumento no teor de sacarose (CIDADE et al., 2006). Um dos atuais desafios do Brasil é aumentar a oferta de álcool combustível. As soluções englobam desde novas variedades de cana-de-açúcar, incluindo

plantas transgênicas, até a simples expansão da área agrícola, além de inovações na linha de produção das usinas. Sinônimo de combustível renovável, que polui menos em comparação com os derivados de petróleo, o etanol voltou a ocupar um lugar de destaque no cenário energético do país e também começou a ser desejado por vários países. No caso brasileiro, os responsáveis pelo renascimento do álcool são os carros bicompostíveis, ou *flex fuel*, que podem ser abastecidos com álcool e gasolina, em qualquer proporção.

Enquanto novas usinas são construídas, outros flancos estão sendo abertos no âmbito da pesquisa tecnológica. As inovações vêm principalmente dos estudos genéticos, como novas variedades e plantas transgênicas. Recentemente num projeto entre o Centro de Tecnologia Canavieira (CTC), a Usina Central de Álcool Lucélia e pesquisadores da Universidade de São Paulo (USP) e Unicamp, financiados pela FAPESP, foram depositados e patenteados 200 genes identificados em diversas variedades de cana que estão relacionados à produção de sacarose, componente fundamental da indústria açucareira e também imprescindível no processo de fermentação, servindo de alimento para a levedura produzir o álcool. (OLIVEIRA; VASCONCELOS, 2006).

No momento atual, a genômica tem contribuído sobremaneira para a análise dos materiais genéticos disponíveis nos centros de pesquisa. As técnicas incluem, entre outras, a clonagem de genes, seqüenciamento de DNA, e a transformação de plantas através da biobalística ou agrotransformação. Além disso, ferramentas de Bioinformática são consideradas imprescindíveis nas análises de expressão gênica, como também os trabalhos em proteômica. Em julho de 1999, com o objetivo de seqüenciar partes escolhidas do DNA da cana-de-açúcar e identificar cerca de 50 mil genes com características de interesse econômico foi executado o Projeto Genoma de ESTs de Cana-de-Açúcar (SUCEST), financiado pela FAPESP- SP.

A identificação e a compreensão dos mecanismos de tolerância à seca são fundamentais no desenvolvimento de novas cultivares comerciais mais tolerantes ao déficit hídrico. A expressão de genes (ainda não caracterizados) em genótipos tolerantes pode ser usada no estudo de mecanismos de tolerância à seca e na identificação de outros genótipos com características similares. Genes induzidos pelo déficit hídrico promovem: tolerância da célula à desidratação; funções de proteção no citoplasma; alterações no potencial osmótico

celular para aumentar a absorção de água; controle da acumulação de íons; regulação adicional de expressão gênica e metabolização de compostos degradados pelo estresse (MARUR, 2005).

Um dos principais objetivos desse trabalho foi analisar os dados produzidos pelo Projeto *Data Mining* (aprovado pela Coordenação do Projeto Genoma Cana - FAPESP), resultado das atividades de seqüenciamento automático de DNA nos laboratórios de São Paulo, Alagoas e Pernambuco, visando a identificação de proteínas relacionadas ao estresse hídrico.

REVISÃO DE LITERATURA

1. Cana-de-açúcar

1.1 - Características gerais

A cana-de-açúcar é uma gramínea perene pertencente ao gênero *Saccharum*, tendo a seguinte classificação botânica, Divisão: *Magnoliophita*; super divisão: *Spermatophyta*; Classe: *Liliopsida*; Sub classe: *Commelinidae*; ordem: *Cyperales*; gênero: *Saccharum*; espécies: *Saccharum officinarum* L., *Saccharum spontaneum* L., *Saccharum robustum* (THE PLANTS database, 2000).

Planta ereta, rizomatosa, formando touceiras, possui colmo cilíndrico, quase sempre fistulosos, espessos, de coloração variável, com nós marcados e entrenós distintos e visíveis. As folhas são simples, alternadas, compridas e estreitas. O fruto é seco do tipo cariopse e com semente de endosperma abundante. O sistema radicular é bem ramificado e não possui raiz pivotante, sendo que a cana planta possui maior quantidade de raízes que a cana soca e cerca de 75% de seu sistema radicular se encontra nos primeiros 50cm de solo. O caule é a parte utilizável, desenvolve-se em touceiras, atinge entre 3 e 6 m de altura, por 2 a 5 cm de diâmetro, caule do tipo colmo, com nós e entrenós. Sob determinadas condições de fotoperíodo, temperatura e umidade, floresce emitindo uma inflorescência ou panícula com flores hermafroditas, que recebem nomes vulgares de flecha ou bandeira, cuja cor, forma e tamanho são específicos de cada cultivar (BACCHI, 1985). No que se refere aos fatores climáticos, o crescimento da cana-de-açúcar mostra-se variável, dependendo da

cultivar e das condições ambientais. A temperatura ideal para o desenvolvimento da cana-de-açúcar situa-se entre 20 e 35°C, sendo que as regiões aptas para o cultivo da cana devem possuir temperatura média anual superior a 20°C (Castro, 2001 apud SILVA, 2004).

1.2 - Origem e dispersão

O lugar onde a cana-de-açúcar germinou pela primeira vez é uma incógnita, devido à quantidade de gramíneas híbridas existentes e à falta de documentação que certifique sua origem (ÚNICA, 2006^a).

Referências indicam que essa gramínea tenha se difundido pela Pérsia, Síria, Egito e Sicília ainda na época de Alexandre Magno, o Grande, então rei da Macedônia, por ocasião da conquista e organização do império Persa em aproximadamente 327 a.C. A expansão de seus domínios propiciou canais mercantis que facilitaram a evasão e influência de artigos, especiarias e costumes entre a Ásia Meridional, sul e sudeste da Europa e norte da África. Segundo MACHADO JR. et al. (1987) apud MELO JR, 2006 a origem da cana-de-açúcar e sua completa história encontram-se perdidas na antiguidade. Foram encontradas referências relatando que se estabeleceu há cerca de 6.000 anos A.C. na Melanésia, Indonésia, Nova Guiné, Índia e China.

Por muito tempo, durante sua difusão pelo continente europeu, o açúcar foi encarado como uma substância rara, utilizada no tratamento de enfermidades das mais diversas origens, chegando a ser incluído em testamento, como o da mulher de Carlos V, imperador do Sacro Império Romano Germânico. Somente após o século XV a cana-de-açúcar passou a ter uma produção em larga escala, tornando-se então, sob a ótica comercial, um produto importante e atraente para o mercado mundial. Em Portugal foi introduzida através da Ilha da Madeira pelo Infante Dom Henrique, ainda no mesmo século. No Brasil a cana-de-açúcar passou a ser cultivada a partir do terceiro decênio do século XVI .

1.3 – Importância sócio-econômica

Entre 111 países produtores de açúcar, 73 cultivam cana-de-açúcar e fornecem 3/4 da produção mundial de açúcar. O Brasil é o maior produtor de cana-de-açúcar do mundo, com os menores custos de produção, e também o maior exportador do produto, seguido pela Índia, o segundo maior produtor, mas com uma produção que equivale à metade da brasileira e a da Austrália. Em média, 55% da cana brasileira produz álcool e 45%, açúcar. Planta-se cana no Centro-Sul e no Norte-Nordeste, o que permite dois períodos de safra. Produz-se, portanto, o ano todo. Dependendo do momento de plantio, a cana demora de ano a ano e meio para ser colhida e processada pela primeira vez (UNICA, 2006b).

A região Centro-Sul com 407,6 milhões de toneladas é responsável por 86,8% da produção nacional, ocupa 81,4% da área a ser colhida (5,0 milhões de hectares) e detém a maior produtividade média do país, de 81.338 kg/ha. A região Sudeste contribui com 325,9 milhões de toneladas, o correspondente a 69,4% da produção nacional e 80% da produção do Centro-Sul, São Paulo é o principal estado produtor com 265,5 milhões de toneladas na safra 2005/2006. A produção nacional de cana-de-açúcar na safra 2005-2006 foi de 436,8 milhões de toneladas, crescimento de 5,1% em relação à safra anterior. A produção da região Norte-Nordeste é de 62,2 milhões de toneladas, correspondente à 13,2% da produção nacional, cultivada numa área de 1,1 milhão de hectares, 18,6% da área a ser colhida no país (CONAB,2006).

A região Nordeste, tradicional produtora de açúcar, vem perdendo posição relativa para a produção de São Paulo desde o lançamento do Proálcool, há 21 anos. Melhores solos, áreas mecanizáveis, fortes investimentos em pesquisa e clima mais regular têm permitido a São Paulo uma produtividade muito superior a Pernambuco e Alagoas. Enquanto os produtores paulistas colhem em média quase 80 kg/ha, a produtividade média no Nordeste não passa de 60 kg/ha. Em função dessa diferença é que vem crescendo a participação da produção paulista no total nacional. Na produção de álcool São Paulo representa 80% e o Nordeste, que no auge do Proálcool representou 20%, hoje não passa de 6%. Já na produção de açúcar as usinas paulistas produzem mais de 50% do total da produção nacional, enquanto as usinas nordestinas não chegam a 30% (PESSOA, 2006). A

tabela 1 demonstra o perfil da produção de cana-de-açúcar desde 1998 a 2006 entre São Paulo e demais regiões (ÚNICA, 2006b).

Tabela 1. Produção de cana-de-açúcar de 1998 a 2006.

PRODUÇÃO DE CANA-DE-AÇÚCAR (Toneladas)					
SAFRA	Região CENTRO/SUL			Região N/NE	BRASIL
	São Paulo	Demais	Total		
1998/1999	199.313.949	70.209.046	269.522.995	45.141.192	314.664.187
1999/2000	194.234.474	69.735.142	263.969.616	36.444.343	300.413.959
2000/2001	148.256.436	58.842.621	207.099.057	49.291.326	256.390.383
2001/2002	176.574.250	67.645.273	244.219.523	47.704.407	291.923.930
2002/2003	190.627.892	77.920.050	268.547.942	50.463.092	319.011.034
2003/2004	207.572.535	91.024.783	298.597.318	60.194.968	358.792.286
2004/2005	229.768.505	97.486.208	327.254.713	57.511.153	384.765.866
2005/2006	242.828.824	94.028.105	336.856.929	49.727.458	386.584.387

Fonte: IBGE

O Estado de Alagoa foi o maior produtor de cana-de-açúcar do Nordeste em 2004, segundo dados do IBGE (2005). No entanto, ainda têm sido baixos os níveis de produtividade agroindustrial e de rendimentos de mão-obra, e vários estudos evidenciam a heterogeneidade das unidades de produção de açúcar e álcool (ANDRADE e BARBOSA, 1996). O setor sucroalcooleiro do Estado de Alagoas tem sua importância reconhecida por ser gerador de emprego e renda, com mais de 90 mil empregos diretos, a partir das suas 26 unidades produtoras de açúcar e álcool. Na safra 05/06, obteve produção superior a 22 milhões de toneladas de cana, transformados em cerca de 1,1 milhão de toneladas de açúcar e mais de 500 mil metros cúbicos de álcool (SINDAÇÚCAR-AL, 2006).

A cana, também matéria-prima para extração de álcool, tem o potencial energético de 1,2 barril de petróleo para cada tonelada. Atualmente, o álcool movimenta 15% da frota automotiva do país. Em 2003/2004, o Brasil produziu 14,4 bilhões de litros de álcool. No ano passado, o volume de embarques alcançou cerca de 800 milhões de litros. Combustível não-poluente, o álcool é um produto que cada vez mais interessa às nações interessadas em reduzir a emissão de gases nocivos à saúde humana (GLOBAL 21, 2006).

O setor sucroalcooleiro do Brasil é o mais competitivo em termos internacionais, por apresentar o melhor desempenho quanto às relações entre rendimentos, produtividade e custos. O grande indutor desse desenvolvimento foi a expansão que o setor experimentou nos últimos 20 anos, como decorrência do aumento da produção do álcool e da produção

do açúcar, atualmente focada no máximo aproveitamento da cana-de-açúcar (OLIVÉRIO, 2002). Esse crescimento deve-se principalmente à instalação de novas usinas para atender à crescente demanda por açúcar e álcool. Do total de cana produzido, 216 milhões de toneladas foram destinados à produção de açúcar e 178 milhões de toneladas de cana para a fabricação de álcool hidratado e anidro, o suficiente para gerar 17 bilhões de litros do produto.

1.4 - Relações Hídricas

Segundo ALEXANDER (1973), cerca de 70% do peso da matéria fresca da cana-de-açúcar é constituída de água. Para cada parte de matéria seca produzida há perda de 250 partes de água. Em solos em condições de coeficiente de murcha, quando a água aparentemente não se encontra disponível, há uma severa redução no crescimento da cana-de-açúcar, mas esta ainda consegue retirar água do solo, mesmo quando seu conteúdo está em torno de 2 a 4% abaixo do coeficiente de murcha. No momento que é fornecido água à planta, ela retoma imediatamente o crescimento em ritmo constante.

O déficit hídrico no solo, durante a maior parte do ano é o maior adversário da produtividade nas áreas produtoras do Nordeste brasileiro. A deficiência hídrica provoca alterações, cuja irreversibilidade no comportamento vegetal vai depender do genótipo, duração e severidade do estresse, estágio de desenvolvimento e história prévia do estresse. As reações das plantas à falta de água no solo são manifestadas por um quadro sintomatológico bastante complexo, sendo que as alterações nas respostas fisiológicas, ocorrem bem antes que os sintomas possam ser percebidos visualmente. O estado de água na planta representa uma integração da demanda atmosférica, potencial de água no solo, densidade e distribuição de raízes, bem como de outras características da planta (CLARK; HILER, 1973).

2. Genoma e mineração de dados (*data mining*)

Os organismos respondem a diferentes estímulos ambientais por meio de alterações no proteoma, conjunto de proteínas das células. Esse processo pode permitir a adaptação a condições adversas, como, por exemplo, choque térmico, exposição a drogas, ataque de patógenos, etc. O fluxo da informação gênica do DNA nos cromossomos (genoma) até o proteoma é intermediado pelo conjunto das moléculas de RNA (transcritoma) (FELIX et al. 2002).

Os programas de melhoramentos genéticos tradicionais esbarram na complexidade do seu genoma, consequência de sua origem e poliploidia (ALMEIDA, 2002). Segundo INGELBRECHT et al. (1999), em virtude de sua origem múltipla, a cana-de-açúcar possui um dos genomas mais complexos do reino vegetal, apresentando um nível variável de ploidia (geralmente de 5x a 14x, em que $x = 5, 6, 8, 10, 12$ ou 14). Essa complexidade dificulta a aplicação de técnicas de melhoramento convencional. Diversos centros de pesquisas, tanto no Brasil como no exterior, há anos geram novas variedades de cana-de-açúcar. O desenvolvimento de variedades resistentes a estes males, obtidas pela modificação genética das plantas, consiste em uma alternativa aos defensivos agrícolas.

A idéia de decifrar o genoma da cana-de-açúcar partiu do Centro de Tecnologia da Copersucar - Cooperativa dos Produtores de Cana, Açúcar e Álcool do Estado de São Paulo, em meados de 1998. Na ocasião, as investigações sobre o material genético da *Xylella fastidiosa* estavam bem adiantadas e o sucesso do projeto fez com que sua financiadora, a Fapesp, anunciasse o interesse em investir em outras pesquisas do gênero. A Copersucar apresentou uma proposta à instituição para incluir a cana-de-açúcar no seu Programa Genoma, com um forte argumento para isso: um quarto de toda a cana-de-açúcar produzida no mundo é colhida no Brasil, sendo que só no Estado de São Paulo responde por 60% desse total. Seria o primeiro projeto nacional a tornar conhecidos os genes de uma planta (PAIVA, 2000).

Nas etapas do referido projeto, os pesquisadores do Genoma Cana buscaram isolar somente os genes envolvidos na síntese de proteínas, desprezando todo o restante do material genético contido no DNA. Para organizar esse acervo, 21 laboratórios de bioinformática envolvidos no projeto receberam cópias de trechos do DNA extraídos de

diferentes tecidos celulares da cana. Através de etiquetas de seqüências transcritas - EST (do inglês, *Expressed Sequence Tags*), os pesquisadores efetuaram o seqüenciamento dos genes expressos, que permite conhecer apenas os chamados genes expressos, ou o genoma funcional da planta (PAIVA, 2000). A cana ideal, mais produtiva, resistente a pragas e doenças, tolerante à seca e a herbicidas e eficiente na absorção de nutrientes poderá ser constituída a partir das descobertas do Genoma Cana.

Motivadas pela grande disponibilidade de recursos computacionais e a facilidade de troca e armazenamento de informações, instituições das mais diversas áreas do conhecimento têm produzido e armazenado eletronicamente uma grande quantidade de dados. Surge, então, a necessidade de técnicas eficientes para análise desses volumes de dados e extração de informações que possam ser úteis ao processo de tomada de decisão. Tais métodos têm contribuído para o surgimento de sistemas computacionais capazes de adquirir novos conhecimentos, novas habilidades e novas maneiras de organizar o conhecimento existente (MITCHELL, 1997 *apud* METZ; MONARD, 2006). Esses sistemas são, na sua maioria, produtos do Aprendizado de Máquina — AM —, cujos algoritmos são amplamente utilizados em tarefas de descoberta de conhecimento de bases de dados (*knowledge discovery from databases* — *KDD*) principalmente para automatizar o processo de análise de dados e extração de conhecimento útil e preferencialmente novo (FAYYAD et al., 1996 *apud* METZ; MONARD, 2006).

“A bioinformática é uma nova disciplina científica com raízes nas ciências da computação, na estatística e na biologia molecular. Esta ferramenta foi desenvolvida para enfrentar os resultados das iniciativas de seqüenciamento de genes, que produzem uma quantidade cada vez maior de dados sobre proteínas, DNA e RNA. Desse modo, os biólogos moleculares passaram a utilizar métodos estatísticos capazes de analisar grandes quantidades de dados biológicos, a predizer funções dos genes e a demonstrar relações entre genes e proteínas” (VOGT, 2006).

Data mining ou mineração de dados consiste em um processo analítico projetado para explorar grandes quantidades de dados (tipicamente relacionados a negócios, mercado ou pesquisas científicas), na busca de padrões consistentes e/ou relacionamentos sistemáticos entre variáveis, validando-os através de padrões detectados a novos subconjuntos de dados. O processo consiste basicamente em 3 etapas: exploração; construção de modelo ou definição do padrão; e validação/verificação (CÔRTEZ, 2006). De tal modo, *data mining* possibilita a descoberta de padrões ou regras que permitam uma melhor compreensão da informação contida nos dados (VESSONI, 2006).

O banco de dados biológico mais utilizado é o GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>), disponível no banco de informações do NCBI (*National Center of Biotechnological Information*) (<http://www.ncbi.nlm.nih.gov>), o qual oferece a coleção mais completa de dados de seqüências nucleotídicas de vários tipos de organismos disponíveis em todo o mundo. O banco disponibiliza diversas ferramentas computacionais com funções específicas, sendo uma das mais utilizadas o BLAST (*Basic Alignment Search Tool*), um algoritmo baseado no método heurístico para a submissão de seqüências e posterior alinhamento, através de homologia dessas seqüências com as depositadas no GenBank. O BLASTn realiza pesquisas em seqüências de nucleotídeos com banco de dados de seqüências também de nucleotídeos (GIBAS; JAMBECK, 2001). O ordenamento se dá por uma análise de trechos semelhantes entre os fragmentos seqüenciados. Quando são encontrados trechos com pelo menos 14 bases semelhantes, eles são interpretados como sendo originários de uma mesma região do genoma. É desta forma que dois fragmentos são alinhados contiguamente, tendo como finalidade principal gerar novos conhecimentos a partir da grande quantidade de dados que vêm sendo obtidos através das seqüências de DNA e proteínas.

O projeto genoma da cana-de-açúcar (SUCEST) produziu uma quantidade enorme de seqüências de DNA. O banco de dados que armazena essas seqüências automaticamente analisa as proteínas que possivelmente são codificadas por cada uma dessas seqüências. Essas proteínas podem ser categorizadas de acordo com diversos critérios: proteínas induzidas por ataque de patógenos, proteínas que se ligam a DNA, etc. O *data mining*

consiste em buscar no banco de dados genes que codifiquem proteínas possivelmente envolvidas em um determinado processo biológico (MENOSSE, 2002).

Os principais bancos de dados encontrados na Internet e que são de interesse em bioinformática são bancos de dados envolvendo seqüências de genes e proteínas: os bancos de seqüências primários e os secundários. Os primários são aqueles que são derivados diretamente dos dados obtidos a partir do seqüenciamento de ácidos nucleicos ou proteínas. Os bancos de dados primários podem conter, além da seqüência em si, outros dados como uma tradução de uma seqüência de um clone de DNA, seqüências padrão (como sítios de fosforilação), promotores e outras anotações semelhantes. Os principais bancos de dados primários são o Genbank, o EBI(EMBL), e o DDBJ. Além dos bancos de dados primários, há outros que são deles derivados. São os bancos de dados secundários, como o **PDB** - http://www.rcsb.org/pdb/Welcome.do;jsessionid=1V20vIPwj2Csu3lLjBDV8Q** (banco de dados de estruturas tridimensionais de macromoléculas biológicas, determinadas por cristalografia de raios X e por ressonância magnética nuclear), o **PIR** - <http://www-nbrf.georgetown.edu/> (seqüência de aminoácidos em proteínas e outras anotações pertinentes), **PROSITE** - <http://www.expasy.ch/prosite/> (um banco de dados de famílias e domínios de proteínas), o **SWISS-PROT** - <http://www.expasy.ch/sprot/> (seqüências de proteínas, contendo diversas anotações) e ainda o **KEGG** - <http://www.genome.jp/kegg/> (Enciclopédia de genes e genomas).

METODOLOGIA

1. Local de execução

O presente estudo foi conduzido no Laboratório de genética Molecular, Genômica e Proteômica – GEMPRO e BIOINFORMAL- localizados no Centro de Ciências Agrárias da Universidade Federal de Alagoas – CECA/UFAL, Campus Delza Gitaí, Rio Largo, Alagoas.

2. Material biológico

O material utilizado foi constituído de quinze placas de bibliotecas genômicas (Tabela 2) fornecidas pelo Laboratório CBMEG da Universidade de Campinas/SP – UNICAMP, contendo os fragmentos de cDNA de cana-de-açúcar inseridos no vetor *pSport* em *Escherichia coli*, preservadas em glicerol 50% e conservadas a -20°C .

As bibliotecas de cDNA foram obtidas de diferentes variedades de cana em diferentes estágios de crescimento, cultivadas na estação experimental da Copersucar (Piracicaba – SP), CBMEG (Universidade Estadual de Campinas / Campinas – SP), Universidade Federal do Rio de Janeiro (Rio de Janeiro – RJ) e Serra do Ouro – Murici, Universidade Federal de Alagoas (Maceió – AL).

Tabela 2. Números de identificação das bibliotecas genômicas utilizadas para seqüenciamento.

NÚMEROS DE IDENTIFICAÇÃO	
NR1 049 RII	NR2 049 RII
NR1 050 RII	NR2 050 RII
NR1 051 RII	NR2 051 RII
NR1 052 RII	NR2 052 RII
NR1 053 RII	NR2 053 RII
NR1 054 RII	NR2 054 RII
NR1 055 RII	NR2 055 RII
	NR2 056 RII

As Bibliotecas foram obtidas de RNA isolado de variedades de cana-de-açúcar comercialmente cultivadas (SP70-1143; SP80-3280²; SP80-87432¹; PB5211 x P57150-4¹; SP83-5077; SP80-185; SP87-396; SP803280 x SP81-5441¹; CB47-89; RB855205; RB845298; RB805228).

3. Crescimento bacteriano

As colônias de bactérias com o plasmídeo foram multiplicadas para obtenção de grande quantidade de cópias do cDNA dos fragmentos a serem seqüenciados. Esta etapa foi desempenhada numa câmara asséptica de fluxo laminar (*Labconco Purifier Class II Total Exhaust*), previamente desinfetada com etanol 70% e esterilizada por luz ultravioleta (UV) por 15min. Com o auxílio do replicador *Boekel* os clones de cada uma das placas com glicerol a 50% foram transferidos para uma placa de crescimento contendo 1,2mL de CircleGrow 4%, 120µL de ampicilina (100mg/mL) (Figura 1). As placas foram identificadas, seladas com adesivo (Sorenson), o qual foi perfurado em dois pontos para facilitar a aeração. As bactérias foram, portanto, incubadas sob agitação constante de 320 rpm em agitador durante 24h à 37°C (Figura 2).

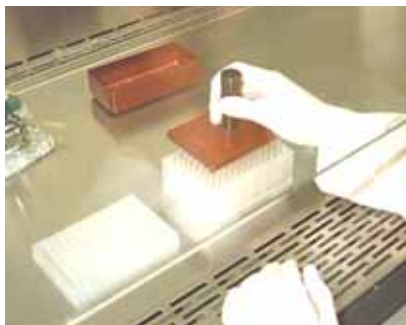


Figura 1. Transferência de clones com replicador *Boeckel*.

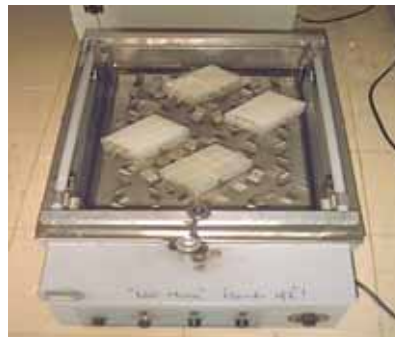


Figura 2. Culturas de bactérias em placas de crescimento sob agitação em agitador.

Após incubação as placas foram centrifugadas por 6min. a 4000 rpm para sedimentar as células. Após conferir e registrar em micro *prep sheet* (Apêndice 1) possíveis anormalidades nas culturas, o adesivo foi removido, o sobrenadante descartado e a placa invertida sobre papel absorvente por 1 min, em seguida novamente selada e mantida a -20°C até a extração do plasmídeo (mini-prep).

4. Extração de DNA plasmidial (Mini-prep)

Esta etapa consistiu em eliminar todos os constituintes celulares da bactéria, que não interessam para o seqüenciamento, do plasmídeo contendo o inserto de cana-de-açúcar. Para tal, foram adicionados a cada poço da placa contendo a cultura bacteriana, 240µL de GET (20% Glicose, 0,5M EDTA, 1M Tris-HCl, pH 7,4), a placa foi selada e agitada em Vortex por 2 mins para ressuspender bem as células. Em seguida as células foram sedimentadas por centrifugação por 6min. a 4000rpm em centrífuga refrigerada (*Eppendorf*). O adesivo (SORENSEN) foi removido, o sobrenadante descartado e a placa invertida em papel absorvente. Em seguida foram adicionados a cada poço 80µL de GET, a placa foi novamente selada e agitada em Vortex por 2min. para ressuspender as células.

Concomitantemente foram adicionados 5µL de RNase (10mg/mL) a cada poço de uma microplaca de 250µL em polipropileno com fundo redondo (tipo Elisa) devidamente identificada. Para esta placa contendo RNase foram transferidos 80µL da suspensão de células e adicionados 80µL de 0,2mol.L⁻¹ NaOH / SDS. A placa foi selada com um novo adesivo e o conteúdo misturado gentilmente por inversão 10 vezes consecutivas. A placa foi incubada por 10 mins a temperatura ambiente sobre a bancada e em seguida centrifugada por alguns segundos (*spin*) até que não houvesse mais solução no adesivo. Logo após foi acrescentado a cada poço, 80µL de KOAc 3mol.L⁻¹ (estocado a 4°C), a placa novamente foi selada e invertida 10 vezes para homogeneizar. Mais uma vez foi dado um *spin* a 4000 rpm para remover gotículas do adesivo, o qual foi logo removido e a placa transferida para uma estufa previamente aquecida a 90°C por exatos 30min. Após esse período de incubação a placa foi selada e arrefecida em gelo picado por 10 mins e posteriormente centrifugada por 4min., 4000 rpm a 20°C. Uma placa de filtro Millipore (MAGV N22) foi fixada com fita adesiva em uma microplaca nova de 96 poços (fundo V), assegurando-se que os filtros e poços receptores estivessem bem alinhados. Transferiu-se todo o volume do sobrenadante (evitando transferir os *debris* celulares, ou seja, sem perturbar o *pellet*) para a placa filtro Millipore e centrifugou-se (sem tampa) por 4 min. a 4000rpm, 20°C ou até que todo o volume descesse para a outra placa (fundo V). Após a transferência do líquido, a placa Millipore foi removida e descartada e foram adicionados 110µL de isopropanol 100% (Merck) ao filtrado, após a placa foi selada com adesivo resistente a álcool e invertida por 20 vezes e centrifugada por 45min., 4000rpm, 20°C. O sobrenadante foi descartado, 200µl de etanol 70% gelado foram adicionados para precipitação do DNA. Centrifugou-se por mais 5 min. a 4000rpm, 20°C. Removeu-se o sobrenadante e invertendo-se a placa sobre papel absorvente e centrifugando por 3min. a 900rpm, 20°C. O DNA secou por 60min. à temperatura ambiente, na placa coberta com papel toalha e em seguida o DNA foi ressuspendido com 60µL de água deionizada (Milli-Q) estéril. A placa selada com novo adesivo foi devidamente identificada e incubada sobre bancada a temperatura ambiente *over night*. Após esse período foram conservadas a -20°C.

5. Análise do DNA plasmidial

Uma alíquota de 5 μ L de DNA extraído com 1 μ L de tampão de carregamento (0,25% azul de bromofenol, 0,25% de xileno cianol e 30% de glicerol) foi analisada através da eletroforese em gel de agarose 1,2%, em tampão TBE 1X (500 mM de Tris-HCl, 60 mM de ácido bórico e 83 mM de EDTA) e corado com brometo de etídio (0,5 μ g/mL), submetido a 3cm⁻¹ por 60 min., visualizado em transiluminador de luz UV (TFX-20M Life Technologies – Gibco BRL) e fotodocumentado com Sistema Digital Kodak Science.

Depois de constatada a eficiência da minipreparação, as amostras puderam ser submetidas à reação de PCR.

6. Reação de PCR (*polimerase chain reaction*)

As reações de seqüenciamento foram preparadas com um quarto do volume recomendado pelo kit ABI PRISM *BigDye terminator Sequence* (Applied Biosystems, Foster City, CA, EUA), acrescentando-se: 2 μ L *BigDye terminator*, 2 μ L do tampão *Save Money* (200mM Tris-HCl pH 9,0; 5mM MgCl₂; H₂O), 1 μ L do oligonucleotídeo T7 (*forward*) (GibcoBRL/Life Technologies), de seqüência (5'- 3'): TAA TAC GAC TCA CTA TAG GG [3.2pMol], 1 μ L DNA [100ng] molde dos clones EST, completando o volume final para 10 μ L com água Milli-Q^{UF} para cada reação. A reação de seqüenciamento foi estabelecida com o seguinte perfil: desnaturação inicial a 96°C por 90 segundos; 25 ciclos de 96°C por 12 segundos, anelamento a 52°C por 6 segundos e extensão a 60°C por 2 min; concluindo a extensão por 4min. a 60°C e finalizando a 4°C. As reações de amplificação foram conduzidas em um termociclador PTC – 200 (*Peltier Thermal Cycler / MJ Research*) para 96 tubos/microplacas.

7. Precipitação e purificação

Para cada poço contendo 10µL da reação de PCR, foram adicionados 90µL de isopropanol 60% (60 µL de isopropanol + 30 µL de água Milli-Q) resultando um volume final de 100µL em cada poço da placa. A placa foi invertida manualmente 5 vezes e deixada em repouso sobre a bancada por 15 min. Logo após foi centrifugada por 40min. a 2000g a 20°C e descartou-se o sobrenadante. Para remover o excesso de álcool a microplaca foi centrifugada em posição invertida sobre papel toalha por 1min., 500g a 20°C. Um volume de 150µL de isopropanol 75% foi adicionado para lavar o *pellet* por inversão da placa 5 vezes seguidas e após centrifugação a 2000g por 10 mins descartando o sobrenadante. As quatro últimas etapas foram repetidas para uma segunda lavagem do precipitado, deixando-os em seguida, secar por 1 hora a temperatura ambiente. Possíveis vestígios de álcool foram eliminados incubando a placa no termociclador (PTC-200) a 90°C por 1 min.

8. Desnaturação

Os DNAs foram ressuspensos em 2µL de tampão de seqüenciamento (80% (v/v) formamida deionizada; 20% de *loading-buffer Blue Dextran*). Posteriormente a placa foi incubada em termociclador a 96°C por 2min. e em seguida esfriada em gelo picado, ficando assim, prontas para o carregamento em gel de poliacrilamida para análise no seqüenciador automático.

9. Seqüenciamento automático

Os fragmentos de DNA marcados (ESTs de cana-de-açúcar) com fluoróforos foram analisados por eletroforese em gel de poliacrilamida, em seqüenciador automático ABI PRISM 377 DNA SEQUENCER (Figura 3), da *Applied Biosystems*.



Figura 3. ABI PRISM 377 DNA SEQUENCER (Applied Biosystems).

Um conjunto de placas de 36cm para seqüenciamento foram lavadas utilizando Alconox 1,5% e água Milli-Q a 90°C. Após secarem, as mesmas foram montadas em cassette para ABI PRISM 377. Os géis para seqüenciamento foram preparados em um Backer utilizando-se: 9g de uréia, 2,5mL de Gel *stock solution* (Applied Biosystems), 2,5mL de TBE 10X e 15,75mL de água Milli-Q autoclavada, homogeneizado manualmente e em seguida filtrados em bomba vácuo. Após a filtração, foram adicionados 125 µL de APS 10% e 17,5 µL de TEMED, homogeneizando-os bem. O gel foi aspirado com seringa de aplicação e injetados nas placas. Deixou-se polimerizar por trinta mins as extremidades hidratadas utilizando papel macio molhado com água Milli-Q. Após duas horas de polimerização, o cassette montado foi colocado no seqüenciador automático ABI PRISM 377, o equipamento foi ligado e em seguida o computador. Para avaliação do gel foi feito *plate check* (Apêndice 2) e comandos de seqüenciamento, utilizou-se o *software collection*. Constatadas as qualidades ótimas do gel e do grupo de placas, deu-se seguimento a montagem completa do equipamento e aplicação das amostras (Figura 4).



Figura 4. A) ABI PRISM 377 pronto para seqüenciamento; B) Detalhe da etapa de carregamento das amostras no gel de eletroforese no seqüenciador automático com auxílio aplicador multicanal.

Os parâmetros (Tabela 3) utilizados no seqüenciamento foram determinados em treinamento promovido pela *Applied Biosystems*. Neste foram abordados todos os aspectos teóricos e práticos que envolvem o uso do equipamento (*hardware e software*), aplicação das amostras, criar *sample sheet* (Apêndice 3), bem como a análise dos dados obtidos.

Tabela 3. Parâmetros eletroforéticos de seqüenciamento ABI PRISM 377

Tipo de corrida	<i>Standard (2X Run Module) 3.0 scans/seg</i>
Velocidade (scans/hora)	100 bases/hora (1.200 <i>scans/ hora</i>)
Well to read	36 cm
Voltagem	1.680V
Tempo de corrida	7 horas
Leitura	~550 bases

10. Análise do gel

Os géis obtidos foram analisados no programa *Sequencing Analysis (Software v5.2 Upgrade from v5.1.1)* que corrige as distorções de eletroforese determinando corretamente as posições das amostras na imagem do gel (Figura 5). Tais distorções podem diminuir a interpretação da informação contida no gel.

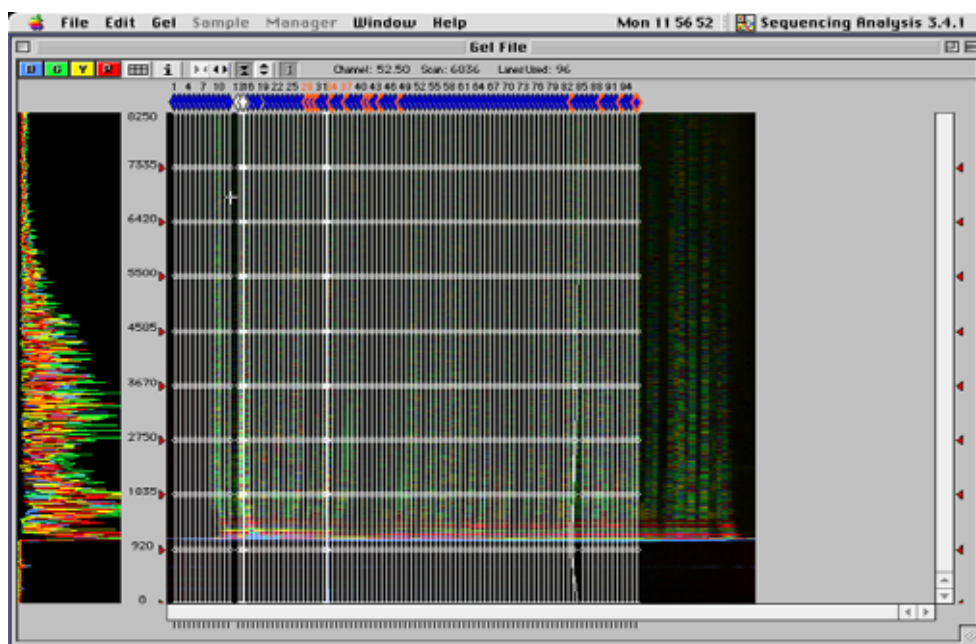


Figura 5. Alinhamento do gel em programa *Sequencing Analysis*.

11. Submissão das seqüências ao banco de dados SUCEST – UNICAMP

Para a submissão on-line das seqüências produzidas no seqüenciamento, os arquivos foram compactados em programa pKZip, selecionando apenas as seqüências e deixando os arquivos correspondentes ao gel, *log run e sample sheet*. Em seguida a página do SUCEST (<http://sucest.lad.ic.unicamp.br>) foi acessada e selecionada a opção *submission reads*. A identificação posterior de cada seqüência EST do projeto foi realizada por meio de código, que representa o laboratório que a submeteu: biblioteca de cDNA de origem; a placa de

clones; a posição na placa e sentido de seqüenciamento (“.g” para 5’ ou “.b” para 3’). O código ainda inclui as iniciais “SC” que corresponde a *sugarcane*. Como por exemplo: “RTI” associada à biblioteca de onde foi obtido o cDNA, neste caso de raízes de cana; “028” corresponde ao número da placa desta biblioteca; “B07” indica a posição na placa e “.g” o sentido do seqüenciamento (5’). As seqüências EST foram analisadas e agrupadas em função da identidade, gerando um consenso para cada agrupamento de seqüências, denominado *cluster*, correspondendo a um gene presumível.

12. Mineração de dados (*data mining*)

A mineração ou garimpagem por genes relacionados à resistência das plantas ao estresse hídrico foi feita na base de dados do GenBank (<http://www.ncbi.nlm.nih.gov>) usando-se ferramentas a busca por palavra chave e FASTA (Conjunto de programas desenvolvidos para procedimentos rápidos de análise de seqüências de DNA). As seqüências da base de dados do SUCEST - UNICAMP foram alinhadas usando algoritmos da família BLAST (Tabela 4) onde, as porcentagens da identidade e da similaridade nestes alinhamentos foram calculadas. O valor de "E" (*e-value*), parâmetro estimado pelo BLAST, expressa a dificuldade para encontrar uma seqüência perfeitamente idêntica, ou seja, quanto menor o valor, menor a chance de tal comparação ter sido encontrada por pura coincidência. Identidades de homologia significativas têm de apresentar um *score* > 200 para esse programa. *Scores* mais baixos podem indicar, eventualmente, baixa homologia, podendo ser classificada como pouco significativa. Os conjuntos foram aceitos como relacionado a um gene particular ao apresentar *e-value* abaixo de 10^{-5} .

O programa BLAST desenvolvido pelo NCBI é um instrumento de análise e identificação de genes. Pode executar busca de seqüências em todos os bancos de dados de DNA em menos de 15 segundos encontrando os melhores *scores* de alinhamentos locais entre a sua seqüência e o banco de dados; baseado em programas que alinham seqüências

não espaçadas. O *score* indica o grau de homologia existente entre as seqüências analisadas e o *score* considerado bom é maior / igual que 200.

Tabela 4. Descrição das ferramentas da família BLAST.

Família de programas BLAST:	
<u>blastp:</u>	compara uma seqüência de aminoácido vs banco de dados de proteínas
<u>blastn:</u>	compara uma seqüência de nucleotídeos vs banco de dados de nucleotídeos
<u>blastx:</u>	compara seis fases de leitura de um nucleotídeo vs banco de dados de proteína
<u>tblastx:</u>	compara todas as seis fases de leitura abertas da seqüência de nucleotídeos vs banco de dados de nucleotídeos

As etapas seguidas na mineração consistiram das seguintes fases:

- a) No navegador Internet Explorer foram abertos os endereços correspondentes ao *sites* do NCBI (www.ncbi.nlm.nih.gov) e do SUCEST (<http://sucest.lad.ic.unicamp.br>) (Figura 6).

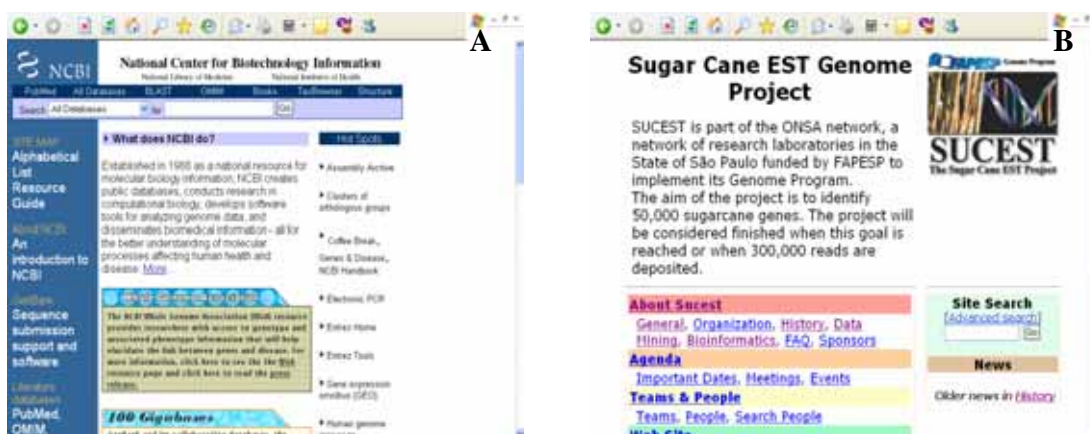


Figura 6. A) Página inicial do NCBI; B) página inicial do SUCEST.

- b) No NCBI, palavras chaves foram utilizadas para as buscas por seqüências homólogas para resistência ao estresse hídrico (Figura 7).

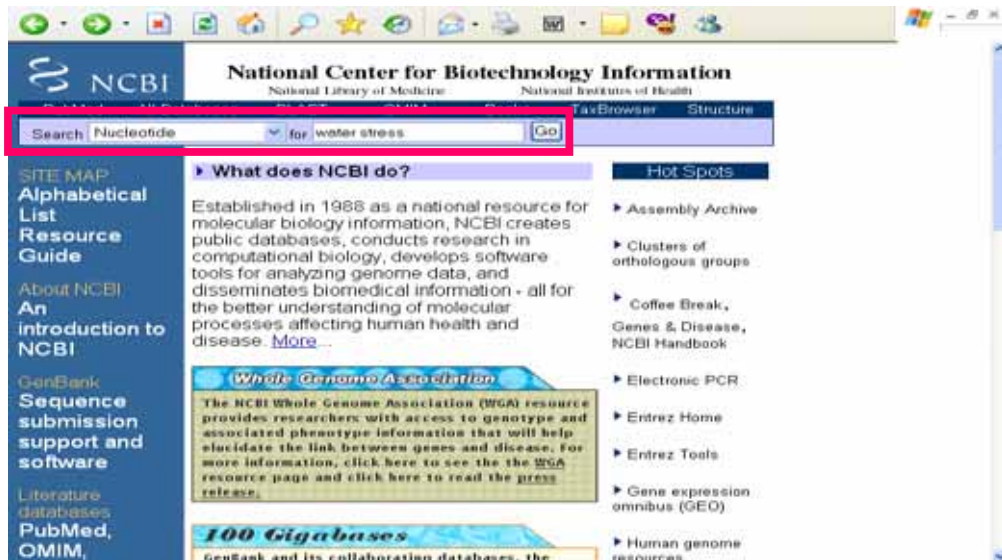


Figura 7. Busca inicial no NCBI - keyword: water stress (em destaque)

- c) Nos resultados da busca selecionou-se a(s) seqüência (s) de interesse e utilizando a ferramenta FASTA, as seqüências foram convertidas ao modo fasta para viabilizar a submissão dos dados no banco do SUCEST. No modo FASTA a seqüência passa a ter um sinal de > na linha anterior, depois deste sinal, na mesma linha, vem a identificação (Figura 8).



Figura 8. Seqüência no modo FASTA.

A seqüência do NCBI no modo FASTA é copiada e transferida para o banco do SUCEST para submissão (Figura 9)

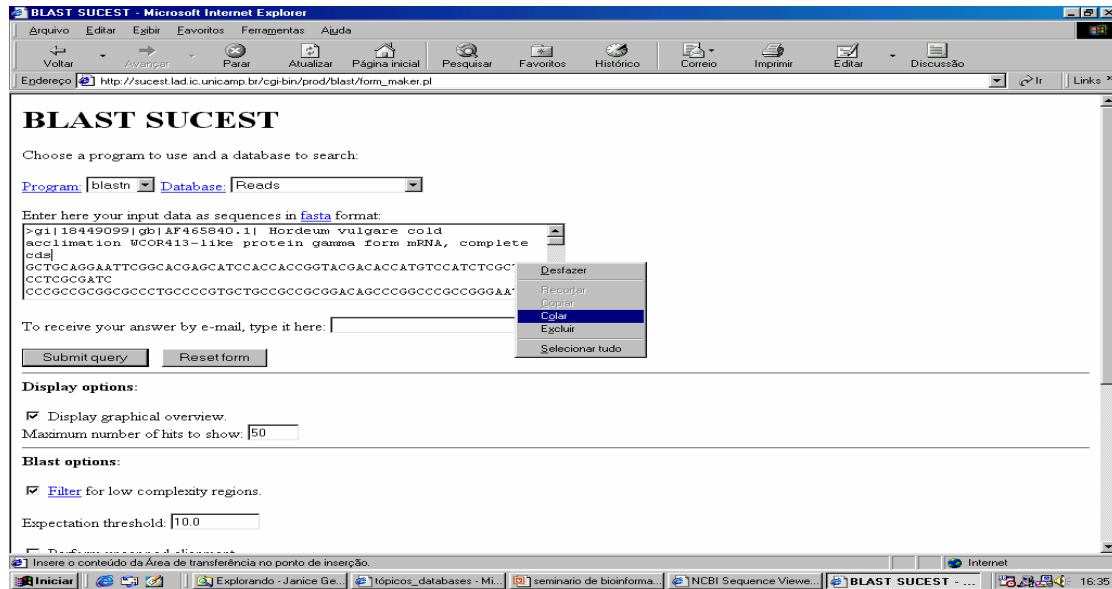


Figura 9. Transferência da seqüência FASTA para submissão ao BLAST no SUCEST.

RESULTADOS E DISCUSSÃO

Etiquetas de seqüências expressas, se constituem numa estratégia eficiente para a descoberta de genes e dos componentes resultantes de sua expressão gênica. Isso é possível devido a freqüência de ESTs em bibliotecas de cDNA, que refletem os níveis de expressão de genes nos tecidos que originaram RNA (OKUBO, et al. 1992). Análises similares *in silico* foram realizadas em transcriptomas de *Eucalyptus* (VICENTINE et al., 2005), em arroz (EWING et al. 1999) *Medicago truncatula* (JOURNET et al. 2002) e *Schistosoma mansoni* (FRANCO et al. 1997). Atualmente, mais de 10 milhões de ESTs podem ser encontradas e analisadas no sub-banco dbEST do GenBank (www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). O conjunto de dados gerados pelo SUCEST, correspondendo a 237.954 ESTs, permite uma visão panorâmica preliminar do perfil de expressão gênica em cana-de-açúcar. A produção de ESTs em larga escala poderá ser extremamente útil nos futuros trabalhos de averiguação de mecanismos moleculares de desenvolvimento e função de genes.

Essa tecnologia tem se revelado eficiente e de baixo custo nos procedimentos de identificação de genes que são expressos em determinados tecidos e células de organismos multicelulares (ADAMS et al., 1991a, HWANG et al., 1997, LIEW et al. 1994, ADAMS et al., 1995b).

Neste trabalho foram estudadas bibliotecas geradas a partir dos principais órgãos da planta em diferentes estágios de desenvolvimento, gerando uma das mais completas bibliotecas de cDNA, provenientes de aproximadamente 12 variedades de cana-de-açúcar cultivadas no território brasileiro.

1. Extração de DNA plasmidial (mini-prep)

Através do método de mini preparação de plasmídeos foram obtidos DNA plasmidial de alto peso molecular. Alíquotas de 5 μ L de DNA foram aplicadas em gel de agarose 1,2% corado com brometo de etídio, e após eletroforese (4V.cm⁻¹), o material genético foi visualizado em transiluminador UV, obtendo-se DNA alta qualidade em todos os clones da biblioteca avaliada (Figura 10).

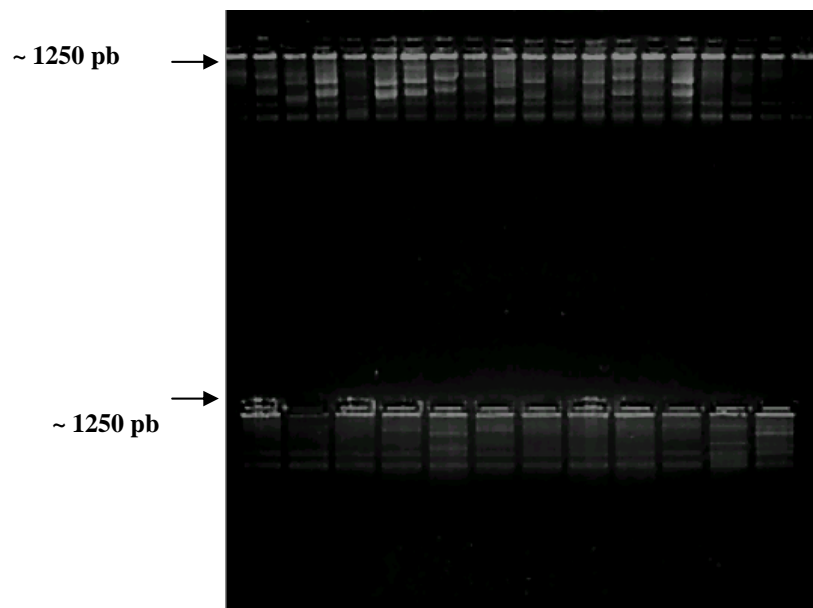


Figura 10. Gel de agarose 1,2% corado com brometo de etídio da biblioteca plasmidial NR1.051 RII.

2. Seqüenciamento de EST da cana-de-açúcar

O seqüenciamento automático realizado em Alagoas, a partir dos plasmídios, produziu seqüências de ESTs individuais de clones de cana-de-açúcar com cerca de 550 nucleotídeos, totalizando 52.800 bases lidas e 1440 *reads* para cada placa de biblioteca de cDNA seqüenciada. A cada seqüenciamento, eletroferogramas foram gerados [representações gráficas da intensidade de fluorescência emitida por cada nucleotídeo de DNA seqüenciado, indicando a confiabilidade dos dados obtidos (Apêndice 4)], e arquivos da imagem do gel e seqüência das bases (Apêndices 5 e 6). As seqüências foram submetidas ao banco do SUCEST ficando assim, disponíveis para a análise de homologia utilizando as ferramentas de bioinformática.

A eletroforese dos fragmentos obtidos com o seqüenciamento automático foi visualizada em plataforma *Apple Mac*. Os géis de poliacrilamida após a eletroforese apresentaram imagens contendo a ilustração das bases no gel e o eletroferograma (Figura 11).

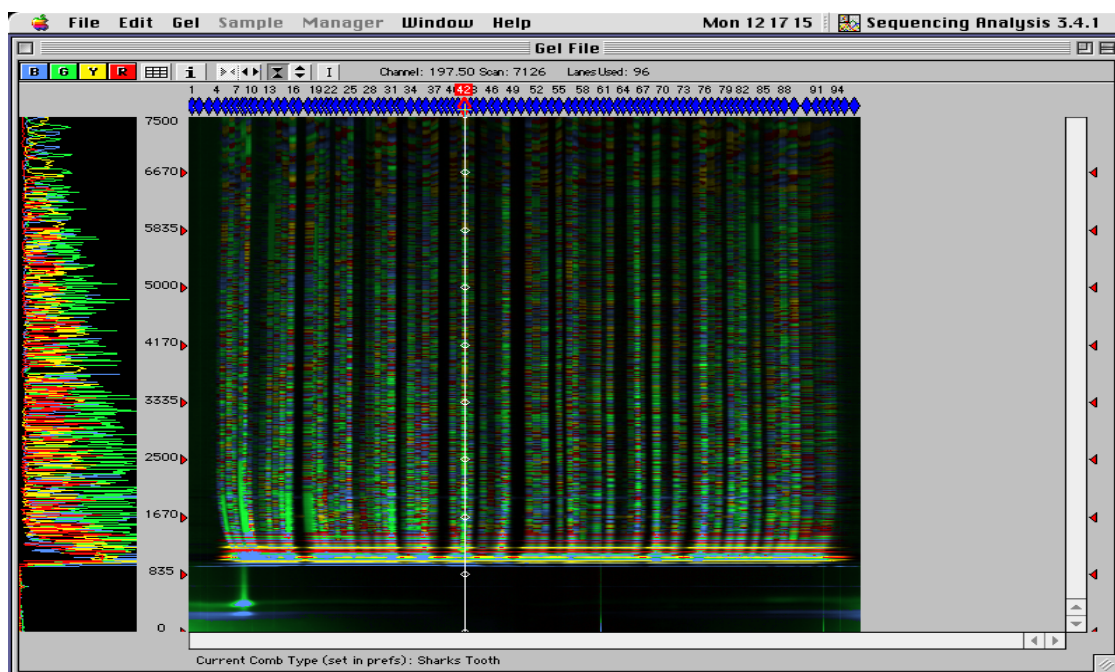


Figura 11. Imagem do gel de eletroforese após seqüenciamento em ABI PRISM 377.

3. Mineração de dados (*data mining*)

Além das seqüências obtidas no GEMPRO, foram avaliadas as depositadas no bando do projeto genoma da cana-de-açúcar (*Sugarcane* EST Project – SUCEST) que analisou 237.954 seqüências de ESTs provenientes de 26 bibliotecas de cDNA produzidas de diversos tecidos de cana. O SUCEST gerou cerca de 43.150 *clusters* (GRIVET e ARRUDA, 2001 e ALMEIDA, 2002) onde se realizaram, no caso do presente trabalho, buscas por genes de resistência ao estresse hídrico.

A mineração ou garimpagem de seqüências de DNA no banco de dados do SUCEST análogas a outras similares as de cana-de-açúcar realizada neste estudo, possibilitou a identificação de 2998 *clusters* com homologia a 52 proteínas putativas que relacionadas com resistência ao estresse hídrico na planta.

Os alinhamentos considerados apresentaram *e-values* abaixo de 10^{-5} e *scores* iguais ou maiores que 200, como mostra a figura abaixo (Figura 12). No alto da figura está a escala de cor usada para alinhamentos e abaixo está a seqüência submetida representada por uma linha preta espessa. O primeiro e último número de bases estão em azul. As seqüências de cana-de-açúcar alinhadas com a que está sendo pesquisada estão em cinza. Os números de bases da seqüência alinhada de encontro com a que está em questão são coloridas de acordo com a contagem do alinhamento. As linhas em vermelho indicam maior homologia, onde o *score* é igual ou maior que 200.

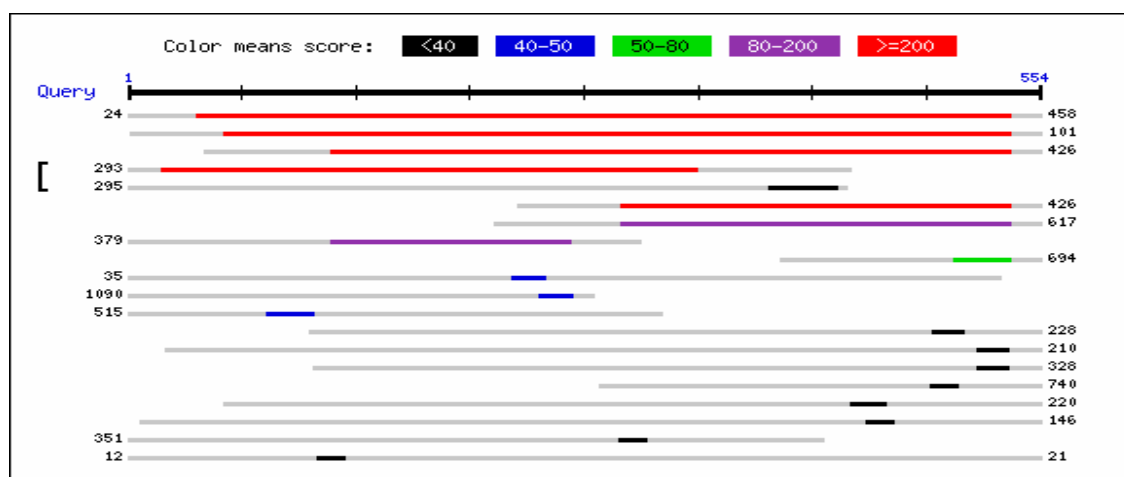


Figura 12. Gráfico de homologia entre seqüências após submissão à ferramenta BLAST.

Dessa forma, os resultados obtidos com as buscas e comparações entre os dados do SUCEST disponível no sítio eletrônico da FAPESP e o banco de domínio público NCBI permitiram a identificação dos seguintes polipeptídios:

1. 1-aminocyclopropane-1-carboxylate oxidase
2. 1-aminocyclopropane-1-carboxylate synthase
3. 60S ribosomal protein L10
4. Abscisic acid-inducible protein kinase [Fragment]
5. ADP-glucose pyrophosphorylase [Fragment]
6. Alpha amylase [Precursor] [Fragment]
7. At1g56600/F25P12_16
8. AtABA1 protein
9. AtPIP5K1
10. cDNA5' similar to water stress inducible
11. Chloroplastic drought-induced stress protein CDSP-34
12. Cold acclimation protein WCOR413-like protein beta form
13. CpABA1 protein
14. Cytosolic pyruvate orthophosphate dikinase
15. DNAK -Type molecular chaperone Bip precursor (similarity)
16. D-TDP-glucose dehydratase
17. Endoplasmic reticulum HSC70-cognate binding protein_[Precursor]
18. Endoxyloglucan transferase
19. Epoxide hydrolase
20. F14J9.1 protein
21. F8A5.2 protein
22. Galactinol synthase [Fragment]
23. Glycine-rich RNA-binding, abscisic acid-inducible protein
24. Hypothetical 41.9 kDa protein
25. Lactoylglutathione lyase
26. Mitogen-activated protein kinase

27. mRNA for putative ubiquitin conjugating enzyme (UBC gene)
28. Neoxanthin cleavage enzyme
29. Phospholipase D2
30. Phospholipase D1
31. Plasma membrane intrinsic protein 2C
32. Polygalacturonase
33. Protein beta form
34. Protein kinase AKINbetagamma-1
35. Protein kinase AKINbetagamma-2
36. Putative cold acclimation protein
37. Putative sucrose-phosphate synthase [Fragment]
38. Ripening – associated membrane protein (clone Pny 507)
39. Sequence 4 from patent W00183756
40. Sequence 5 from patent W00183756
41. Similar to rice water stress induced protein
42. T13D8.32 protein
43. Tubulin
44. Tubulin (Rip3)
45. Tubulin beta chain [Fragment]
46. Water stress-induced ER5 protein
47. Water stress-induced protein (WSI729)
48. Water-stress induced tonoplast intrinsic protein (*Pisum sativo*)
49. Water-stress induced tonoplast intrinsic protein(Tomato)
50. Wilm's tumor suppressor homolog
51. WSI724 protein induced by water stress
52. WSI76 protein induced by water stress

Os alinhamentos de seqüências similares dentro do SUCEST foram variáveis quanto ao número de *clusters* homólogos para cada um dos polipeptídios identificados. No total, 2.998 *clusters* relativos às seqüências análogas estudadas foram distribuídos respectivamente com as proteínas identificadas, de acordo com a tabela 5.

Tabela 5. *Clusters* homólogos relativos às seqüências análogas das proteínas relacionadas à resistência ao estresse hídrico.

CLUSTERS HOMÓLOGOS	BEST HIT (PROTEÍNAS PUTATIVAS)
1722	TIPW_PEA P25794 WATER-STRESS INDUCED TONOPLAST INTRINSIC PROTEIN
2	GALACTINOL SYNTHASE
8	WATER STRESS-INDUCED PROTEIN, PUTATIVE, PREDICTED
30	PUTATIVE UBIQUITIN-CONJUGATING ENZYME (UBC GENE)
77	TUBULIN (RIP3)
99	D-TDP-GLUCOSE DEHYDRATASE
41	PROTEIN KINASE AKINGETAGAMMA -2
45	PROTEIN KINASE AKINBETAGAMMA-1
43	COLD ACCLIMATION PROTEIN WCOR413-LIKE PROTEIN
22	CDNA5' SIMILAR TO WATER STRESS INDUCIBLE PROTEIN
27	WATER STRESS-INDUCED PROTEIN (WSI729
27	ENDOPLASMIC RETICULUM HSC70-COGNATE BINDING PROTEIN PRECURSOR (BIP)
32	ADP-GLUCOSE PYROPHOSPHORYLASE (WAL2)
5	ADP-GLUCOSE PYROPHOSPHORYLASE (WAL1)
14	TOMATO RIPENING MEMBRANE PROTEIN, CLONE PTOM75
26	TOMATO RIPENING MEMBRANE PROTEIN, CLONE PNY507
49	GLYCINE-RICH PROTEIN
41	WATER STRESS INDUCED PROTEIN
2	WATER STRESS-INDUCED PROTEIN, PUTATIVE
28	PROTEIN KINASE
229	INTRINSIC PROTEIN 2C (WATER-STRESS INDUCED TONOPLAST INTRINSIC PROTEIN)
7	D-TDP-GLUCOSE DEHYDRATASE
2	IDENTICAL TO RICE WATER STRESS INDUCE PROTEIN
2	NEOXANTHIN CLEAVAGE ENZYME
3	POLYGALACTURONASE
4	PUTATIVE PROTEIN
2	GALACTINOL SYNTHASE
39	WSI76 PROTEIN INDUCED BY WATER STRESS
7	ALPHA AMYLASE PRECURSOR
6	MITOGEN-ACTIVATED PROTEIN KINASE

12	PYRUVATE ORTHOPHOSPHATE DIKINASE
78	BETA-TUBULIN
1	FIBRILLIN HOMOLOG CDSP34 PRECURSOR
6	PUTATIVE SUCROSE SYNTHASE
7	PHOSPHOLIPASE D1
7	PHOSPHOLIPASE D2
17	TUBULIN BETA CHAIN
1	LGUL_LYCES LACTOYLGLUTATHIONE LYASE
5	CYTOSOLIC PYRUVATE ORTHOPHOSPHATE DIKINASE
1	CHLOROPLAST DROUGHT-INDUCED STRESS PROTEIN
12	PUTATIVE IMBIBITION PROTEIN
50	ENDOXYLOGLUCAN TRANSFERASASE
4	GLYOXALASE I
39	1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE
85	WATER STRESS-INDUCED ER5 PROTEIN
5	PUTATIVE UBIQUITIN-CONJUGATING ENZYME
27	ENDOPLASMATIC RETICULUM HSC70-COGNATE BINDING PROTEIN PRECURSOR

4. Identificação eletrônica das seqüências

A identificação das proteínas foi obtida obedecendo-se a um protocolo de buscas desenvolvido no BIOINFORMAL (Laboratório de Bioinformática da Universidade Federal de Alagoas), apresentado a seguir:

1. Página inicial do banco de dados do NCBI (Figura 13): Nessa página pode se encontrar as informações e diretrizes para a busca e identificação de seqüências de nucleotídeos ou aminoácidos eventualmente relacionadas com características fenotípicas de interesse agroindustrial.

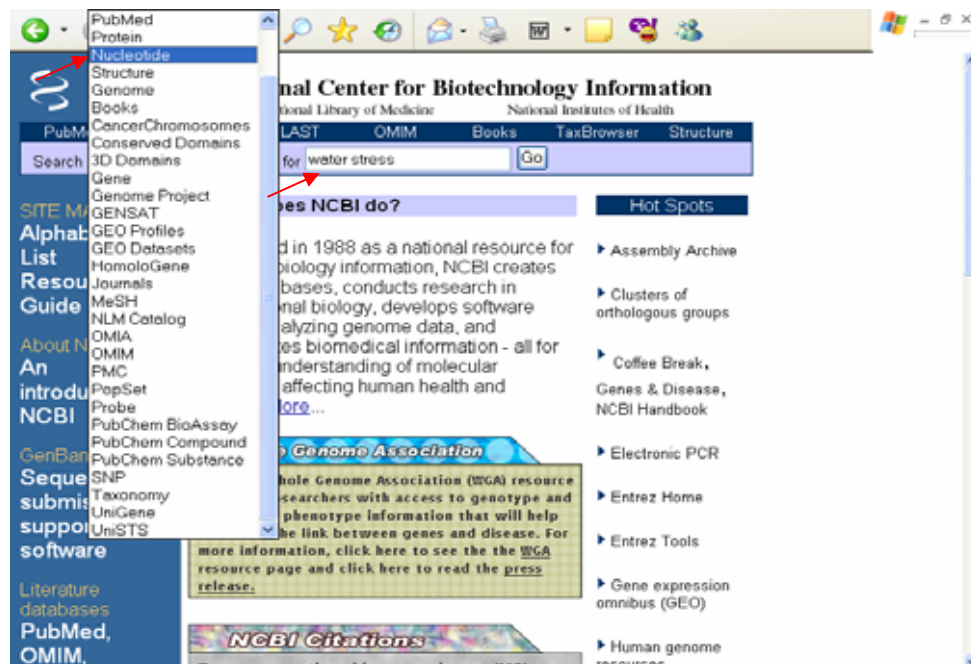


Figura 13. Página inicial do NCBI apresentando a indicação do tipo de busca (nucleotídeos) e palavra chave (*water stress*) utilizada na mineração de dados (seta).

A conclusão dos procedimentos de busca no banco de nucleotídeos do NCBI, relacionados com a palavra chave *water stress* resultou em 12.647 hits (Figura 14).



Figura 14. Resultados da busca no banco de nucleotídeos do NCBI utilizando a palavra chave *water stress* no campo de pesquisa (*search*) (seta).

A próxima etapa correspondeu a escolha progressiva dos itens resultantes da busca e a ulterior seleção da ferramenta FASTA com vistas ao processamento dos dados disponíveis no banco SUCEST (Figura 15).

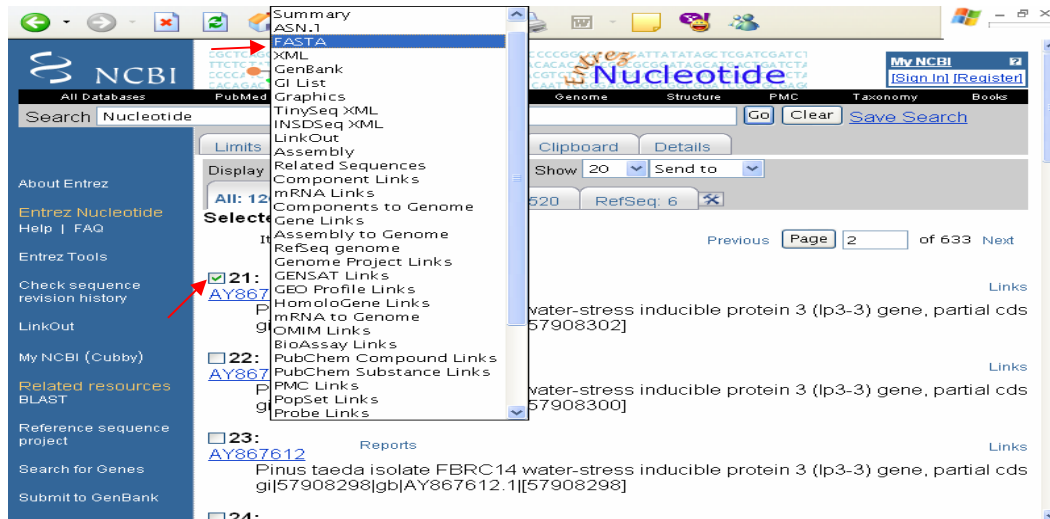


Figura 15. Item selecionado para adaptação da seqüência ao modo FASTA (setas).

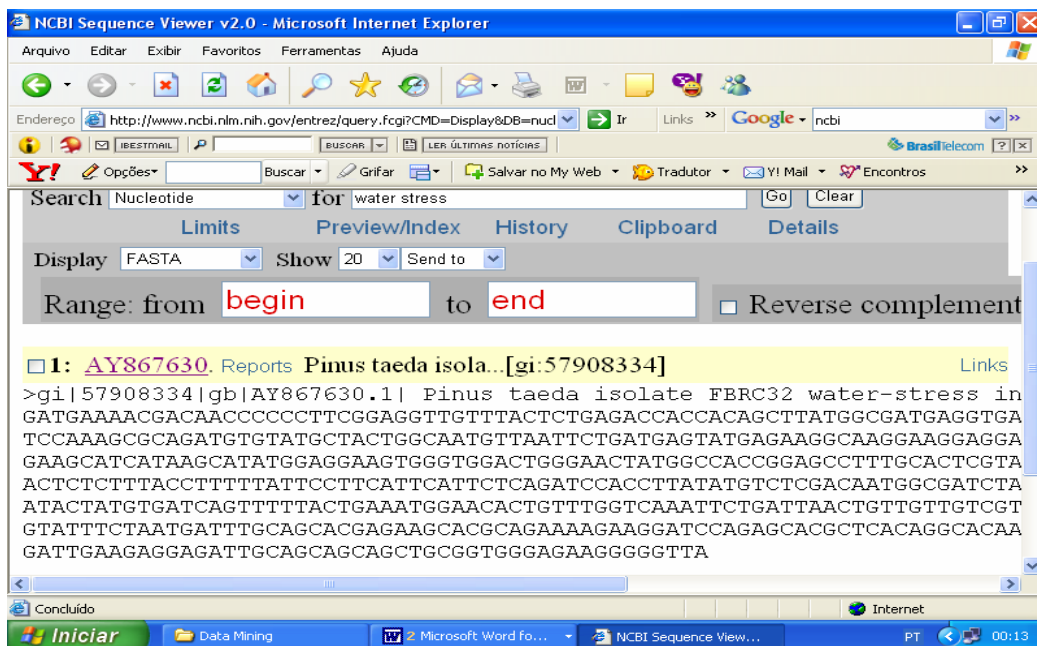


Figura 16. Resultado da utilização da ferramenta FASTA. A seqüência relacionada à resistência ao estresse hídrico apresenta um sinal de > seguida da identificação.

A seqüência de cada organismo identificada no NCBI como sendo relacionada à resistência ao estresse hídrico, já no modo FASTA (Figura 16), foi selecionada, copiada e submetida à ferramenta BLAST no banco de dados do SUCEST (Figura 17). Para acesso a ferramenta BLAST, uma senha fornecida pelo LNCC / UNICAMP aos participantes do projeto é solicitada na página inicial.

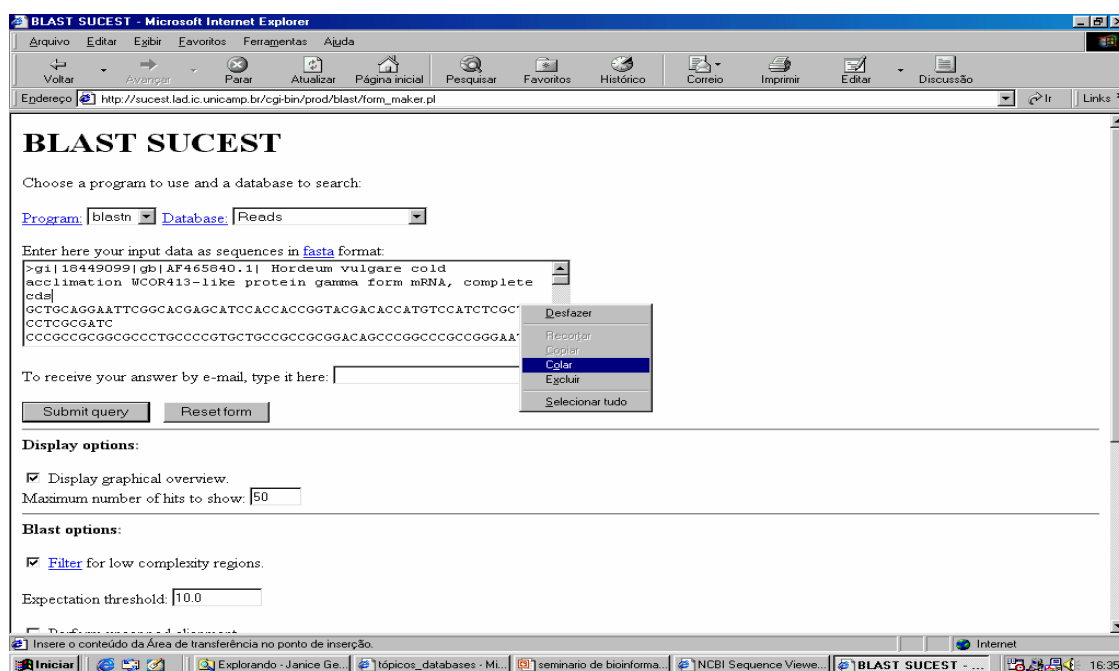


Figura 17. Página inicial da ferramenta BLAST do SUCEST (acesso restrito) para análise de homologia (BLASTn). A seqüência submetida para análise (*query*) foi comparada com seqüências de cana-de-açúcar (*subject*).

Foram evidenciadas as produções de alinhamentos significativos (altos *scores* e baixos *e-values*) das seqüências de cana-de-açúcar alinhadas (Figuras 18 e 19).

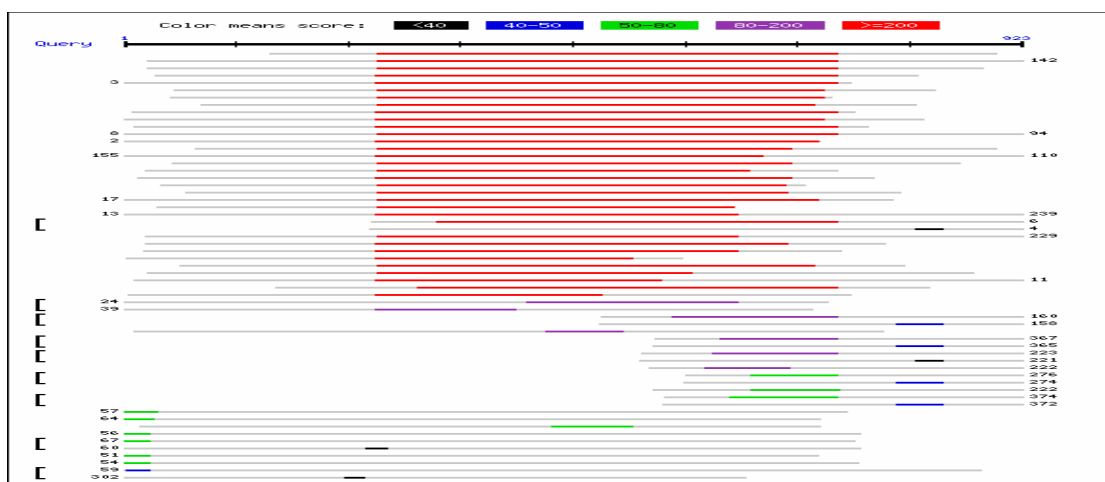


Figura 18. Resultado da análise de similaridade. Histograma gerado pela ferramenta BLASTn, onde se observa alta homologia entre as seqüências confrontadas. Linhas vermelhas indicam *scores* > 200.

Sequence	Score (bits)	E Value
SCSBST3096C11.g	1300	0.0
SCCCR22001D09.g	1239	0.0
SCEQSB1019A04.g	1197	0.0
SCCCR3008C10.g	1164	0.0
SCCGST3126F04.g	1116	0.0
SCQGAN2109F11.g	1112	0.0
SCCCR3014H09.g	1108	0.0
SCJLFL4099H10.g	1096	0.0
SCCCCL4014E06.g	1094	0.0
SCMCSB1109G02.g	1080	0.0
SCJFFL1C04H02.g	1078	0.0
SCSGFL403F10.g	1074	0.0
SCSFAD1125D09.g	1065	0.0
SCCCR3004H06.g	1051	0.0
SCCCFL4092D05.g	1047	0.0
SCVPLB2090B09.g	1045	0.0
SCSFFL4082G09.g	1043	0.0
SCBGLF5082B11.g	1039	0.0
SCUTR23073G06.g	1039	0.0
SCBGLR1113D02.g	1039	0.0
SCQGHR1014F09.g	1033	0.0
SCSGFL4034F08.g	1021	0.0
SCCCAM2001F10.g	1019	0.0
SCEZFL4041C03.g	1015	0.0
SCSGAD1005F01.g	1013	0.0
SCSBFL4012A07.g	1011	0.0
SCSGAD1006H05.g	1011	0.0
SCEQSB1C03G07.g	1007	0.0
SCCCFL4120B04.g	1003	0.0
SCCCFL4095G04.g	1003	0.0
SCQSLB1049F01.g	1001	0.0
SCCCLR1068B09.g	999	0.0
SCCCR3068B07.g	995	0.0

Figura 19. Valores de *scores* e *e-values* comprovando alta homologia entre as seqüências (destaque).

Com base nos resultados gerados na etapa acima, iniciou-se o processo de identificação das proteínas putativas encontradas, as quais foram submetidas aos bancos de dados secundários de domínio público. Neste estudo, foi selecionada uma das proteínas para exemplificar as etapas de análise.

5. Identificação de proteínas *in silico*

A etapa subsequente gerou resultados relativos à identificação e descrição das proteínas provavelmente relacionadas com a resistência ao estresse hídrico. O protocolo obedeceu as seguintes etapas:

Acesso a página inicial do banco de proteínas SWISS-PROT TrEMBL (Figura 20): Nessa página pode se encontrar as informações e diretrizes para a busca e identificação de proteínas eventualmente relacionadas com a indução a resistência ao estresse hídrico. Sendo assim, neste programa de pesquisa, as proteínas putativas selecionadas e avaliadas foram as seguintes:

5.1 - WSI76 protein induced by water stress

Na análise *in silico*, verificou-se que a proteína WSI76 (WSI = *Water Stress Induced*) induzida por estresse hídrico, isolada de *Oryza sativa* (TaxID: 4530) participa do grupo das transferases (Figura 21), com a função de transferir grupos Hexosil por processos de biossíntese de carboidratos (Figura 22). Esta proteína é composta de 328 aminoácidos (aa), com peso molecular de 37698 Da. Está incluída na família glicosil transferase 8, que compreende enzimas transferidoras de resíduos de açúcar para moléculas doadoras. A estrutura 2D das cadeias “A” com 261 resíduos foi obtida de *Oryctolagus cuniculus*, expressa em *Escherichia coli* com a resolução de 1,90Å. A cadeia apresenta 11 fitas β e 17 α -hélices, dois ligantes GOL-glicerol de fórmula: $2(C_3H_8O_3)$ e 245 moléculas de água. A WSI76 participa da rota metabólica da pirimidina.

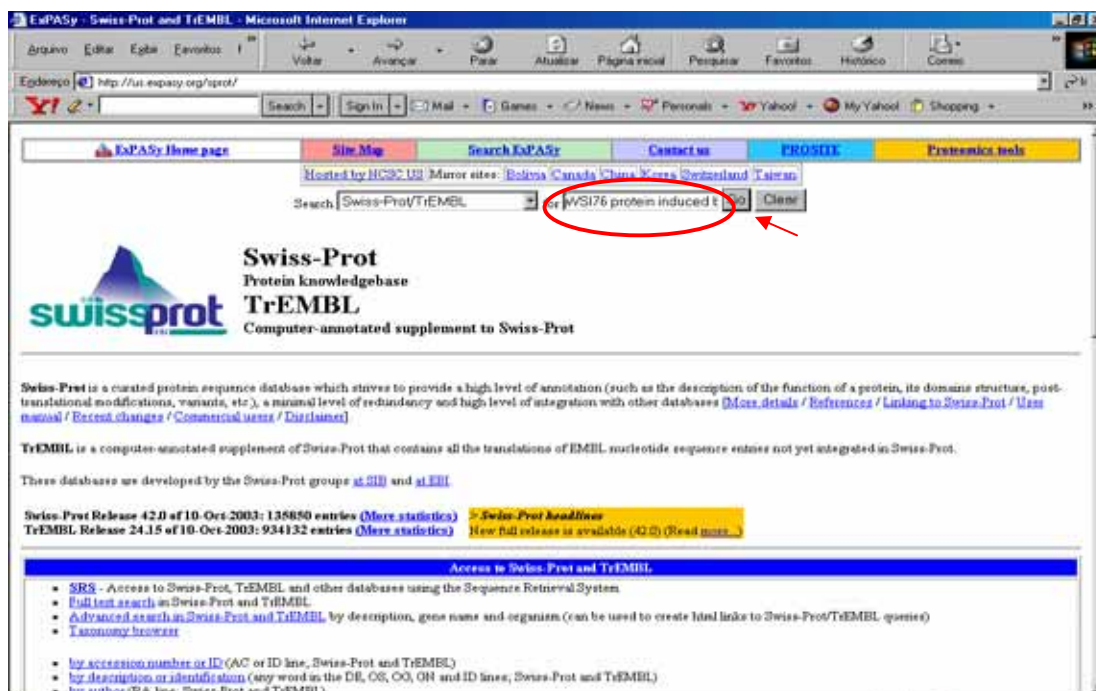


Figura 20. Submissão da proteína WSI76 como palavra chave no banco de dados SWISS-PROT/TrEMBL (seta).

Nesta página, *links* para outros bancos (Figura 21) de proteínas estão disponíveis e interligados, dando seqüência à pesquisa a cada acesso (Figura 22).

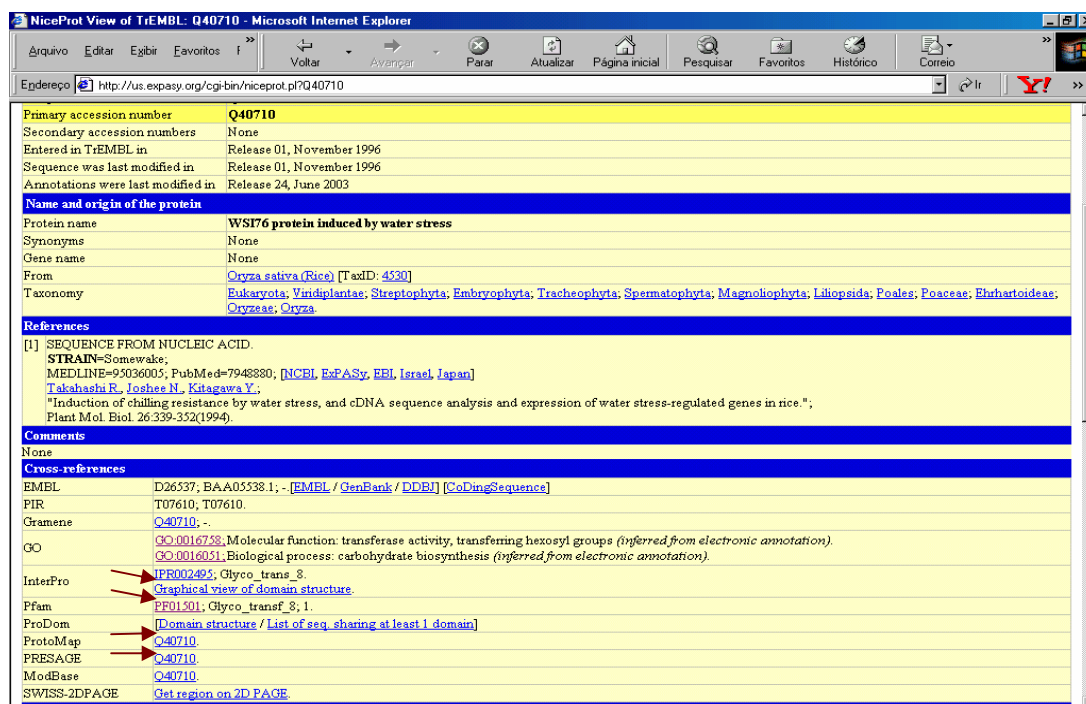


Figura 21. Resultado da busca com informações e *links* que dão acesso ao InterPro, Pfam entre outros (setas).

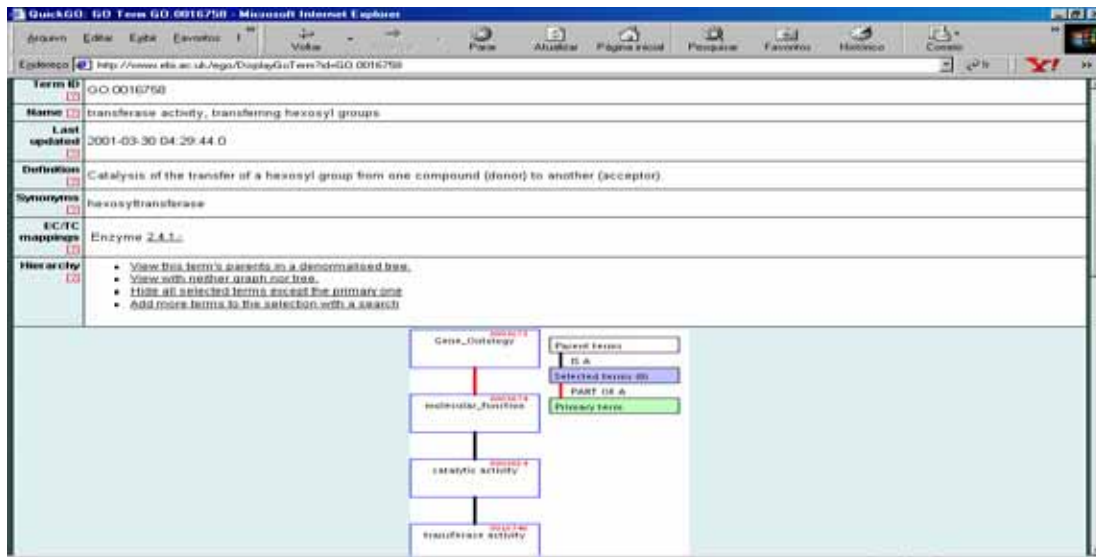


Figura 22. Página de resultados com informações sobre a proteína no banco de dados do EBI acessado através de link no Swiss-Prot.

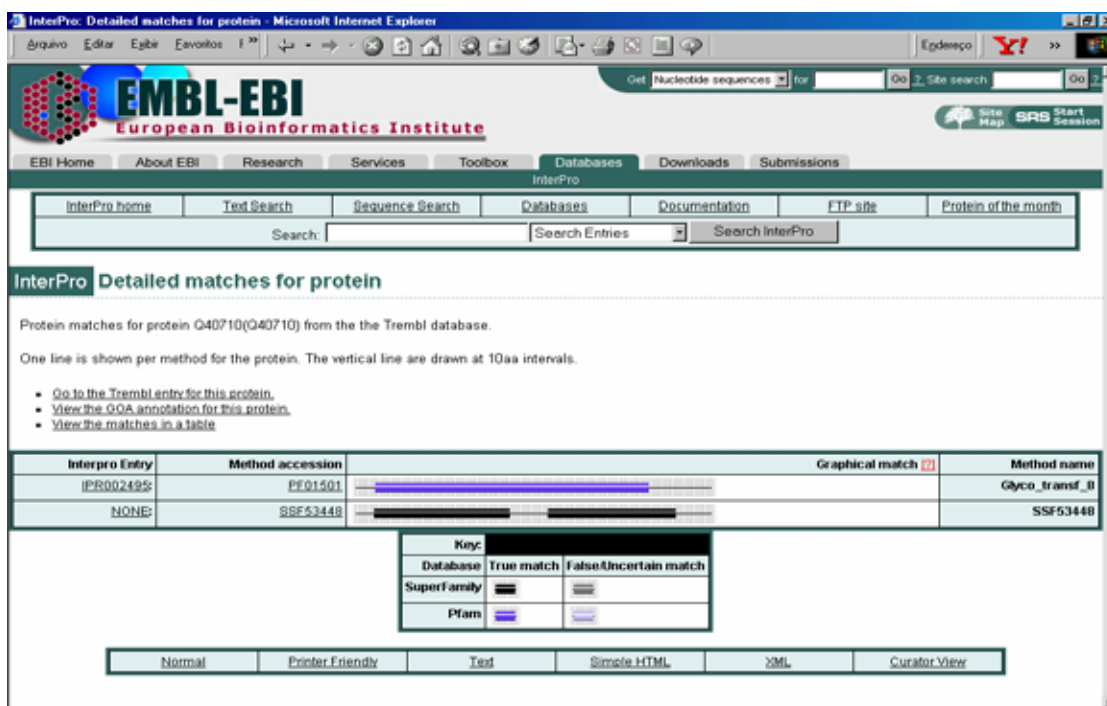


Figura 23. Resultado ilustrando domínio conservado da proteína.



Pfam Protein families database of alignments and HMMs

Home Keyword Search Protein Search Browse Pfam DNA Search Taxonomy Rf Help Glyco_transf_8 family

Accession number: PF01501

Glycosyl transferase family 8

This family includes enzymes that transfer sugar residues to donor molecules. Members of this family are involved in lipopolysaccharide biosynthesis and glycogen synthesis. This family includes Lipopolysaccharide galactosyltransferase [REAL_ECOLI](#), lipopolysaccharide glucosyltransferase 1 [REAL_ECOLI](#) and glycogenin glucosyltransferase [GLYO_HUMAN](#).

INTERPRO description (entry IPR002405)

The biosynthesis of disaccharides, oligosaccharides and polysaccharides involves the action of hundreds of different glycosyltransferases. These are enzymes that catalyse the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds. A classification of glycosyltransferases using nucleotide diphospho-sugar, nucleotide monophospho-sugar and sugar phosphates (EC: 2.4.1.-) and related proteins into distinct sequence based families has been described [MEDLINE:97474370]. This classification is available on the CAZY (Carbohydrate-Active Enzymes) web site P000007032. The same three-dimensional fold is expected to occur within each of the families. Because 3-D structures are better conserved than sequences, several of the families defined on the basis of sequence similarities may have similar 3-D structures and therefore form 'clans'.

Glycosyltransferase family 8 (CAZY:GT_8), lipopolysaccharide glucosyltransferase 1 (EC:2.4.1.58), glycogenin glucosyltransferase (EC:2.4.1.186), inositol 1- α -galactosyltransferase (EC:2.4.1.122). These enzymes have a distant similarity to family GT24.

QuickGO

FUNCTION :	transferase activity, transferring hexosyl groups (GO:0016748)
PROCESS :	carbohydrate biosynthesis (GO:0016051)

Alignment **Domain organisation**

Seed (78) Full (227) Seed (78) Full (227) Curated (4)

Figura 24. Visualização da estrutura 2D da cadeia "A" na página do Pfam acessada através de *link* do Swiss -Prot.



PDBsum **1IB3**

PDB id: 1IB3

Transferase

Title: Crystal structure of rabbit muscle glycogenin

Structure: Glycogenin-1, Chain: a, Engineered, yes

Source: *Oryctolagus cuniculus*, Rabbit, Tissue: skeletal muscle. Expressed in: escherichia coli

Resolution: 1.50Å R-factor: 0.210 R-free: 0.248

Authors: B.J.Oblinns, P.J.Rosch, T.D.Huley

Date: 26-Apr-02

PDB header records

SWISS-PROT code: GLYG_RABIT [Swiss-prot](#) [Pfam](#)

Enzyme class from PDB file: EC:2.4.1.186 [E.C.->PDB](#)

Related entries: [1IB0](#) PDB id [1IB0](#) is the original 3.4a mod structure solution of glycogenin with a partner of dimers in the asymmetric unit and p4(7)(1)2 space group. [1IB2](#) PDB id [1IB2](#) is a 1.96a complex of glycogenin with ms2 + and udp-glucose, i222 space group.

Links:

- [PDB](#)
- [DCA](#)
- [MMQB](#)
- [JMB Jense](#)
- [STING](#)
- [BRASS](#)
- [PQS](#)
- [CATH](#)
- [scop](#)
- [PSSP](#)
- [PROCHECK](#)
- [WHATCHECK](#)
- [PROMOTIF](#)
- [Relibase](#)

Figura 25. Página do PDBsum, onde foi visualizada a estrutura cristalizada da proteína com *links* de acesso para outras informações.

E.C.2.4.1.186 Glycogenin glucosyltransferase. - Microsoft Internet Explorer

PDBsum Enzymes E.C.2.4.1.186

[E.C.2.4.1.186](#) **Transferases.**

[E.C.2.4.1.186](#) **Glycosyltransferases.**

[E.C.2.4.1.186](#) **Hexosyltransferases.**

E.C.2.4.1.186 Glycogenin glucosyltransferase.

Reaction: UDP-glucose + glycogenin = UDP + glucosylglycogenin

Other name(s): Glycogenin

Comments: The glycogenin subunit of E.C.2.4.1.11 catalyzes this reaction, i.e. the enzyme catalyzes its own autoglucosylation. Five molecules of glucose can be transferred to one molecule of glycogenin. The product acts as a primer for the reaction catalyzed by glycogen synthase.

Links to other enzyme databases: [ExPaSy](#) [KEGG](#) [WIT](#) [BRENDA](#)

There are 3 PDB entries in enzyme class E.C.2.4.1.186:

- [1100](#)
Structure: 1. Glycogenin-1.
Source: 1. *Oryctolagus cuniculus*. Rabbit. Tissue: skeletal muscle. Expression_system_common: bacteria. Expression_system_vector_type: plasmid
- [1102](#)
Structure: 1. Glycogenin-1.
Source: 1. *Oryctolagus cuniculus*. Rabbit. Tissue: skeletal muscle. Expression_system_common: bacteria. Expression_system_vector_type: plasmid
- [1103](#)
Structure: 1. Glycogenin-1.
Source: 1. *Oryctolagus cuniculus*. Rabbit. Tissue: skeletal muscle. Expression_system_common: bacteria. Expression_system_vector_type: plasmid

PDBsum Enzymes E.C.2.4.1.186

Figura 26. Página do PDBsum contendo informações e link de acesso que leva ao KEGG.

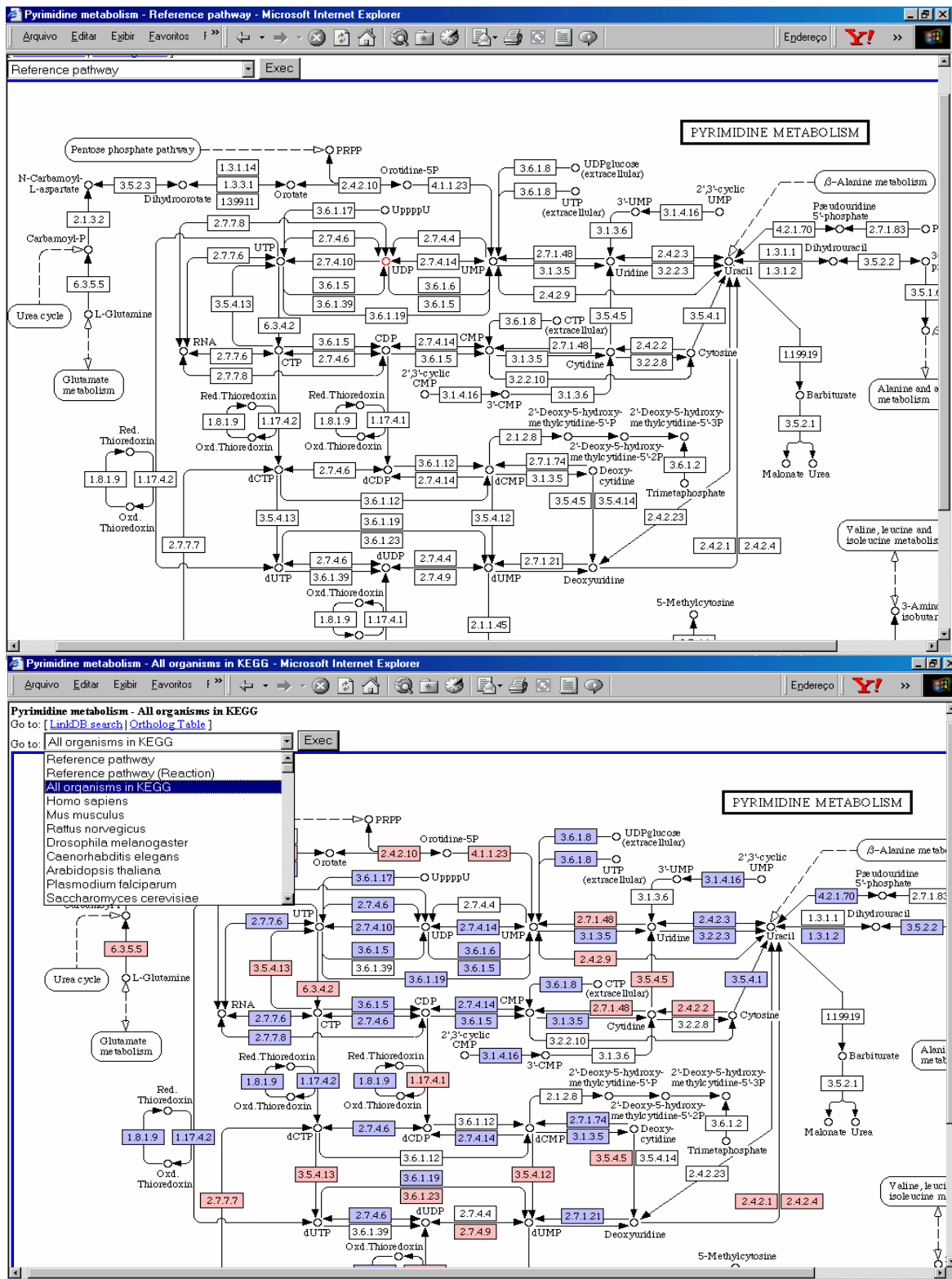


Figura 27. Rota metabólica do metabolismo da pirimidina a qual a enzima está inserida.

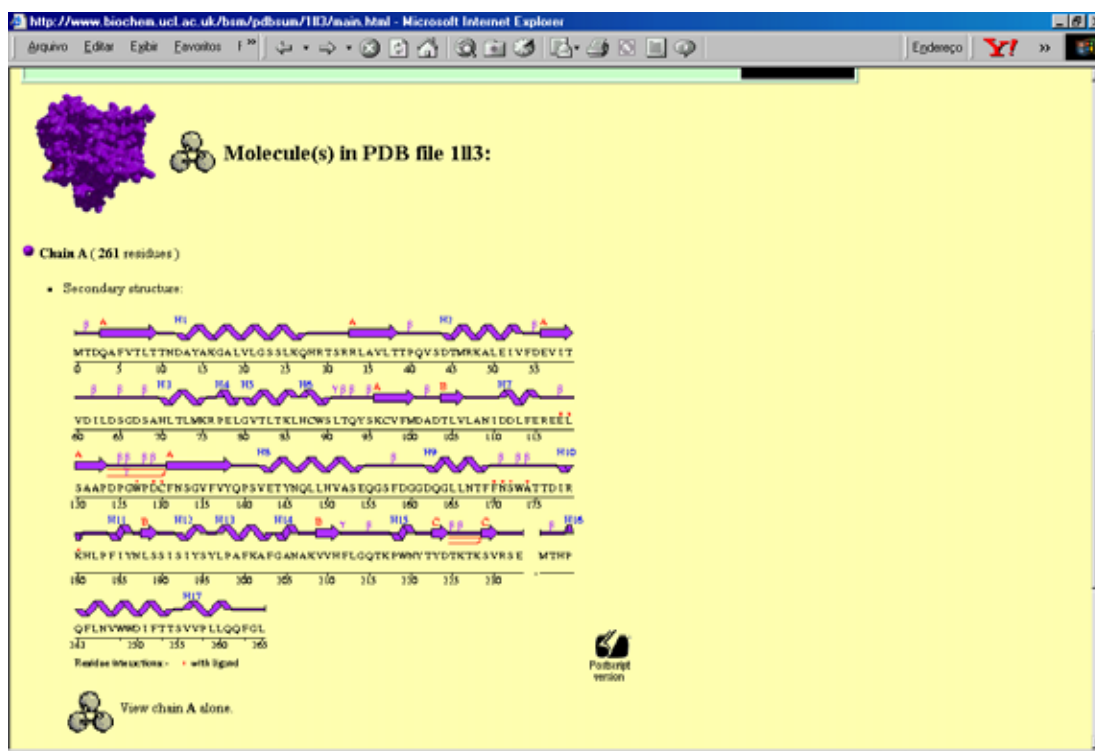


Figura 28. Representação esquemática da cadeia “A” (estrutura secundária).

PDB id: 1113

Transferase

Title: Crystal structure of rabbit muscle glycogenin

Structure: Glycogenin-1, Chain: a, Engineered: yes

Source: *Oryctolagus cuniculus*, Rabbit, Tissue: skeletal muscle, Expressed in: *escherichia coli*.

Resolution: 1.90 Å, R-factor: 0.210, R-free: 0.248

Authors: B.J. Osborne, P.J. Roach, T.D. Hurley

Date: 26-Apr-02

SWISS-PROT code: GLYG_RABIT | Swiss-prot | Pfam

Enzyme class from PDB file: E.C.2.4.1.106 | E.C.->PDB

Related entries: 1100 PDB id 1100 is the original 3.4a mad structure solution of glycogenin with a pentamer of dimers in the asymmetric unit and p4(7)2(1)2 space group. 1102 PDB id 1102 is a 1.90a complex of glycogenin with ms2+ and wsp-glucose, 1222 space group.

Links:

- PDB
- OCA
- MMDB
- IMB Jena
- STING
- GRASS
- PQS
- CATH
- scop
- FSSP
- PROCHECK
- WHATCHECK
- PROMOTIF
- Relibase

Figura 29. Acesso à visualização da estrutura 3D da proteína na página do PDBsum.

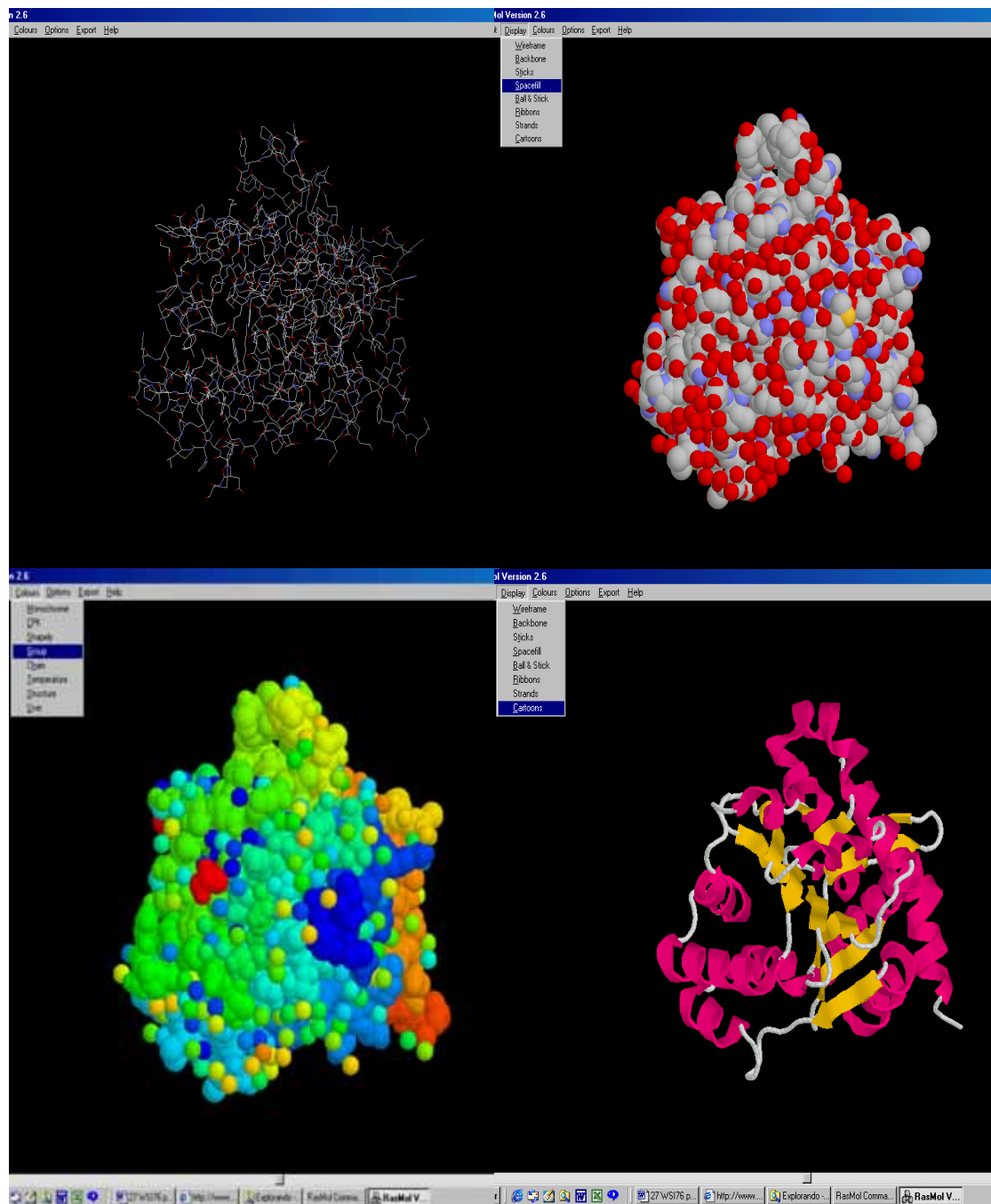


Figura 30. Visualização da estrutura 3D da cadeia A da proteína WS76 no PDBsum através do programa RASMOL. Neste programa é possível girar, movimentar a estrutura em vários ângulos, mudar apresentação, entre outras funções.

O protocolo acima descrito foi adotado para os procedimentos de avaliação das seqüências de DNA relacionadas com as 52 proteínas identificadas. Devido ao grande volume de informações obtidas com a mineração de dados, apenas cinco proteínas putativas serão descritas a seguir incluindo sua rota metabólica e estrutura 3D.

5.2 - 41,9 KDa ou gene T25K17,60

Esta proteína apresenta atividade de transferase, participa transferindo grupos Hexosil por processo de biossíntese de carboidratos. Foi isolada de *Arabidopsis thaliana* (TaxID: 3702). Esta proteína é composta de seqüência parcial de 332aa, com peso molecular de 37266 Da. A proteína putativa 41,9 KDa pertence à família glicosil transferase 8, que inclui enzimas que transferem resíduos de açúcar para moléculas doadoras. Membros dessa família estão envolvidos com a biossíntese de lipopolissacarídeos e síntese de glicogênio, também inclui lipopolissacarídeos glicosiltransferase, lipopolissacarídeos glicosiltransferase 1 e glicogenina glicosiltransferase. A estrutura 2D das cadeias "A" com 225 resíduos foi obtida de *Oryctolagus cuniculus*, apresenta 11 fitas β e 17 α -hélices, participa da rota metabólica da pirimidina (Figuras 31) com resolução 1,90Å (Figura 32).

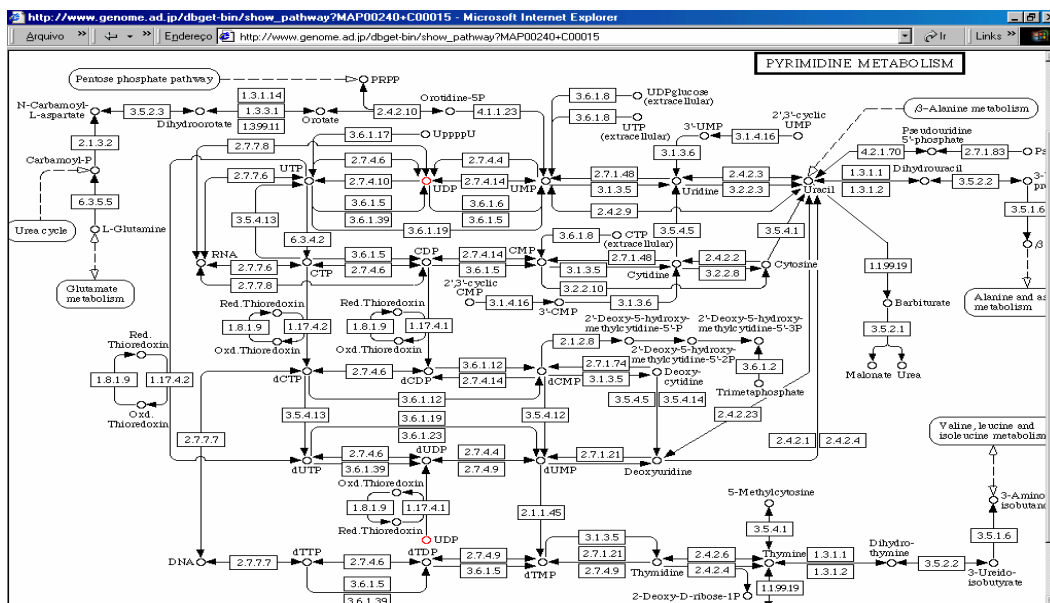


Figura 31. Rota metabólica da pirimidina, a qual a proteína 41,9 KDa está inserida.

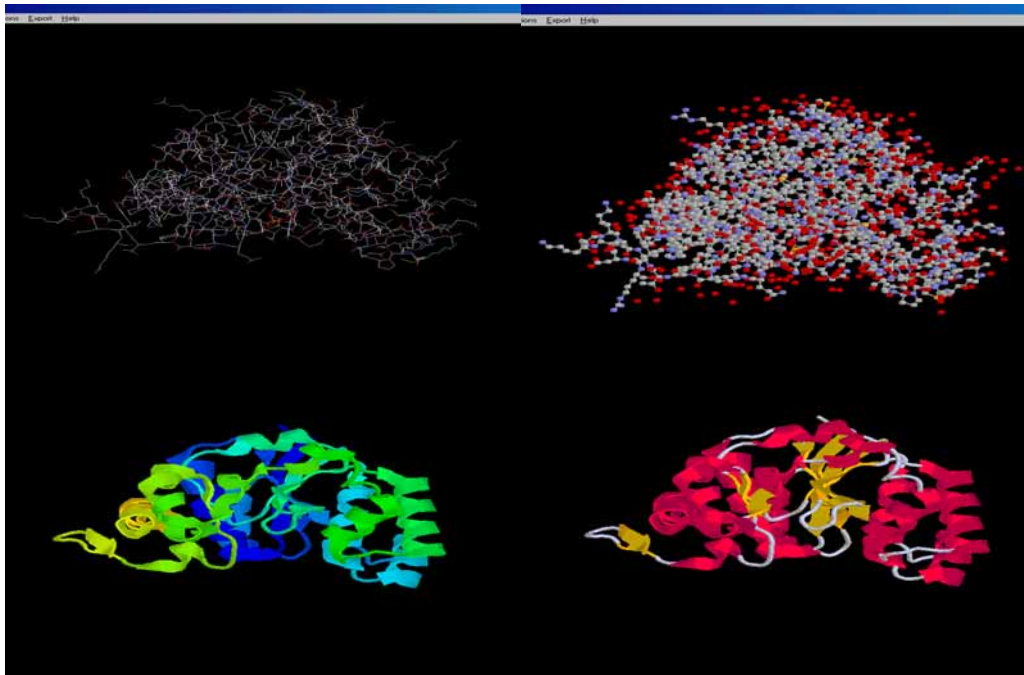


Figura 32. Visualização da estrutura 3D da cadeia A da hipotética 41,9 KDa (gene T25K17,60) através do programa VRML do PDBsum.

5.3- *Alpha amylase [precursor] (fragment)*

A enzima precursora da alfa amilase ou gene CUS AMY2 apresenta atividade alfa-amilase por processo de metabolismo de carboidrato. Faz parte do grupo das hidrolases no sistema imunológico. Foi isolada de *Cuscuta reflexa* (TaxID: 4129). Esta proteína é composta de seqüência parcial de 264aa, com peso molecular de 29093 Da, está na classe das hidrolases com função de atividade alfa-amilase, família alfa-amilase e apresenta um domínio catalítico. A estrutura 2D apresenta 8 cadeias (A, B, C, D, E, F, G e H) com 496 resíduos as quatro primeiras cadeias e 120 resíduos nas demais. Foi obtida de *sus scrofa*, com resolução de 1,60Å. As cadeias apresentaram um total de 140 fitas β e 96 α -hélices. A rota metabólica a qual participa é a do amido e da sacarose (Figuras 33). A proteína contém 4 íons metal Cl, 4 íons metal Ca e 2807 moléculas de água (Figura 34).

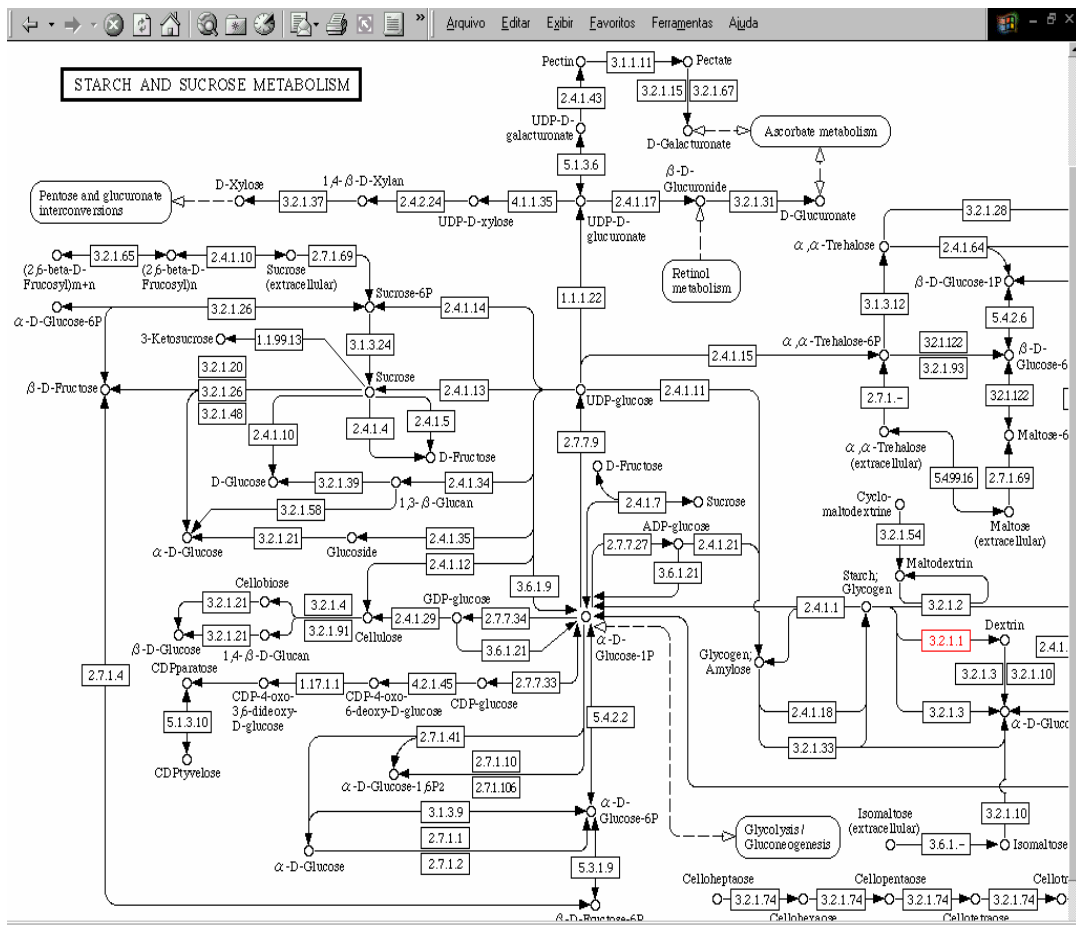


Figura 33. Rota metabólica da sucrose onde *alpha amylase* atua.

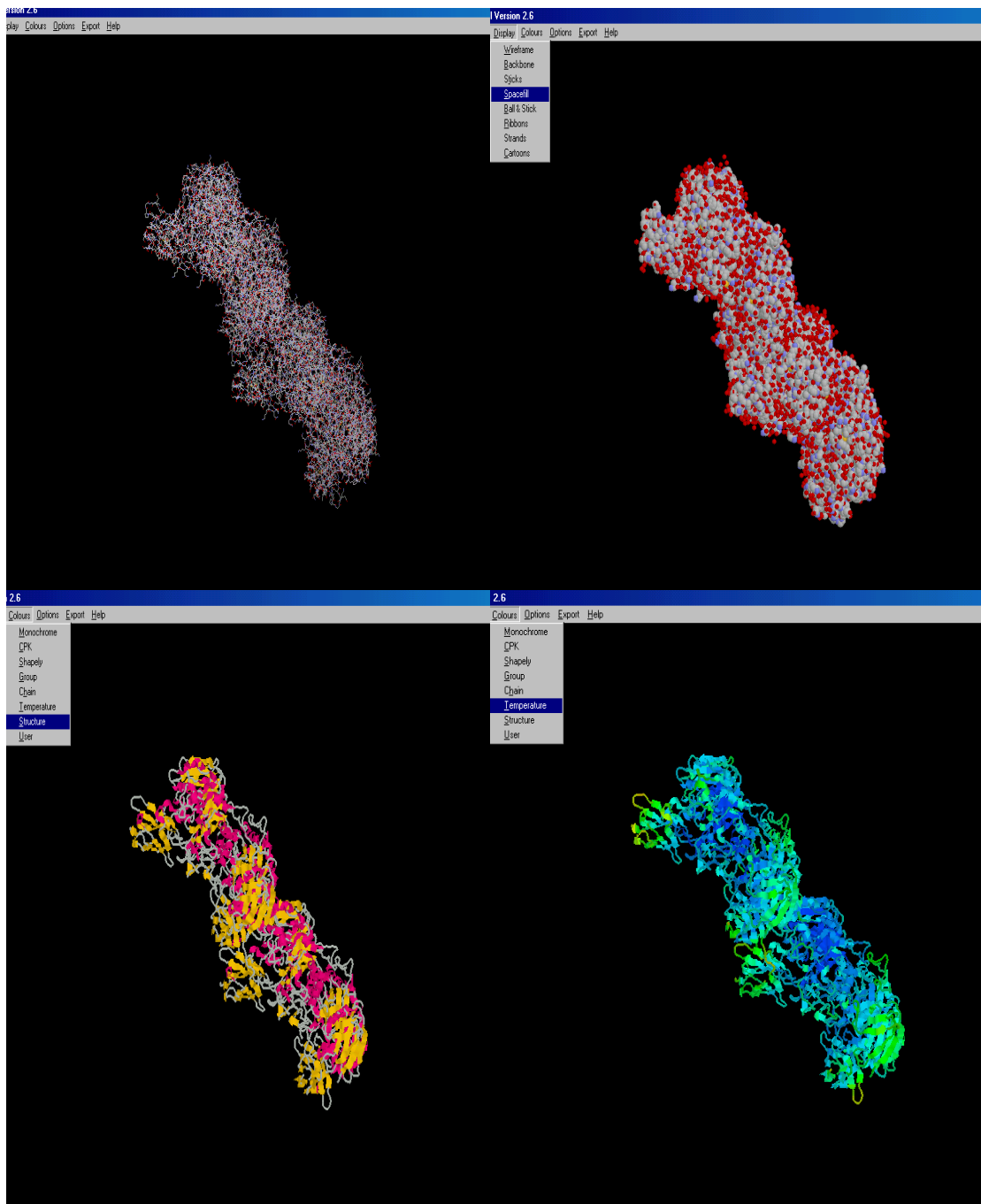


Figura 34. Estrutura 3D da cadeia A da proteína alpha-amilase (Gene.CUS AMY2)

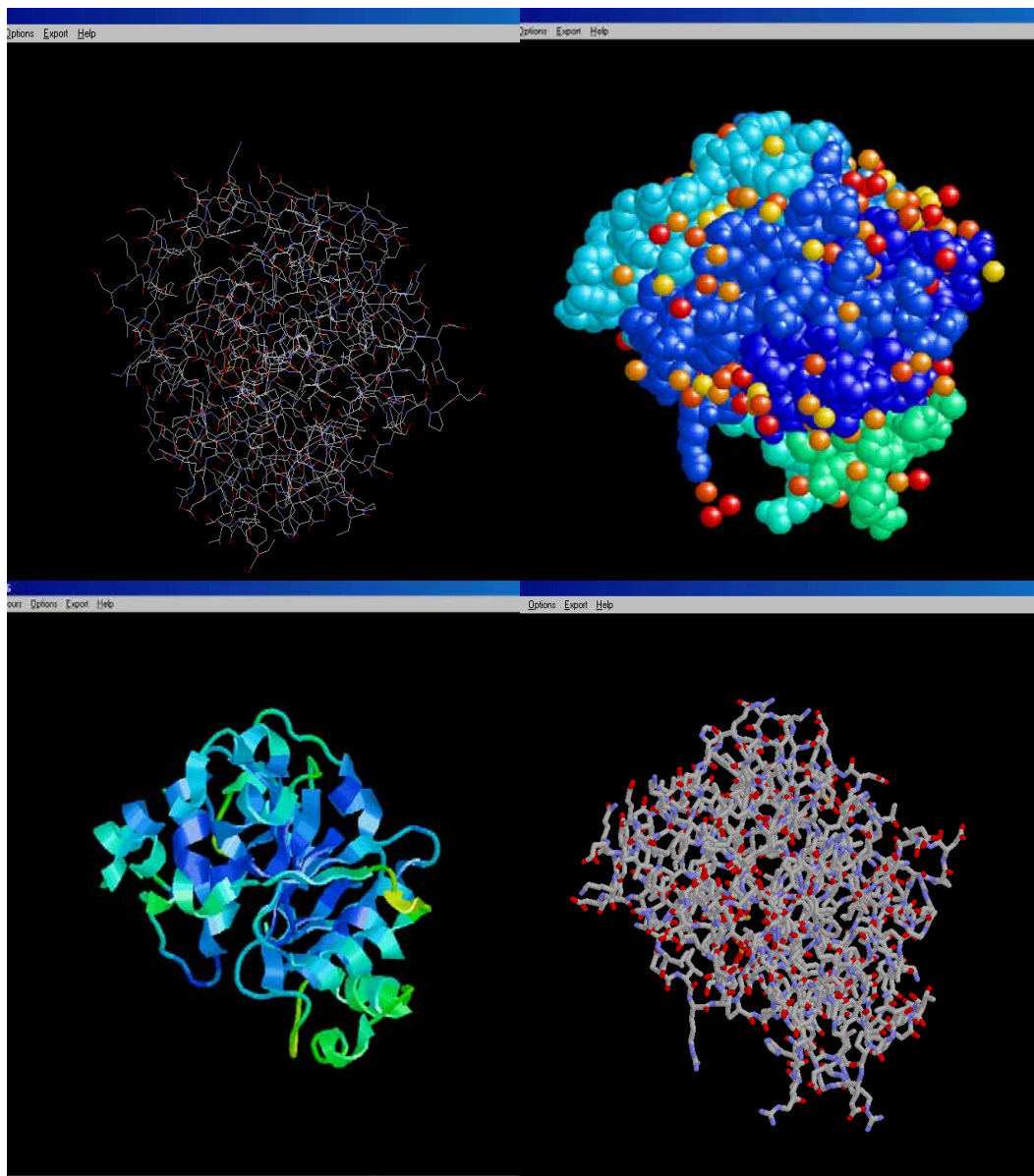


Figura 36. Visualização da estrutura 3D da cadeia A da galactinol synthase através do programa VRML do PDB.

5.5 - Cytosolic pyruvate orthophosphate dikinase

Cytosolic pyruvate orthophosphate dikinase ou gene *PPDK1* apresenta função de atividade quinase através de processo de fosforilação. Foi isolada de *Oryza sativa* (TaxID: 4530). É composta de 947 aa, com peso molecular de 102801 Da. A proteína cataliza a conversão reversível de ATP para AMP, pirofosfato e fosfoenolpiruvato (PEP), resíduos na terminação N para o trânsito de peptídeo o qual é indispensável para o transporte da proteína precursora. Funciona na direção da formação de PEP, o qual é o acceptor primário do dióxido de carbono em plantas C4 e CAM.

A visualização 2D das cadeias “A” com 869 resíduos foi obtido da bactéria *Clostridium symbiosum* e expressa em *Escherichia coli* com resolução de 2,60 Å (Figura 37). A cadeia apresenta 31 fitas β e 48 α-hélices. A proteína contém 2 ligantes SO₄ – íon sulfato: 2(O₄ S₁ 2-) e 227 moléculas de água, participa da rota metabólica do piruvato.

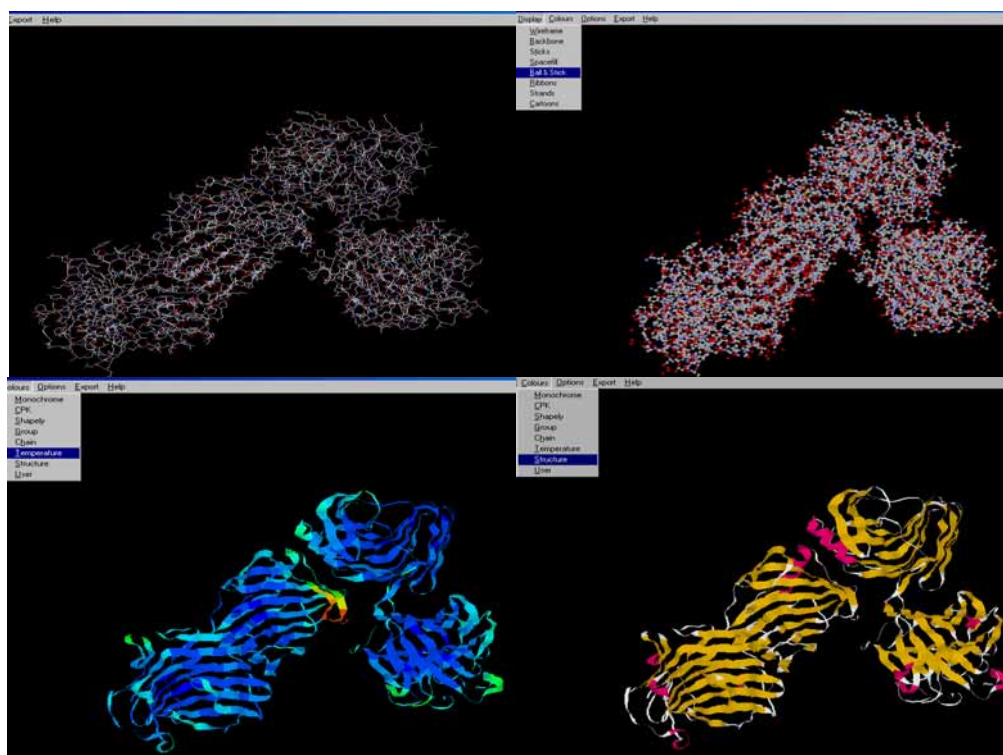


Figura 37. Visualização da estrutura 3D cadeia A da proteína cytosolic pyruvate orthophosphate dikinase através do programa VRML do PDB.

5.6 - *Lactoylglutathione lyase*

Esta proteína cataliza a conversão de hemimercaptal formado de metilglioxal e glutationa para S-lactoelglutationa. Foi isolada de *Cicer orietinum* (TaxID: 3827). É composta de 186aa, com peso molecular de 21088 Da. Faz parte da superfamília da dioxigenase. A glioxalase I catalisa o primeiro passo da rota glioxal. S-lactoelglutationa é então convertida pela glioxalase II para o ácido láctico. A glioxalase I é uma enzima ambígua na qual anexa uma molécula de zinco por subunidade. A seqüência de glioxalase I é bem conservada. O domínio é encontrado em outras proteínas relacionadas, incluindo a proteína de resistência bleomicina e dioxigenase, por exemplo: 4- hidroxifenilpiruvato dioxigenase.

Observou-se que em mamíferos e bactérias a enzima tem um comprimento em torno de 130 a 180 resíduos, enquanto que em fungo essa seqüência é duas vezes mais longo. Nesses organismos, a enzima é construída em repetições de um domínio homólogo. Tem sido observados dois padrões de assinatura para essa família. A primeira localiza-se em um N-terminal, enquanto que a segunda é localizada na sessão central da proteína e contém uma histidina conservada que pode estar implicada na adição do íon zinco. A proteína contém 4 íons metal FEL- formula $4(\text{H}_6\text{O}_3 \text{FE}_1 3+)$ e 1093 moléculas de água e participa do metabolismo da tirosina (Figura 38). A estrutura 2D apresenta 4 cadeias (A, B, C e D) com um total de 1276 resíduos obtida de *Brevibacterium fuscum* e expressa em *Escherichia coli* com resolução de 1,60Å, na qual a cadeia “A” apresenta 17 fitas β e 7 α -hélices (Figura 39).

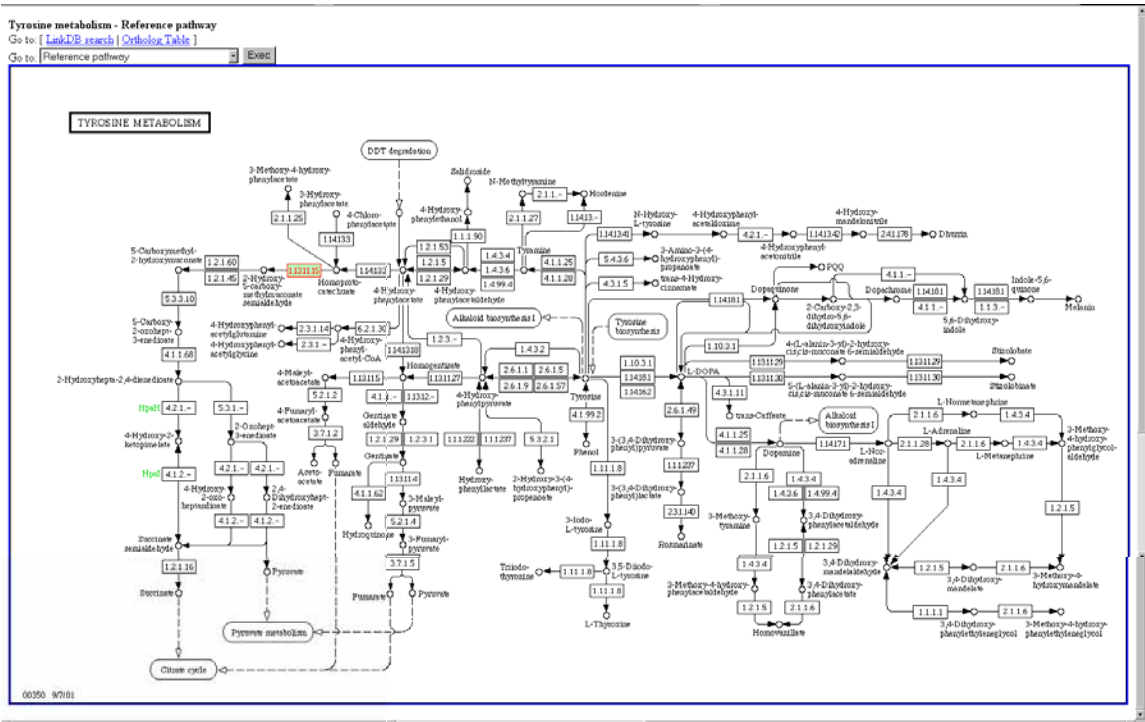


Figura 38. Rota metabólica mostrando a participação da *lactoilglutathione lyase* no metabolismo da tirosina.

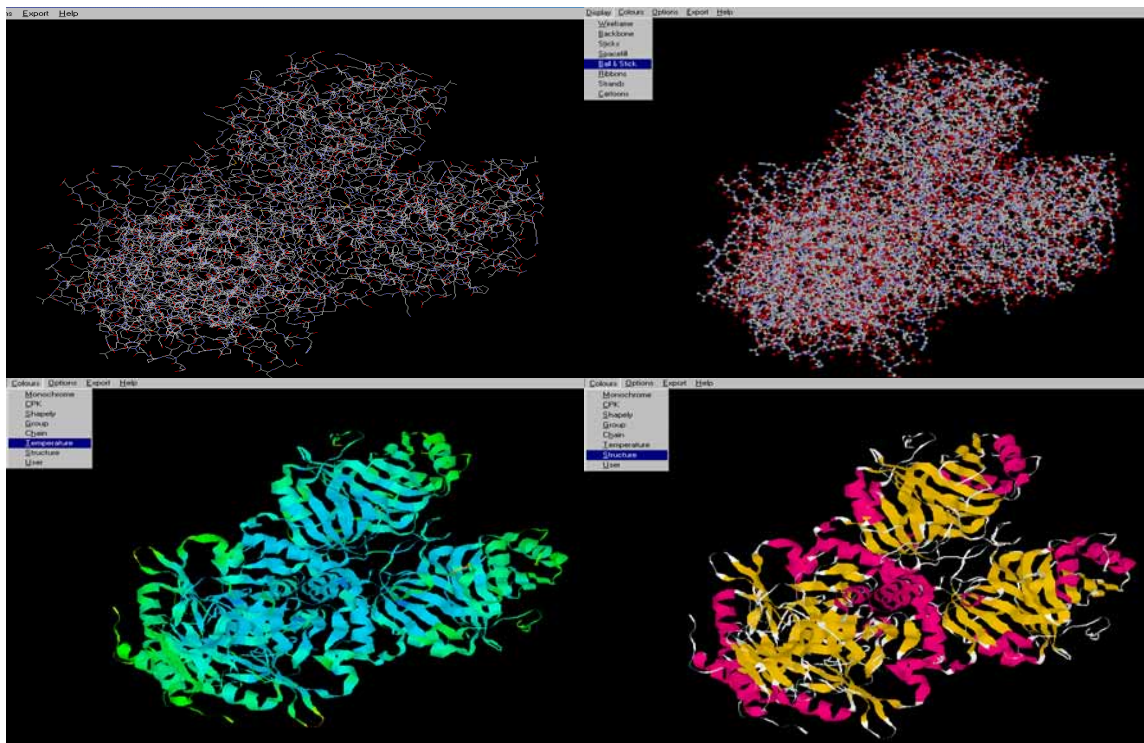


Figura 39. Visualização da estrutura 3D da cadeia A da proteína *lactoilglutathione lyase* através do programa VRML do PDB.

Os resultados obtidos nesse trabalho podem ser comparados com outras análises *in silico* realizadas em outros gêneros de plantas como, por exemplo, *Eucalyptus* (VICENTINI et al. 2005). Essa avaliação permitiu a obtenção de informações de diferentes proteínas produzidas e a identificação de genes que mostraram padrões de expressão diferenciados nas amostras estudadas. A análise de transcriptoma em *Eucalyptus* se constituiu num banco de dados de grande interesse para futuros estudos incluindo a descoberta de genes específicos em determinados tecidos. O conjunto de dados gerados pelo SUCEST permitiu o desenvolvimento de inúmeros sub-projetos de mineração de dados incluindo aquele proposto pelo grupo de pesquisa da Universidade Federal de Alagoas, que versava pela identificação de genes relacionados com a resistência ao estresse hídrico em cana. Os resultados produzidos por VICENTINI et. al, 2005, mostram que genes que são expressos na maioria dos tecidos podem se constituir em excelentes alvos para a clonagem de promotores ambíguos, que podem ser úteis em muitas aplicações. No caso específico da cana-de-açúcar, nesse presente trabalho, será necessária a avaliação de *per se* dessas proteínas relacionadas com a resistência ao estresse hídrico. O estudo dessas proteínas putativas será a próxima etapa deste projeto, cujo tema integra a programação da Rede Nacional de Proteoma recém estruturada no Laboratório de Genética Molecular, Genômica e Proteômica – GEMPRO/UFAL. O objetivo final será a construção de variedades de cana-de-açúcar transgênicas, obtidas através da inserção de genes que codificam para as proteínas identificadas nesta análise *in silico*. As primeiras atividades de implantação de programas de transgênia em cana estão sendo planejadas pela Rede Interinstitucional de Desenvolvimento do Setor Sucroalcooleiro – RIDESA, da qual, fazem parte oito Instituições Federais de Ensino Superior, inclusive a Universidade Federal de Alagoas. Para tanto, foi financiado pela FINEP a implantação do Laboratório de Proteômica da UFAL, e do Banco de DNA de Germoplasmas da Serra do Ouro, totalmente financiado pela RIDESA e Banco do Nordeste.

CONCLUSÕES

Os protocolos de mini preparação de plasmídeos (mini-prep) e da análise genômica dos clones foram eficientes e produziram resultados consistentes para as 15 placas da biblioteca genômica avaliada;

O seqüenciamento automático produziu 1440 *reads* que foram utilizados para análise *in silico* proposta no presente trabalho;

A ferramenta BLASTn foi eficiente na avaliação de parâmetros estatísticos do alinhamento de seqüências de DNA permitindo a identificação de 2998 clusters com homologia para proteínas putativas relacionadas a resistência ao déficit hídrico;

A proteína tipw_pea p25794 water-stress induced tonoplast intrinsic protein galactinol synthase teve 1722 clusters homólogos;

A análise *in silico* evidenciou a existência de pelo menos 52 seqüências de DNA no SUCEST que podem estar relacionadas à resistência ao estresse hídrico.

REFERÊNCIAS

ADAMS, M. D.; KELLEY, J. M., GOCAYNE, J.D., DUBNICK, M., POLYMERPOULOS, M. H., XIAO, H.; MERRIL, C.R., WU, A., OLDE, B., and MORENO, R.F. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252** (5013): 1651-1656, 1991a.

ADAMS, M. D., KERLAVAGE, A. R., FLEISCHMANN, R.D., FULDNER, R.A., BULT, C. J., and WHITE, O. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377** (6547 Suppl): 3-174, 1995b.

ALEXANDER A. G. **Sugarcane physiology**. Amsterdam: Elsevier, p.752, 1973.

ALMEIDA, R. S; Identificação e caracterização de genes de transportadores de fosfato em cana-de-açúcar (*Saccharum spp*). (**Dissertação Mestrado**) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, p. 43,2002.

ANDRADE, J. C. T.; BARBOSA, G. V. S. Controle de qualidade do corte de cana-de-açúcar (*Saccharum spp.*) na Usina Caeté na safra 95/96. In: CONGRESSO NACIONAL DA STAB, 6. **Anais...** Maceió: STAB, p.571-579. 1996.

Apostila do curso RBQ-5758 - Tópicos em Bioinformática. Disponível em: <http://www.rbi.fmrp.usp.br/bioinfo/bioinfo1.htm> acesso em: 15/06/2006.

BACCHI, O. O. S. **Ecofisiologia da Cana-de-açúcar**. Instituto do Açúcar e do Alcool – Programa Nacional de Melhoramento da Cana-de-açúcar / PLANALSUCAR, 20p. Piracicaba-SP, 1985a.

Cana doce. Disponível em: http://www.pas.pt/canaldoce/hist_acucar/conteudos/hist_conteudos.html Acesso em: 08/07/2006.

Cana-de-açúcar . Disponível em: <http://br.geocities.com/atine50/cana/cana.htm> Acesso em: 08/07/2006.

Cana-de-Açúcar, uma das riquezas da cidade de Sertãozinho/SP – Um pouco de história. Disponível em: <http://www.sertaozinho.com/cana-de-acucar.shtml> Acesso em: 09/07/2006.

CASTRO, P. R. C. Utilização de reguladores vegetais no sistema de produção da cana-de-açúcar. In: Curso: Fisiologia vegetal aplicada à cana-de-açúcar. STAB-Leste, Maceió, 10p. 2001.

CIDADE D. A. DE P.; GARCIA, R. DE O.; DUARTE A. C.; SACHETTO-MARTINSE, G.; MANSUR, E. Morfogênese *in vitro* de variedades de cana-de-açúcar. **Pesq. Agropec. Bras.**, Brasília, v.41, n.3, p.385-391. Março, 2006.

CLARK, R. N., HILER, E. A. Plant measurements as indicators of crop water deficit. **Crop. Sci.**, v. 13, p. 466-469, 1973.

CONAB - Avaliação da Safra Agrícola de cana-de-açúcar 2006/2007 – Primeiro Levantamento – maio/2006.

CÔRTEZ. S. C. DA. **Data Mining - Conceitos, Técnicas, Ferramentas e Aplicações**. Disponível em: <http://www.cce.puc-rio.br/informatica/dataminingcentro.htm> Acesso em: 10/07/2006.

CLUBE DO FAZENDEIRO. Cana-de-açúcar – Características gerais. Disponível em <http://www.clubedofazendeiro.com.br/guiaprod/GuiaProduto.asp?codigo=3&Item=1#> Acesso em: 09/07/2006.

EWING, RM, KAHLA B.A., POIROT, O. LOPEZ, F., AUDIC S. and CLAVERIE J. M. Large –scale statistical analyses of rice ESTs reveal correlated patterns of genes expression. **Genome Res.** 9:950-959, 1999.

FAYYAD, U., PIATETSKY-SHAPIRO, G., AND SMYTH, P. Knowledge discovery and data mining: Towards a unifying framework. In *Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 82–88 ,1996.

FRANCO G. R., RABELO, E. M, AZEVEDO V. PENNA H. B., ORTEGA J. M., SANTOS T. M. MEIRA, W. S. RODRIGUES N.A., DIAS C. M., HARROP R. WILSON A., SABER, M., ABDEL-HAMID H., FARIA M.S., MARGUTTI M.E., PARRA J.C. and PENNA S. D. Evaluation of cDNA libraries from different developmental stages of *Schistosoma masoni* for production of expressed sequence tags (ESTs). **DNA res** 4: 231:40, 1997.

FELIX, J.M., DRUMMOND, R.D., NOGUEIRA, F.T.S., DE ROSA JUNIOR, V.E., JORGE, R.A., ARRUDA, P., MENOSSE, M. (2002). Genoma Funcional: Uso de arranjos de DNA em nylon para a análise de expressão gênica em larga escala. **Biotecnologia Ciência & Desenvolvimento** 24:60-67. Disponível em: <<http://cafe.cbmeg.unicamp.br/pub/felix-gf.pdf>> . Acesso em: 14/07/2006.

FURTADO, D.C. DE M., *Análise funcional de genes de resistência ao estresse hídrico em cana-de-açúcar*, 2004.

GIBAS, J; JAMBECK, P. **Desenvolvendo bioinformática: ferramentas de software para aplicações em biologia**. Tradução: Cristina de Amorim Machado, Rio de Janeiro: Campus, 2001.

GRIVET, L.; GLASZMAN, J. C.; ARRUDA, P. Sequence polymorphism from EST data in sugarcane; a fine analysis of 6-phosphogluconate dehydrogenase genes. **Genetics and Molecular Biology**. v.24, p.161-167, 2001. In: ALMEIDA, R. S. Identificação de genes de transportadores de fósforo em cana-de-açúcar (*Saccharum spp.*) (Dissertação de Mestrado) Escola Superior de Agricultura Luiz de Queiroz. Piracicaba, p.43, 2002.

GLOBAL 21. **INFORMES SETORIAIS: AGRONEGÓCIO**. Disponível em: <<http://www.global21.com.br/informessetoriais/setor.asp?cod=12>> Acesso em: 10/07/2006.

HWANG, D.M., DEMPSEY, A.A., WANG, R.X., REZVANI, M., BARRANS, J.D., DAI, K.S. WANG, H.Y., MA, H., CUKERMAN, E., LIU, Y.Q., GU, J. R., ZHANG, J.S., TSUI, S.K., WAYE, M.M., FUNG, K.P., LEE, C.Y., and LIEW, C.C. A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. **Circulation** 96 (12): 4146-4203, 1997.

IBGE: Instituto Brasileiro de Geografia e Estatística. Produção Agrícola Municipal 2005. Disponível em: <<http://www.sidra.ibge.gov.br/>> Acesso em: 08/07/ 2006.

INGELBRECHT, I. L.; IRVINE, J.E.; MIRKOV. Posttranscriptional gene silencing in transgenic sugarcane; Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome. **Plant physiology**, v. 119, p. 1187-1197, 1999.

Interpretação dos clusters. Disponível em: <bibliotecadigital.sbc.org.br/download.php?paper=376> Acesso em: 10/07/2006.

JOURNET, E. P., VAN, T. D., GOUZY, J. CRESPEAU H. CARREAU V., FARMER, M. J. , NIEBEL, A. SCHIEX , T., JAILLON, O. CHATAGNEIR O. GODIARD L. MICHELI F. KAHN D. GIANINAZZI-PEARSON V., and GAMAS P. Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. **Nucleic Acids Res** 30: 5579-5592, 2002.

LIEW, C.C., HWANG, D.M., FUNG, Y. W., LAURENSSEN, C., CUKERMAN, E., TSUI, S., and LEE, C.Y. A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *Proc Natl Acad Sci U.S.A.* 91 (22): 10645-10649, 1994.

MACHADO JR., G. R., SILVA, W. M. & IRVINE, J. E. **Sugarcane breeding in Brazil: the Copersucar program.** In: Copersucar Int. Sugarcane Breeding Workshop. São Paulo: Copersucar, p. 217-232,1987.

MARUR, C. J. **Estresse hídrico em plantas e suas conseqüências.** Área de Ecofisiologia – IAPAR, 2005 Disponível em: < <http://www.alcpar.org.br/tecnica/download.htm>> Acesso em: 10/07/2006.

METZ, J., MONARD, M. C. **Clustering hierárquico: uma metodologia para auxiliar na interpretação dos clusters.** Disponível em: <bibliotecadigital.sbc.org.br/download.php?paper=376> acesso em: 10/07/2006.

MELO JÚNIOR, A. M. DE.; Cultivo da cana-de-açúcar na usina caeté - safra 05/06. (Trabalho de conclusão de curso) CECA/UFAL,Rio Largo, 2006.

MENOSSE, M. - **Data mining em cana.** (Linha de pesquisa) CBMEG –UNICAMP, 2002. Disponível em:< <http://cafe.cbmeg.unicamp.br/principal/pesquisa.htm>> Acesso em: 14/07/2006.

MITCHELL, T. M. MACHINE LEARNING. McGraw-Hill Series in Computer Science. 1997.

NOTICIA_____Doce Genoma. Soluções para melhorar a produtividade. Disponível em:< <http://www.comciencia.br/reportagens/cana/cana3.htm>> Acesso em: 08/07/2006.

NOTÍCIAS _____ Produção nacional de cana-de-açúcar é a maior da história, diz **CONAB.** Disponível em:< <http://www.anba.com.br/noticia.php?id=9718>> Acesso em 08/07/2006.

OKUBO, K. HORI N., MATOBA R., NIYAMA T., FUKUSHIMA A., KOJIMA Y. and MATSUBARA K. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. **Nat Genet** 2:173-179. 1992.

OLIVEIRA, M. DE; VASCONCELOS, Y. Revolução no canavial. **Revista Pesquisa FAPESP** - Abril 2006 - Edição 122 Tecnologia Capa/Agroindústria. Disponível em: <<http://www.revistapesquisa.fapesp.br/pop/imprimir.php?id=2939&bid=1>> Acesso em: 09/07/2006.

OLIVÉRIO, J. L. Evolução tecnológica do setor sucroalcooleiro: a visão da indústria de equipamentos. In: CONGRESSO NACIONAL DA STAB, 8. **Anais...** Pernambuco: STAB, p.737-745. 2002.

PESSOA, A. Cana-de-açúcar. Disponível em: <<http://www.mre.gov.br/cdbrasil/itamaraty/web/port/economia/agric/producao/cana/apresent.htm>> Acesso em: 10/07/2006.

Programa Nacional de Melhoramento da Cana-de-açúcar / PLANALSUCAR, 20p. Piracicaba-SP, 1985.

REPORTAGEM_____PAIVA, R. Por dentro da cana. Grupo de 120 pesquisadores em 21 laboratórios paulistas empenha-se em desvendar o patrimônio genético da planta. *Globo Rural*, Ed. Globo S/A, 2000. Disponível em:<http://globorural.globo.com/barra.asp?d=/edic/177/rep_ciencia.htm> Acesso em 10/07/2006.

RESENDE, D.C.; LIMA, D.S.; MIRA, R. C. **Genome System**. TCC, Universidade Católica de Goiás, 2002.

SALES, J.S., Projeto de desenvolvimento e pesquisa em tecnologia de bioinformação genômica, 2003.

SILVA, P. P. DA. Avaliação de quinze safras da usina triunfo e predição de produtividade em função da precipitação pluviométrica. (Trabalho de Conclusão de Curso). CECA/UFAL. Rio Largo, 2004.

SILVA, J.G. B. S. DA, Análise por BLASTx e FASTA3 das seqüências de ESTs produzidas pelo Genoma Cana, 2004.

SINDAÇÚCAR-AL. Disponível em: <http://www.sindacucar-al.com.br/www/prodagricmedia.htm> > Acesso em 10/07/2006.

THE PLANTS database, database (Version 5.1.1) 2000. National Plant Data Center, NRCS, USDA, Baton Rouge, LA 70874-4490 USA. Disponível em: <http://plants.usda.gov>. Acesso:17/12/2006.

ÚNICA (União da Agroindústria Canavieira de São Paulo) - Agroindústria da cana-de-açúcar: Alta competitividade canavieira. Disponível em: http://www.unica.com.br/pages/agroindustria_alta.asp Acesso em: 08/07/2006 a.

ÚNICA. **PRODUÇÃO DE CANA-DE-AÇÚCAR.** Disponível em:<
http://www.orplana.com.br/corpo_estatisticas_producao.asp> Acesso em: 10/07/2006 b.

VESSONI, F. **Introdução à mineração de dados.** MV2 Sistemas de Informação. Disponível em:<<http://www.mv2.com.br/datamining.doc>> Acesso em: 10/07/2006.

VICENTINI, R., SASSAKI, F., GIMENES, M.A., MAIA, I. G. and MENOSSI, M. In silico evaluation of the *Eucalyptus* transcriptome. **Genetics and Molecular Biology**, 28, 3 (suppl), 487-495, 2005.

VOGT C. **Bioinformática, genes e inovação.** Disponível em: <
<http://www.comciencia.br/reportagens/bioinformatica/bio01.shtml>> Acesso em:
09/07/2006.

APÊNDICES

Apêndice 1. Micro-Prep Sheet**Placa:** _____ **Data de Inóculo:** ___/___/___

Inoculado por _____

Tempo de incubação: de _____ até _____ (___ horas)

Micro-Prep realizada por _____

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

Legenda: X = sem pellet, 100% de crescimento

Placa preservada em glicerol

Utilização do replicador Boekel

Reação de PCR 1 feita por _____ **Data:** ___/___/___Reação de PCR 2 feita por _____ **Data:** ___/___/___Reação de PCR 3 feita por _____ **Data:** ___/___/___Purificação 1 feita por _____ **Data:** ___/___/___Purificação 2 feita por _____ **Data:** ___/___/___Purificação 3 feita por _____ **Data:** ___/___/___

Cont....**Apêndice 1. Micro-Prep Sheet**

Gel de Sequenciamento:

GEL 1 Data ___/___/___ 1 Manhã 1 Tarde
Carregado por _____ amostras de _____ a _____

GEL 2 Data ___/___/___ 1 Manhã 1 Tarde
Carregado por _____ amostras de _____ a _____

GEL 3 Data ___/___/___ 1 Manhã 1 Tarde
Carregado por _____ amostras de _____ a _____

Gel de Agarose corrido por _____ Data: ___/___/___

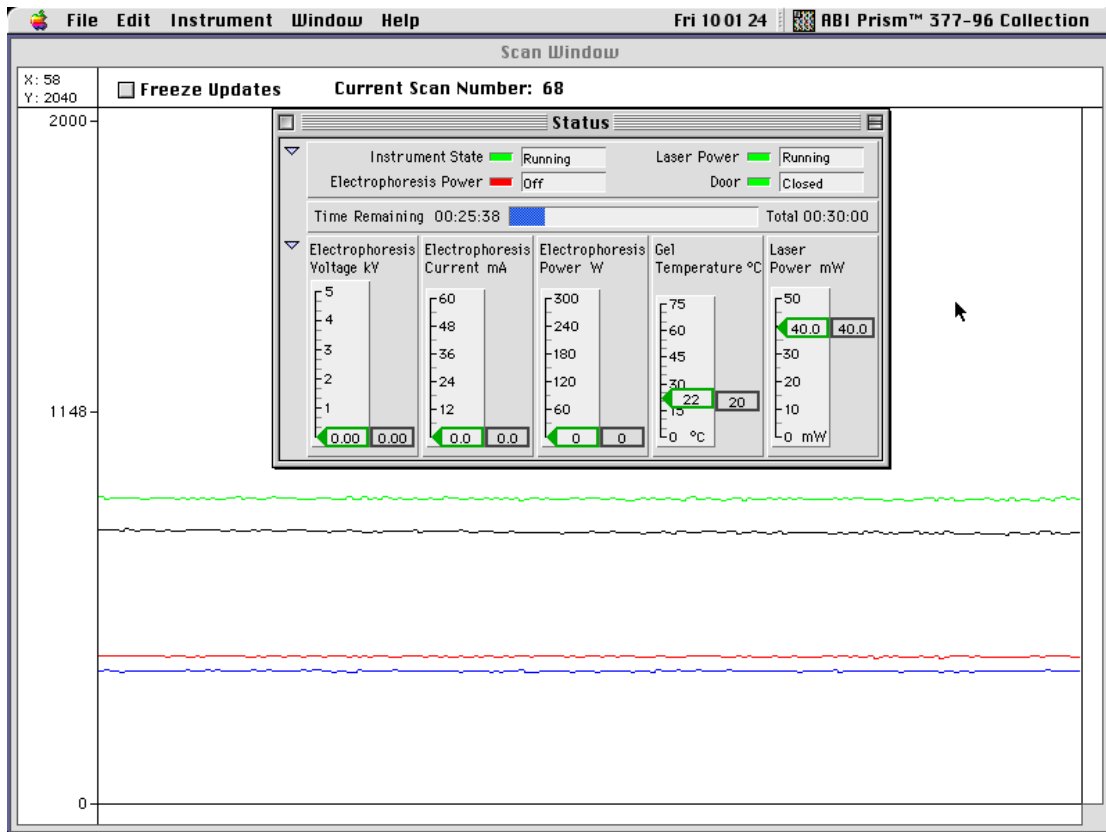
Foto do gel:

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

Legenda: X = Sem DNA, 100% de aproveitamento

Observações: _____

Apêndice 2. *Plate check* (verificação da placa), as linhas paralelas indicam ótimas condições da placa e do gel para seqüenciamento.



Apêndice 3. Sample sheet (planilha de localização das amostras).

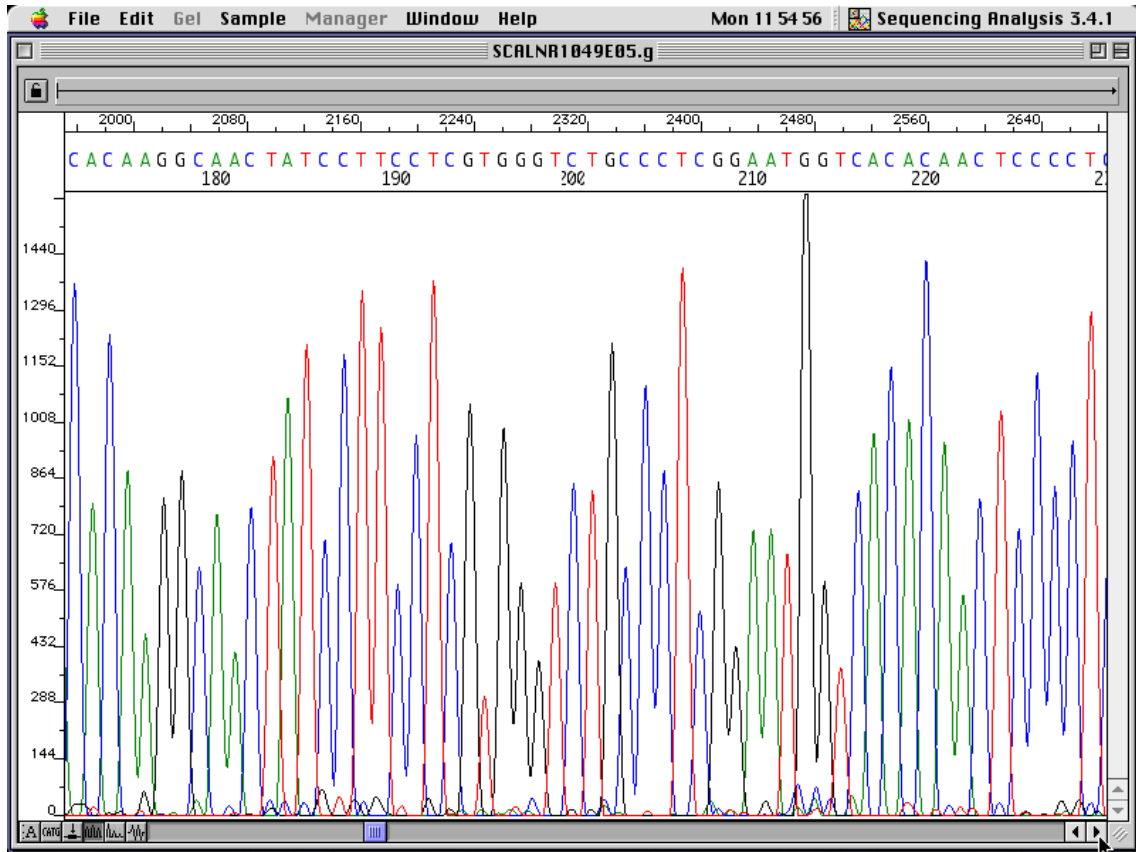
The screenshot shows a software window titled 'Sample Sheet "SCALNR1050/04-01/11/29 01.58 PM"'. The window contains a table with the following columns: Sample Name, DyeSet/Primer, Instrument File, Project Name, and Comments. The table lists 20 samples, numbered 1 to 20. Each row contains a sample name (e.g., A01.g, B01.g, C01.g), a dye set/primer (DT {BD Set Any-Prim}), an instrument file (BD2 Matrix), and a project name (indicated by a dropdown arrow). The 'Comments' column is empty for all samples.

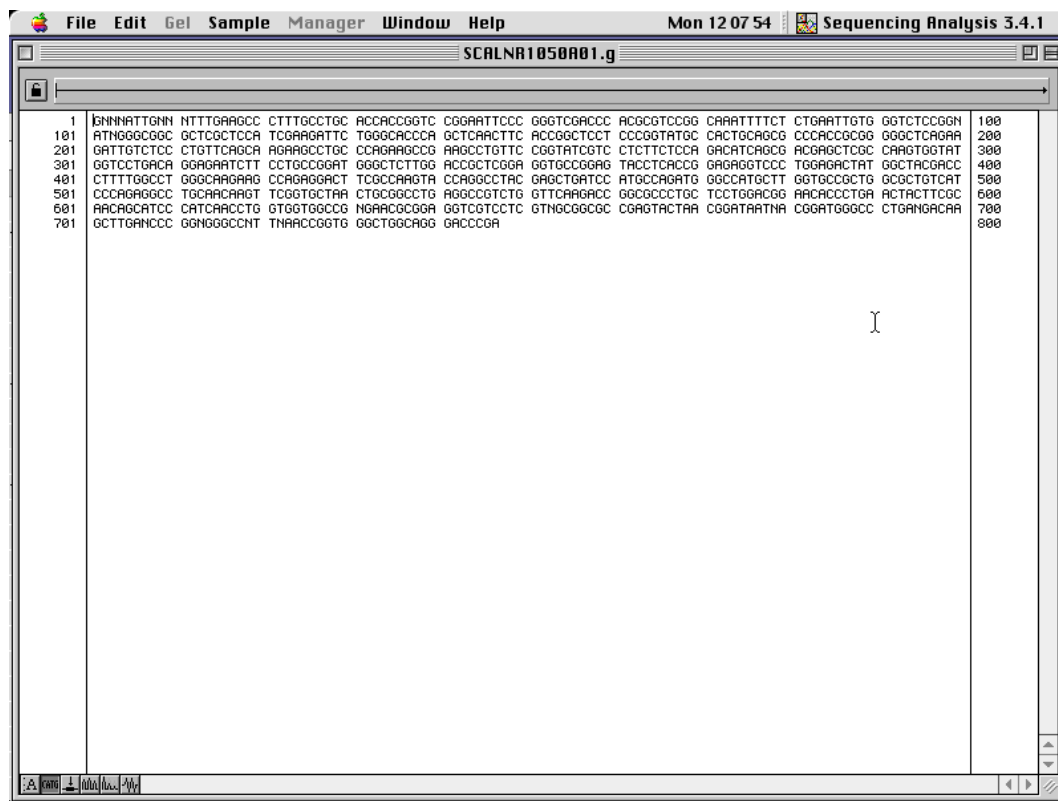
Sample Name	DyeSet/Primer	Instrument File	Project Name	Comments
A01.g	DT {BD Set Any-Prim}	BD2 Matrix		
A07.g	DT {BD Set Any-Prim}	BD2 Matrix		
A02.g	DT {BD Set Any-Prim}	BD2 Matrix		
A08.g	DT {BD Set Any-Prim}	BD2 Matrix		
A03.g	DT {BD Set Any-Prim}	BD2 Matrix		
A09.g	DT {BD Set Any-Prim}	BD2 Matrix		
B01.g	DT {BD Set Any-Prim}	BD2 Matrix		
B07.g	DT {BD Set Any-Prim}	BD2 Matrix		
B02.g	DT {BD Set Any-Prim}	BD2 Matrix		
B08.g	DT {BD Set Any-Prim}	BD2 Matrix		
B03.g	DT {BD Set Any-Prim}	BD2 Matrix		
B09.g	DT {BD Set Any-Prim}	BD2 Matrix		
C01.g	DT {BD Set Any-Prim}	BD2 Matrix		
C07.g	DT {BD Set Any-Prim}	BD2 Matrix		
C02.g	DT {BD Set Any-Prim}	BD2 Matrix		
C08.g	DT {BD Set Any-Prim}	BD2 Matrix		
C03.g	DT {BD Set Any-Prim}	BD2 Matrix		
C09.g	DT {BD Set Any-Prim}	BD2 Matrix		
D01.g	DT {BD Set Any-Prim}	BD2 Matrix		
D07.g	DT {BD Set Any-Prim}	BD2 Matrix		

Sample sheet

1	1A	25	1E	49	4A	73	4E
2	7A	26	7E	50	10A	74	10E
3	2A	27	2E	51	5A	75	5E
4	8A	28	8E	52	11A	76	11E
5	3A	29	3E	53	6A	77	6E
6	9A	30	9E	54	12A	78	12E
7	1B	31	1F	55	4B	79	4F
8	7B	32	7F	56	10B	80	10F
9	2B	33	2F	57	5B	81	5F
10	8B	34	8F	58	11B	82	11F
11	3B	35	3F	59	6B	83	6F
12	9B	36	9F	60	12B	84	12F
13	1C	37	1G	61	4C	85	4G
14	7C	38	7G	62	10C	86	10G
15	2C	39	2G	63	5C	87	5G
16	8C	40	8G	64	11C	88	11G
17	3C	41	3G	65	6C	89	6G
18	9C	42	9G	66	12C	90	12G
19	1D	43	1H	67	4D	91	4H
20	7D	44	7H	68	10D	92	10H
21	2D	45	2H	69	5D	93	5H
22	8D	46	8H	70	11D	94	11H
23	3D	47	3H	71	6D	95	6H
24	9D	48	9H	72	12D	96	12H

Apêndice 4. Eletroferograma: representações gráficas da intensidade de fluorescência emitida por cada nucleotídeo de DNA seqüenciado, indicando a confiabilidade dos dados obtidos



Apêndice 5. Sequência de 750 nucleotídeos gerada no sequenciamento automático.

The image shows a screenshot of a software window titled "Sequencing Analysis 3.4.1" with a menu bar (File, Edit, Gel, Sample Manager, Window, Help) and a status bar (Mon 12 07 54). The window title bar also displays "SCALNR1050A01.g". The main area contains a DNA sequence of 750 nucleotides, displayed in a grid format with line numbers on the left and right margins. The sequence is as follows:

1	ENNNATTGNN	NTTTGARGCC	CTTTGCCTGC	ACCACCGGTC	CGGATTCCC	GGGTCGACCC	ACGGTCGGG	CAATTTTCT	CTGATTGTG	GGTCTCCGN	100
101	ATNGGGCGGC	GCTCGCTCCA	TGGAGATTTC	TGGGCACCCA	GTCARCTTC	ACCGGTCCTC	CCCGGATGTC	CACCTGAGCG	CCGACCGCGG	GGGTCAGAA	200
201	GATTGTCTCC	CTGTTCAACA	AGAGCCCTGC	CCAGAGCCCG	AGCCCTGTTT	CGGTATCGTC	CTCTCTCCCA	GACATCAGCG	ACGACCTCGC	CAAGTGGTAT	300
301	GCTCCTGACA	GGAGATCTT	CCTGCCGAT	GGCTCTTGG	ACCGCTCGGA	GSTGCCGAG	TACCTACCCG	GAGAGTCCC	TGGAGCTAT	GGTACAGCC	400
401	CTTTTGGCCT	GGCAGAGAG	CCAGAGACT	TGCCARAGTA	CCAGCCCTAC	GAGCTGATCC	ATGCCAGATG	GGCATGCTT	GGTCCCGCTG	GGCTGTCTAT	500
501	CCAGAGGCC	TGCARAGT	TGGTGTCTA	CTGCCGCTG	AGCCCGCTG	GTTCAAGACC	GGGCCCTG	TCCTGGAGCG	AGACCCCTGA	ACTACTTCGC	600
601	ACAGCATCC	CATCACCCTG	GTGGTGGCCG	NGAGCGGGA	GSTCGTCTC	GTNGCGGCG	CGAGTACTA	CGGATATNA	CGGATGGCC	CTGANGACAA	700
701	GCTTGANCCC	GGNGGGCNT	TNACCCTG	GGCTGGCAGG	GACCCGA						800

Apêndice 6. Resultados de dez proteínas da busca por homologia no Blastn do SUCEST, utilizando a palavra chave water stress ou estresse hídrico.

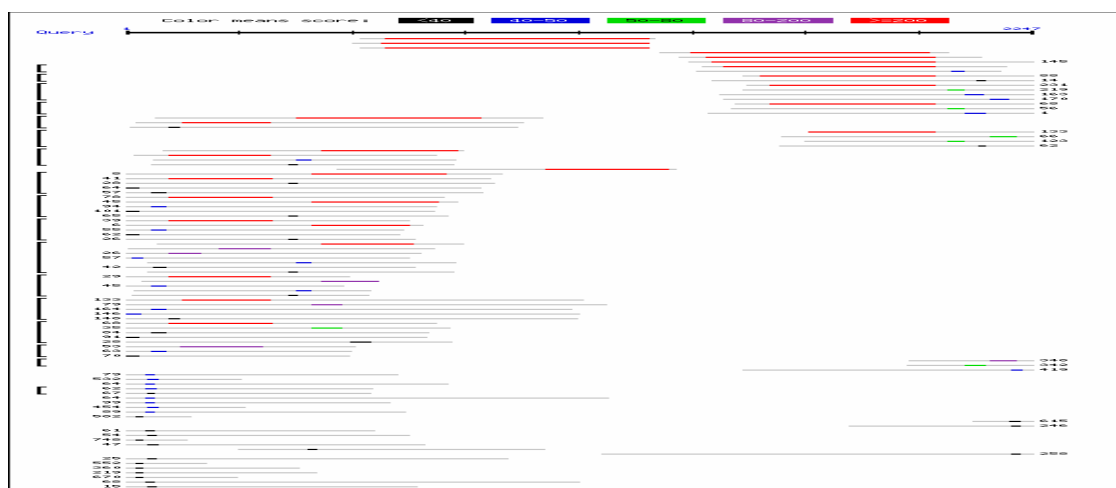
Query= gi|11121507|emb|AJ295156.1|PCO295156 *Phragmites communis* mRNA
for d-TDP-glucose dehydratase (tgp gene)
(1562 letters)



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
AJ295156	SCSBST3096C11.g	0.0	<u>1300</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCCCRZ2001D09.g	0.0	<u>1239</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCEQSB1019A04.g	0.0	<u>1197</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCCIRT3008C10.g	0.0	<u>1164</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCQGST3126F04.g	0.0	<u>1116</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCQGM2109F11.g	0.0	<u>1112</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCCST3C14H09.g	0.0	<u>1108</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCJLFL4099H10.g	0.0	<u>1096</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCCCL4014E06.g	0.0	<u>1094</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCMCSB1109G02.g	0.0	<u>1080</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCJFFL1C04H02.g	0.0	<u>1078</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCSGFL4C03F10.g	0.0	<u>1074</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCSFAD1125D09.g	0.0	<u>1065</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCCST3C04H08.g	0.0	<u>1051</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCCFL4092D05.g	0.0	<u>1047</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCVPLB2090B09.g	0.0	<u>1045</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCSFFL4082G09.g	0.0	<u>1043</u>	<i>Phragmites communis</i>	d-TDP-glucose dehydratase
AJ295156	SCBGFL5082B11.g	0.0	<u>1039</u>	<i>Phragmites communis</i>	d-TDP-glucose

					dehidratase
AJ295156	SCUTRZ3073G06.g	0.0	<u>1039</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCBGLR1113D02.g	0.0	<u>1039</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCQGHR1014F09.g	0.0	<u>1033</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCSGFL4034F08.g	0.0	<u>1021</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCAM2001F10.g	0.0	<u>1019</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCEZFL4041C03.g	0.0	<u>1015</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCSGAD1005F01.g	0.0	<u>1013</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCSBFL4012A07.g	0.0	<u>1011</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCSGAD1006H05.g	0.0	<u>1011</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCEQSB1C03G07.g	0.0	<u>1007</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCFL4120B04.g	0.0	<u>1003</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCFL4095G04.g	0.0	<u>1003</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCQSLB1049F01.g	0.0	<u>1001</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCLR1068B09.g	0.0	<u>999</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCST3C06B07.g	0.0	<u>995</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCST3C13B05.g	0.0	<u>995</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCEPRZ3129B04.g	0.0	<u>989</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCUTSB1031G08.g	0.0	<u>989</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCAM1C01B08.g	0.0	<u>985</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCMCST1054G06.g	0.0	<u>985</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCQSST3114A07.g	0.0	<u>983</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCVPAM2065D01.g	0.0	<u>979</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCFL4003E05.g	0.0	<u>977</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCACST3161C12.g	0.0	<u>977</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCMCST1054A09.g	0.0	<u>977</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCACL6010C03.g	0.0	<u>975</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCEQLB1064G12.g	0.0	<u>973</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCST3C10F02.g	0.0	<u>967</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCJFST1048B05.g	0.0	<u>965</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCRLSB1044A10.g	0.0	<u>961</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCSBST3095H02.g	0.0	<u>959</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase
AJ295156	SCCCST3006B07.g	0.0	<u>959</u>	<i>Phragmites communis</i>	d-TDP-glucose dehidratase

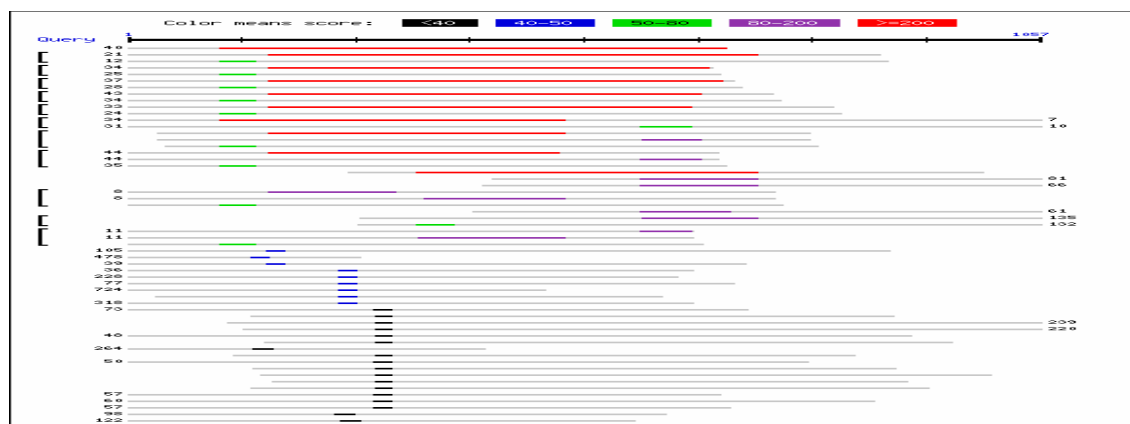
Query= gi|11139545|gb|AF276085.1|AF276085 protein kinase AKINbetagamma-1 mRNA, complete cds



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
AF276085	SCRLAD1098B08.g	0.0	<u>995</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCPILB2024A07.g	0.0	<u>977</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCJFAD1C09B02.g	0.0	<u>965</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCRLFL4103H07.g	0.0	<u>918</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCEZRZ1016F01.g	0.0	<u>878</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCQGAD1066E02.g	0.0	852	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCQGLR1087C12.g	0.0	<u>829</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCRLLR1038B01.g	0.0	<u>670</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCQGFL4077H12.g	0.0	<u>648</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCRLFL1008H04.g	0.0	<u>646</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCJFST1011B06.g	e-153	<u>545</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA

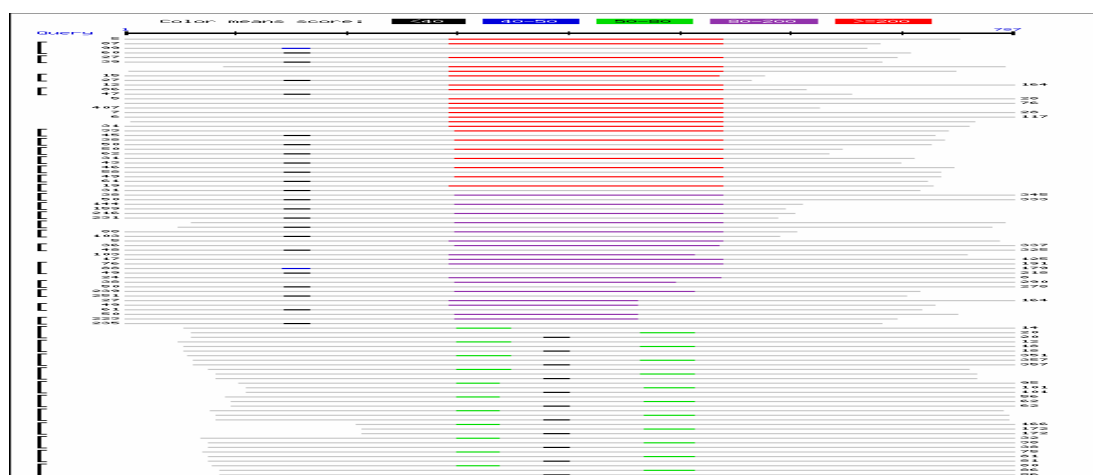
AF276085	SCJFAD1C09B02.b	e-140	<u>504</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCSBFL4065C12.g	e-120	<u>434</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCEZFL5091D11.g	e-110	<u>402</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCRFST3145A06.g	6e-99	<u>365</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCCCCL6024A07.g	3e-88	<u>329</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCEZFL4047E02.g	3e-88	<u>329</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCEPRZ3129B06.g	6e-68	<u>262</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCRUFL1115C04.g	6e-65	<u>252</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCEQFL5047G05.g	9e-58	<u>228</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA
AF276085	SCCCCL6024A05.g	3e-57	<u>226</u>	<i>Zea mays</i>	protein kinase AKINbetagamma-1 mRNA

Query= gi|10121844|gb|AF283006.1|AF283006 *Oryza sativa* cold acclimation protein WCOR413-like protein mRNA, complete cds



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
AF283006	SCSGRT2061G01.g	e-106	<u>387</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCRFST3143E07.g	e-105	<u>385</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCBGLR1120F12.g	e-103	<u>377</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCAGLR2033G02.g	2e-97	<u>359</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCEQRT1027F03.g	4e-95	<u>351</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCBGRT3011A03.g	2e-88	<u>329</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCEZHR1087E06.g	2e-78	<u>295</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCSFFL8048B02.g	3e-77	<u>291</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCVPRZ2035D10.g	2e-75	<u>285</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein
AF283006	SCRLFL4059C01.g	3e-71	<u>272</u>	<i>Oryza sativa</i>	Cold acclimation protein WCOR413-like protein

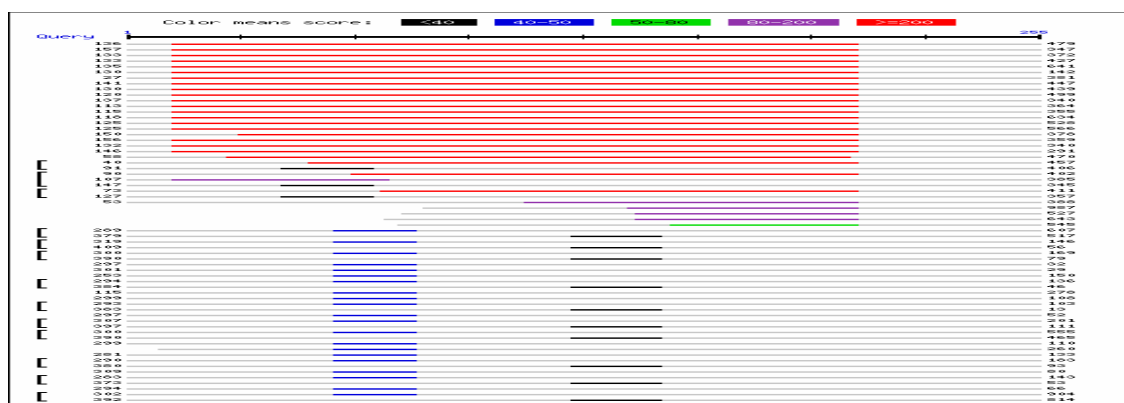
Query= gi|8335995|gb|BE040671.1|BE040671 OF04G10 OF *Oryza sativa* cDNA5' similar to water stress inducible protein, mRNA sequence



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
BE040671	SCBGLR1119B07.g	5e-54	<u>214</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCBGLR1119A06.g	5e-54	<u>214</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCBGLR1097E11.g	5e-54	<u>214</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCACLR1128A06.g	5e-54	<u>214</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCACLR1127G06.g	5e-54	<u>214</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCRUFL3065B07.g	3e-52	<u>208</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCVPRT3087B07.g	1e-51	<u>206</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCAGFL3022A04.g	1e-51	<u>206</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCCART2004G01.g	1e-51	<u>206</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible
BE040671	SCUTLR2023H08.g	1e-51	<u>206</u>	<i>Oryza sativa</i>	cDNA5' similar to water stress inducible

BE040671	SCJLLR2013B10.b	1e-51	<u>206</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCJLLR2013B10.g	1e-51	<u>206</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCCCLR2C02H09.g	1e-51	<u>206</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCBGLR1099A11.g	1e-51	<u>206</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCQSLB1050G12.g	1e-51	<u>206</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCAGSD2041E01.g	2e-50	<u>202</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCCCLB1C05C05.g	2e-50	<u>202</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCCCLB1C04D05.g	2e-50	<u>202</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCACLR2014A04.g	2e-50	<u>202</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCSBLB1036G03.g	2e-50	<u>202</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCEPLB1041D08.g	2e-50	<u>202</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible
BE040671	SCQSAM1033H03.g	7e-50	<u>200</u>	<i>Oryza sativa</i>	cdNA5' similar to water stress inducible

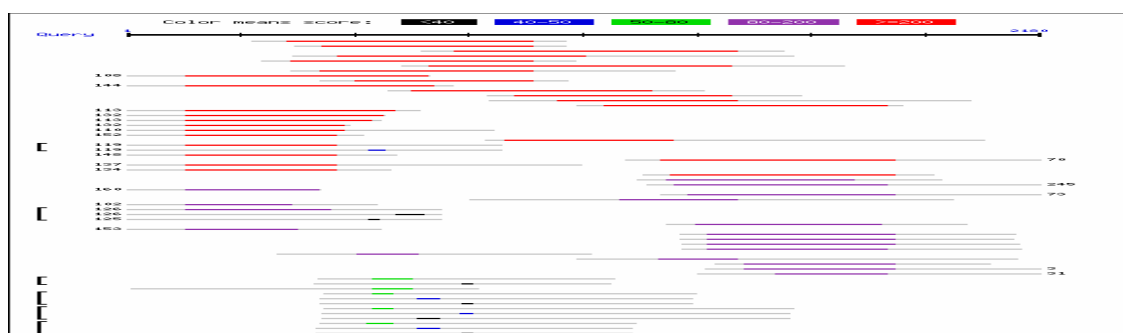
Query= gi|1000805|gb|H55025.1|H55025 HHU53a Sorghum bicolor cv. TX430 Sorghum bicolor cDNA clone HHU53 5' similar to water stress-induced protein (WSI729), rice, mRNA sequence



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
H55025	SCACRZ3036G09.g	5e-88	<u>325</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCQSHR1023D12.g	5e-88	<u>325</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCAGRT3049G05.g	1e-85	317	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCEZAD1C07F07.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCEZAD1079F08.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCCART1C01E11.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCJFRZ3C05G07.b	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCJFRZ3C05G07.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCACRZ3034A05.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCJFRT1011E08.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCACLR2007A07.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCSFHR1046F08.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCRFHR1009D01.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCQGHR1011C10.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCJFHR1C04F10.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCJFHR1034G10.g	1e-85	<u>317</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCJFHR1034G10.b	8e-84	<u>311</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCCART2C06A09.g	3e-83	<u>309</u>	<i>Sorghum bicolor</i>	water stress-induced protein

					(WSI729)
H55025	SCSFRT2069A08.g	3e-83	<u>309</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCRFRZ3055G08.g	2e-78	<u>293</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCJFRT1011E08.b	4e-73	<u>276</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCEPSD1070A05.g	4e-61	<u>236</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCAGRT3046G01.g	2e-60	<u>234</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)
H55025	SCVPRZ2041C03.g	9e-53	<u>208</u>	<i>Sorghum bicolor</i>	water stress-induced protein (WSI729)

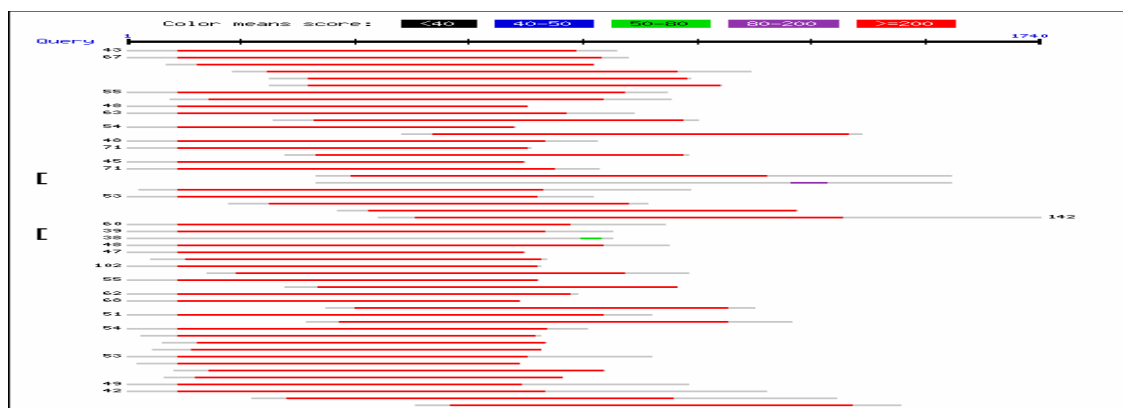
Query= gi|2642237|gb|AF031241.1|AF031241 Glycine max endoplasmic reticulum HSC70-cognate binding protein precursor (BIP) mRNA, complete cds



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
AF031241	SCBGAD1029B10.g	e-130	<u>468</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCUTFL1057G06.g	e-122	<u>442</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCJFST1047F01.g	e-120	<u>436</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCQGFL4076A11.g	e-118	<u>430</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCSBFL4065H10.g	e-117	<u>426</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCRFRZ3058E12.g	e-115	<u>418</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCSFFL4018A03.g	e-110	<u>402</u>	<i>Glycine max</i>	protein precursor (BIP)

AF031241	SCVPST1061F02.g	e-105	<u>387</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCUTSD1086B10.g	e-105	<u>385</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCRFAM2128G06.g	9e-98	<u>361</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCSGHR1071C09.g	4e-97	<u>359</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCMCSB1113D08.g	5e-93	<u>345</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCEQSB1C10D05.g	8e-92	<u>341</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCMCL6054G06.g	3e-88	<u>329</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCJFST1018G02.g	2e-77	<u>293</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCSFFL4083B06.g	4e-72	<u>276</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCSFBS1072E09.g	4e-69	<u>266</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCVPST1059D02.g	9e-64	<u>248</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCJFST1012C05.g	9e-64	<u>248</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCQSRT2032C09.g	9e-64	<u>248</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCEQHR1078A03.g	2e-61	<u>240</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCJLLR1054F05.g	9e-61	<u>238</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCBFAM2022G09.g	9e-61	<u>238</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCCCRT3008D10.g	2e-58	<u>230</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCEQRZ3024A10.g	5e-56	<u>222</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCEPAM1019E12.g	1e-53	<u>214</u>	<i>Glycine max</i>	protein precursor (BIP)
AF031241	SCJLFL4098A03.g	8e-52	<u>208</u>	<i>Glycine max</i>	protein precursor (BIP)

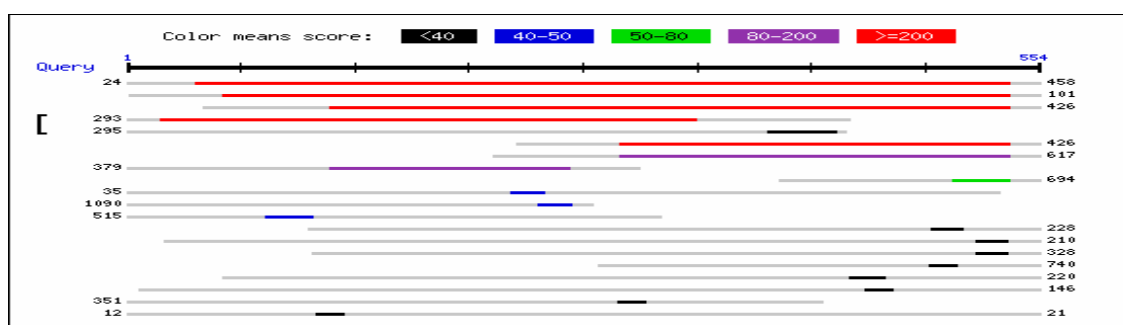
Query= gi|2613142|gb|AF030548.1|AF030548 *Oryza sativa* tubulin (RiP3)mRNA, complete cds



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
AF030548	SCQGHR1011C08.g	0.0	1019	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJFRZ3C06A02.g	0.0	966	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCCCFL4090F02.g	0.0	960	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCVPFL1133B11.g	0.0	956	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCEQRT3C05B06.g	0.0	932	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCACSB1123B09.g	0.0	932	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCSBST3097G12.g	0.0	930	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJLSB1067A10.g	0.0	924	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCCCFL4090E02.g	0.0	918	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCSBFL4063F11.g	0.0	912	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCEQRT3C05C10.g	0.0	910	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJFFL1C02E06.g	0.0	904	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCRFRZ3057A10.g	0.0	904	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCRFBAM2130E03.g	0.0	902	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJLAM2092B10.g	0.0	900	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCCCAD1003G03.g	0.0	898	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCAGFL8042F03.g	0.0	892	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCVPFL1134F05.g	0.0	890	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCCCFL5093G11.g	0.0	890	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCVPFL1071B01.g	0.0	888	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJLFL4100D03.g	0.0	884	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCSFSB1071G02.g	0.0	884	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCRFRZ3057B06.g	0.0	884	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCBFRT1070A09.g	0.0	881	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJFHR1031E05.g	0.0	881	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCVPFL1068E02.g	0.0	879	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCSBST3100D04.g	0.0	877	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCQSR2036F02.g	0.0	875	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCMCR2105G07.g	0.0	875	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCRFBST1045D12.g	0.0	871	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCRFFL4009E03.g	0.0	867	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJLFL1049G09.g	0.0	867	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCSBFL1045B07.g	0.0	867	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCRFRZ3058F05.g	0.0	867	<i>Oryza sativa</i>	tubulin (RiP3)

AF030548	SCSBFL5014F12.g	0.0	865	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCAGLV1041D07.g	0.0	863	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCQGSB1143C05.g	0.0	863	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJLFL1048E03.g	0.0	861	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCJLFL1050A02.g	0.0	859	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCCCST1005F06.g	0.0	859	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCSGRT2064C10.g	0.0	859	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCBGR1049A05.g	0.0	857	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCCCSD1001C06.g	0.0	855	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCMCRZ3067B03.g	0.0	855	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCEZRZ3100H03.g	0.0	855	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCSGST3117G02.g	0.0	851	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCQGRZ3014B12.g	0.0	847	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCEQAM2039D04.g	0.0	845	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCEQSB1C01E05.g	0.0	841	<i>Oryza sativa</i>	tubulin (RiP3)
AF030548	SCRFRZ3058D12.g	0.0	841	<i>Oryza sativa</i>	tubulin (RiP3)

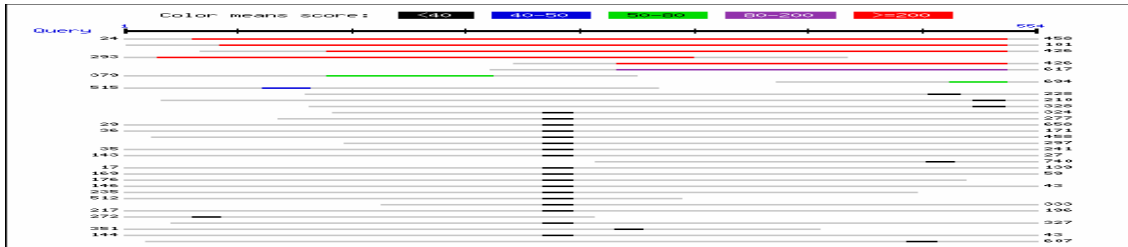
Query= gi|1432142|gb|U61179.1|TAU61179 *Triticum aestivum* ADP-glucose pyrophosphorylase (WAL2) mRNA, partial cds



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
U61179	SCRLFL4103H07.g	0.0	983	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCRLAD1098B08.g	0.0	<u>948</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCPILB2024A07.g	0.0	<u>930</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCEZRZ1016F01.g	0.0	<u>924</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCJFAD1C09B02.g	0.0	<u>910</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCQGAD1066E02.g	0.0	<u>902</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase

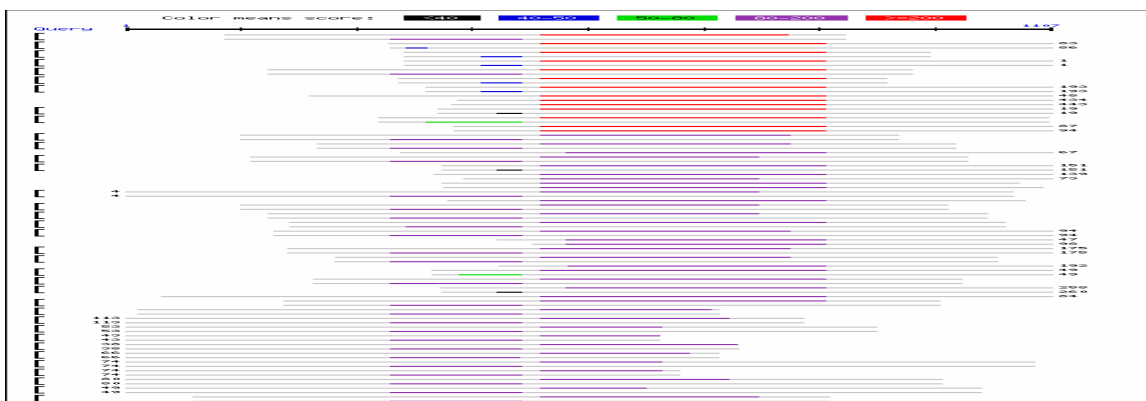
U61179	SCQGLR1087C12.g	0.0	<u>868</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCRLLR1038B01.g	0.0	<u>702</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCQGFL4077H12.g	0.0	<u>694</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCRLFL1008H04.g	0.0	<u>692</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCJFST1011B06.g	e-158	<u>561</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCJFAD1C09B02.b	e-148	<u>529</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCSBFL4065C12.g	e-118	<u>428</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCEZFL5091D11.g	e-103	<u>379</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCRFST3145A06.g	1e-99	<u>367</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCCCCL6024A07.g	3e-88	<u>329</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCEZFL4047E02.g	1e-81	<u>307</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCEPRZ3129B06.g	7e-74	<u>281</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase
U61179	SCRUFL1115C04.g	5e-53	<u>212</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase

Query= gi|1432140|gb|U61178.1|TAU61178 *Triticum aestivum* ADP-glucose pyrophosphorylase (WAL1) mRNA, partial cds (554 letters)



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
U61178	SCVPLB2089F09.g	e-132	<u>472</u>	<i>Triticum aestivum</i>	ADP-glucose Pyrophosphorylase
U61178	SCEPFL3080F12.g	e-117	<u>424</u>	<i>Triticum aestivum</i>	ADP-glucose Pyrophosphorylase
U61178	SCBFLR1039F10.g	e-109	<u>398</u>	<i>Triticum aestivum</i>	ADP-glucose Pyrophosphorylase
U61178	SCRLSB1043H12.g	4e-78	<u>293</u>	<i>Triticum aestivum</i>	ADP-glucose Pyrophosphorylase
U61178	SCMCRT3084F01.g	1e-56	<u>222</u>	<i>Triticum aestivum</i>	ADP-glucose pyrophosphorylase

Query= gi|395045|emb|X73848.1|LETRAMP2 *L. esculentum* mRNA for tomato ripening membrane protein, clone pNY507 (1107 letters)



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
X73848	SCMCCL6058D08.g	3e-53	<u>212</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73848	SCSGLV1010C11.b	1e-52	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73848	SCEQRT3018D10.g	1e-52	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73848	SCEZAD1C05E12.g	1e-52	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73848	SCRLFL1011F12.g	1e-52	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73848	SCRUAD1060E08.g	1e-52	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73848	SCQGSD2044H03.g	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73848	SCQSFL3037B05.b	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73848	Scqgst3126h12.G	2e-50	<u>202</u>	<i>L.Esculentum</i>	Tomato Ripening Membrane Protein
X73848	Sceqrz3021e03.G	2e-50	<u>202</u>	<i>L.Esculentum</i>	Tomato Ripening Membrane Protein
X73848	Scezrz1015g12.G	2e-50	<u>202</u>	<i>L.Esculentum</i>	Tomato Ripening Membrane Protein
X73848	Scvprt2080e01.G	2e-50	<u>202</u>	<i>L.Esculentum</i>	Tomato Ripening Membrane Protein
X73848	Sccclb1021h10.G	2e-50	<u>202</u>	<i>L.Esculentum</i>	Tomato Ripening Membrane Protein
X73848	Sccclb1021g10.G	2e-50	<u>202</u>	<i>L.Esculentum</i>	Tomato Ripening Membrane Protein

					Protein
--	--	--	--	--	---------

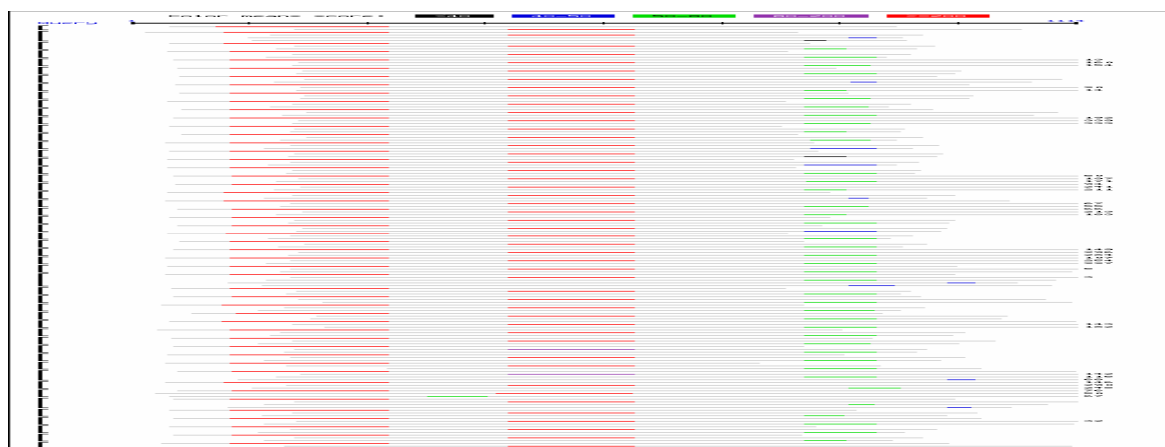
Query= gi|395043|emb|X73847.1|LETRAMP1 *L.esculentum* mRNA for tomato ripening membrane protein, clone pTOM75 (859 letters)



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
X73847	SCMCCL6058D08.g	2e-53	<u>212</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCSGLV1010C11.b	8e-53	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCEQRT3018D10.g	8e-53	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCEZAD1C05E12.g	8e-53	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCRLFL1011F12.g	8e-53	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCRUAD1060E08.g	8e-53	<u>210</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCQGS2044H03.g	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCQSFL3037B05.b	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCQGST3126H12.g	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCEQRZ3021E03.g	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein

X73847	SCEZRZ1015G12.g	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCVPRT2080E01.g	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCCCLB1021H10.g	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein
X73847	SCCCLB1021G10.g	2e-50	<u>202</u>	<i>L.esculentum</i>	Tomato Ripening Membrane Protein

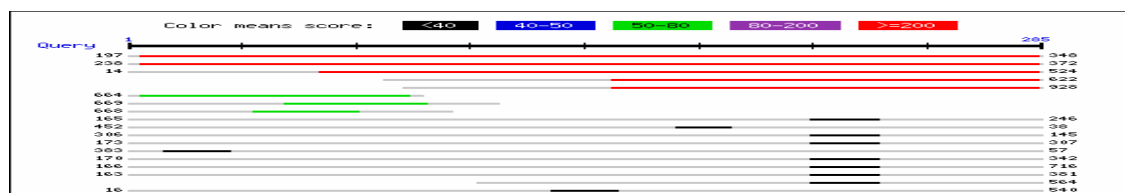
Query= gi|22312|emb|X12564.1|ZMGRP Maize ABA-inducible gene for glycine-rich protein (ABA = abscisic acid) (1114 letters)



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
X12564	SCEZFL4042C04.g	4e-86	<u>321</u>	Maize	glycine-rich protein
X12564	SCCCFL4089B09.g	1e-79	<u>299</u>	Maize	glycine-rich protein
X12564	SCSBFL1045D01.g	1e-79	<u>299</u>	Maize	glycine-rich protein
X12564	SCSGLV1005C03.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCCCSD1001G05.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCEQLB2019D06.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCRFFL5038D01.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCJFST1011A01.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCJLRZ3074A05.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCEPRZ3085D03.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCJFRZ2007A06.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCJFRT2055A03.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCACLR2029A08.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCSFHR1046D10.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCRFHR1006C02.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCSGCL6071E07.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCRFAM2070G01.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCRFAM2070E01.g	9e-78	<u>293</u>	Maize	glycine-rich protein
X12564	SCBFLB2093H03.g	3e-77	<u>291</u>	Maize	glycine-rich protein
X12564	SCBFFL1142B04.g	3e-77	<u>291</u>	Maize	glycine-rich protein
X12564	SCBGFL5080F02.g	3e-77	<u>291</u>	Maize	glycine-rich protein
X12564	SCJFST1015A11.g	3e-77	<u>291</u>	Maize	glycine-rich protein
X12564	SCJFRZ2011H03.g	3e-77	<u>291</u>	Maize	glycine-rich protein
X12564	SCSFLR2009A08.g	3e-77	<u>291</u>	Maize	glycine-rich protein
X12564	SCCCLR2C01A07.g	3e-77	<u>291</u>	Maize	glycine-rich protein
X12564	SCRFAM2071A09.g	1e-76	<u>289</u>	Maize	glycine-rich protein

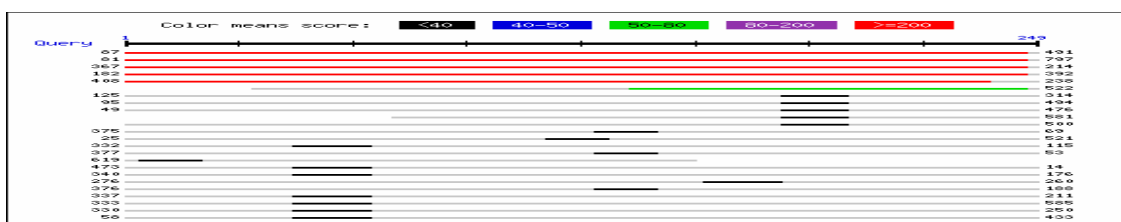
X12564	SCRFLV1037F03.g	2e-75	<u>285</u>	Maize	glycine-rich protein
X12564	SCJFRZ2012H09.g	2e-75	<u>285</u>	Maize	glycine-rich protein
X12564	SCJFRZ2012D05.g	2e-75	<u>285</u>	Maize	glycine-rich protein
X12564	SCJFRZ2005F05.g	2e-75	<u>285</u>	Maize	glycine-rich protein
X12564	SCJFRZ2005F04.g	2e-75	<u>285</u>	Maize	glycine-rich protein
X12564	SCJLRT2051D11.g	2e-75	<u>285</u>	Maize	glycine-rich protein
X12564	SCCLR1121A05.g	9e-75	<u>283</u>	Maize	glycine-rich protein
X12564	SCACSB1122G03.g	3e-74	<u>281</u>	Maize	glycine-rich protein
X12564	SCQGAM2026F07.g	3e-74	<u>281</u>	Maize	glycine-rich protein
X12564	SCAGLV1040F04.g	3e-71	<u>272</u>	Maize	glycine-rich protein
X12564	SCCCLB2002F08.g	3e-71	<u>272</u>	Maize	glycine-rich protein
X12564	SCCCFL5060D08.g	3e-71	<u>272</u>	Maize	glycine-rich protein
X12564	SCJFFL1C02D08.g	3e-71	<u>272</u>	Maize	glycine-rich protein
X12564	SCCCT2C02H10.g	3e-71	<u>272</u>	Maize	glycine-rich protein
X12564	SCJFRZ2009G02.g	1e-70	<u>270</u>	Maize	glycine-rich protein
X12564	SCBFRZ2049D07.g	1e-70	<u>270</u>	Maize	glycine-rich protein
X12564	SCEQHR1082B11.g	1e-70	<u>270</u>	Maize	glycine-rich protein
X12564	SCRFRZ3058G04.g	8e-69	<u>264</u>	Maize	glycine-rich protein
X12564	SCJLRT1023D06.g	8e-69	<u>264</u>	Maize	glycine-rich protein
X12564	SCRFAM2127G09.g	8e-69	<u>264</u>	Maize	glycine-rich protein
X12564	SCAGFL8013C06.g	3e-68	<u>262</u>	Maize	glycine-rich protein
X12564	SCACLR2022D08.g	3e-68	<u>262</u>	Maize	glycine-rich protein
X12564	SCJFRZ2005H12.g	2e-66	<u>256</u>	Maize	glycine-rich protein

Query= gi|1298583|gb|W21647.1|W21647 zEST00935 Maize Leaf, Stratagene #937005 Zea mays cDNA clone csuh00935 5' end similar to water stress induced protein, mRNA sequence (285 letters)



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
W21647	SCBGLB2011B01.g	e-122	<u>438</u>	<i>Zea mays</i>	Water Stress Induced Protein
W21647	SCCCLR1075C10.g	e-122	<u>438</u>	<i>Zea mays</i>	Water Stress Induced Protein
W21647	SCQSSB1059A10.g	e-101	<u>369</u>	<i>Zea mays</i>	Water Stress Induced Protein
W21647	SCSFLV1046C01.g	3e-62	<u>240</u>	<i>Zea mays</i>	Water Stress Induced Protein
W21647	SCVPLB2C06C07.g	3e-62	<u>240</u>	<i>Zea mays</i>	Water Stress Induced Protein

Query= gi|1298574|gb|W21638.1|W21638 zEST00924 Maize Leaf, Stratagene #937005 Zea mays cDNA clone csuh00924 5' end similar to water stress induced protein, mRNA sequence (249 letters)



ACCESSION	SUGARCANE CLUSTER	E-VALUE	SCORE	ORGANISM	PROTEIN
W21638	SCSFLV1046C01.g	e-115	<u>416</u>	<i>Zea mays</i>	water stress induced protein
W21638	SCVPLB2C06C07.g	e-115	<u>416</u>	<i>Zea mays</i>	water stress induced protein
W21638	SCBGLB2011B01.g	e-115	<u>416</u>	<i>Zea mays</i>	water stress induced protein
W21638	SCQSSB1059A10.g	e-111	<u>402</u>	<i>Zea mays</i>	water stress induced protein
W21638	SCCCLR1075C10.g	e-107	<u>391</u>	<i>Zea mays</i>	water stress induced protein