UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA

**Elucidação do destino metabólico de glicose no fungo
filamentoso *Trichoderma reesei* por análise EST
(Expressed Sequence Tags) e “microarrays” de cDNA**

FELIPE SANTIAGO CHAMBERGO ALCALDE

TESE DE DOUTORADO

Orientador: Prof. Dr. Hamza El-Dorrry

São Paulo

01/03/2002

Ficha Catalográfica

Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Chambergo Alcalde, Felipe Santiago
C443e Elucidação do destino metabólico de glicose no fungo
filamentoso *Trichoderma reesei* por análise EST (Expressed
Sequence Tags) e “microarrays” de cDNA / Felipe Santiago
Chambergo Alcalde. -- São Paulo, 2002.
108p.

Tese (doutorado) - Instituto de Química da Universidade
de São Paulo. Departamento de Bioquímica.

Orientador : El-Dorry, Hamza Fahmi Ali

1. Biologia molecular 2. Expressão gênica 3.
Trichoderma : Micologia I. T. II. El-Dorry, Hamza
Fahmi Ali, orientador.

574.88 CDD

**‘Elucidação do Destino Metabólico de
Glicose no Fungo Filamentoso *Trichoderma
reesei* por Análise EST (Expressed Sequence
Tags) e “Microarrays” de cDNA’**

FELIPE SANTIAGO CHAMBERGO ALCALDE

Tese de Doutorado submetida ao Instituto de
Química da Universidade de São Paulo como parte
dos requisitos necessários à obtenção do grau de
Doutor em Ciências – Área: Bioquímica.

Aprovado por:

Prof. Dr. HAMZA FAHMI ALI EL DORRY
IQ – USP
(Orientador e Presidente)

Profa. Dra. BIANCA SILVANA ZINGALES
IQ – USP

Profa. Dra. ALINE MARIA DA SILVA
IQ – USP

Prof. Dr. FRANCISCO GORGÔNIO DA NÓBREGA
UNIVAP

Prof. Dr. ARMANDO MORAES VENTURA
ICB - USP

SÃO PAULO
01 DE MARÇO 2002.

In Memoriam

ANDRÉS FELIPE

E

SUSANA CECILIA

A mis Padres Carlos y Luisa, mi Esposa Estela Ynés y mis pequeños Sergio y Ricardo, que en todo momento confiaron en mi y desde la distancia me dieron la fuerza necesaria para seguir adelante, a ellos mi eterno Amor.

Agradecimentos

Ao Professor Dr. Hamza El-Dorry, pela amizade, confiança, orientação e apoio durante todo o desenvolvimento deste trabalho, no qual os dois apreendemos muito.

Aos Profs. Dr. João Pedro Simon Farah, Prof. Dr. José Abrahão-Neto e Dr. Jaime Leyton pela sua amizade e importante colaboração na realização de este trabalho.

À meus irmãos Alberto, Juan, Pedro, Luis, César, Lizeth e Luisa, ao Sr. Mario Valencia (q.e.p.d.) e Sra. Ysidora Morante, a Arecio, Julia, Mario, Baltazar e Célia pelo incentivo, e fé depositada na minha pessoa.

Aos amigos, Dr. Euclides Matheucci Jr. e Augusto Sávio Peixoto Ramos pelo companheirismo, amizade, apoio e pelas boas idéias e frases filosóficas que davam um significado ao trabalho do dia a dia.

Aos colegas Eric Bonaccorsi, Ari José Scattone Ferreira, Augusto Peixoto Ramos e Dr. José Ribamar Ferreira Júnior pela sua participação e decidido apoio mostrando uma vez mais que a “A união faz a força”.

Á Dra. Nathalie Cella, Daniela Steconci, Sílvia Regina Blanco que iluminarem com seu sorriso e conversa os momentos certos do dia a dia.

Á Wilton José, pelo companheirismo á Zilda e Adriano pela ajuda indispensável e sincera amizade.

Á todos os Professores do Departamento de Bioquímica, colegas e funcionários do Instituto de Química que no dia a dia me fizeram sentir como em minha “terra”.

Ao Conselho Nacional de Pesquisa (CNPq) e a Fundação de Amparo á Pesquisa do Estado de São Paulo (FAPESP), pelas bolsas de Mestrado e Doutorado direto, respectivamente, concedidas.

ÍNDICE

Summary.....	ix
Resumo.....	x
Abreviaturas.....	xi
Introdução.....	1
Metabolismo Aeróbico e Anaeróbico de Glicose.....	4
<i>Trichoderma</i>	9
Análise estrutural e funcional do Genoma.....	11
Genômica estrutural: Expressed Sequence Tags (ESTs).....	13
Genômica funcional: Análise global de expressão gênica.....	16
Objetivos.....	21
Materiais e métodos	22
Reagentes e equipamentos.....	22
Manipulação de DNA e técnicas em geral.....	22
Cepas e condições de cultura.....	22
Extração de DNA e RNA	23
Construção da biblioteca de cDNA.....	24
Seqüenciamento e análise computacional.....	24
Microarray e análise computacional dos dados.....	27
Northern blot.....	38
Atividade enzimática de álcool desidrogenase (ADH).....	38
Resultados.....	40
Estabelecimento da base de dados EST de <i>T. reesei</i>	40
Montagem e avaliação da biblioteca de cDNA.....	40
Seqüenciamento em larga escala.....	45
Anotação e construção da base de dados EST.....	46
Cinética do destino de Glicose em <i>T.reesei</i> , determinada por análise transcricional através de “Microarray”.....	49
Obtenção de dados.....	63
Normalização de dados.....	65

Agrupamento hierárquico (“Hierarchical Clustering”)	
para identificação dos genes regulados por glicose.....	69
Análise de expressão dos genes envolvidos no metabolismo de Glicose.....	73
Discussão.....	82
Biblioteca de cDNA	82
Base de dados EST de <i>T. reesei</i>	83
Análise de expressão gênica.....	87
Conclusões	95
Referências.....	97
Curriculum vitae.....	108

Índice de Figuras

Figura 01 - Representação esquemática de glicólise, TCA, gliconeogênese e via alternativa do metabolismo de piruvato.....	2
Figura 02 - Representação esquemática da redução de piruvato a Lactato ou Etanol e a regeneração do NAD ⁺ em organismos eucariotos.....	3
Figura 03 - Perfil geral da passagem do metabolismo anaeróbico (fermentação) de glicose a metabolismo aeróbico (respiração) em <i>S. cerevisiae</i>	5
Figura 04 - Padrão de expressão dos genes de Glicólise, TCA, gliconeogênese e a via alternativa do metabolismo de piruvato durante a transição do metabolismo anaeróbico (fermentação) de glicose ao metabolismo aeróbico (respiração) em <i>S. cerevisiae</i>	8
Figura 05 - Métodos de construção de bibliotecas de cDNA para seqüenciamento e geração de dados de EST.....	14
Figura 06 - Metodologia de análise de expressão gênica por cDNA microarray.....	18
Figura 07 - Perfil da lâmina de microarray.....	33
Figura 08 - Estrutura de Cy3, Cy5, amino allyl-dUTP e dUTP ligado à Cy5.....	34
Figura 09 - Eletroforese em gel de agarose, 0,8%, de produto de PCR dos insertos de 24 clones escolhidos aleatoriamente da biblioteca TrEST-A de <i>T. reesei</i>	41
Figura 10 - Classificação dos 1.151 ESTs transcritos únicos de <i>T. reesei</i>	48
Figura 11 - Página de Internet.....	50
Figura 12 - Representação esquemática de glicólise, TCA, gliconeogênese e via alternativa de metabolismo de piruvato (Pyruvate bypas) apontando os genes identificados em <i>T. reesei</i>	58
Figura 13 - Eletroforese em gel de agarose 1% do produto purificado dos PCR dos cDNAs, representando uma das 4 placas de 384 poços utilizadas neste estudo....	59
Figura 14 - Concentração de glicose e densidade celular durante o crescimento de <i>T. reesei</i>	61
Figura 15 - Eletroforese em gel desnaturante de agarose 1%, do RNA extraído das aliquotas retiradas a distintos tempos (Fig.14).....	61
Figura 16 - Imagens da fluorescência da lâmina A4.....	62
Figura 17 - Imagem composta da lâmina A4.....	64

Figura 18 - Análise estatístico da razão correspondente aos 32 elementos do gene controle Q-fago lambda, depositados na lâmina A1 (A), A2 (B), A3 (C) e A4	66
Figura 19 - Agrupamento hierárquico de genes de <i>T. reesei</i>	71
Figura 20 - Perfil de expressão de genes reprimidos por glicose.....	72
Figura 21 - Imagem nas lâminas A1, A2, A3 e A4, dos 14 genes selecionados na Figura 20.....	74
Figura 22 - Comparação do perfil de expressão gênica de genes da glicólise, gliconeogênese, ciclo do TCA e via alternativa de piruvato (piruvate bypass) durante o esgotamento de glicose em <i>T. reesei</i> e <i>S. cerevisiae</i>	79
Figura 23 – Perfil da expressão gênica, analisada por Northern blot, de 11 genes selecionados para validação dos resultados obtidos por microarray.....	80

Índice de Tabelas

Tabela 01 - Oligonucleotídeos empregados na amplificação de genes de <i>T. reesei</i> , incluídos no microarray.....	28
Tabela 02 - Análise de qualidade da biblioteca TrEST-A.....	43
Tabela 03 - Número de ESTs de <i>T. reesei</i> com similaridade á seqüências depositadas na base pública de dados GenBank.....	43
Tabela 04 - Distribuição dos ESTs de <i>T. reesei</i>	47
Tabela 05 - Genes com maior número de ocorrências.....	47
Tabela 06 - Formato de submissão de seqüência EST ao dbEST GenBank.....	53
Tabela 07 - ESTs de <i>Trichoderma reesei</i> envolvidos na glicólise, ciclo de TCA, via alternativa do metabolismo de piruvato e gliconeogênese.....	55
Tabela 08 - Determinação de atividade ADH em <i>T. reesei</i>	57
Tabela 09 - Valores de fluorescência obtidos para os primeiros 20 elementos da lâmina A4.....	57
Tabela 10 - Normalização de dados de fluorescência da lâmina A4, correspondente aos 14 genes selecionados na figura	76
Tabela 11 - Valores normalizados correspondentes aos 14 genes selecionados na figura 20.....	77

SUMMARY

Despite the intense interest in the metabolic regulation and evolution of the ATP-producing pathways, the long-standing question of why most multicellular microorganisms metabolize glucose by respiration rather than fermentation remains unanswered. One such microorganism is the cellulolytic fungus *Trichoderma reesei* (*Hypocrea jecorina*).

Using EST analysis and cDNA microarrays, we find that in *T. reesei* expression of the genes encoding the enzymes of the TCA is programmed in a way that favors the oxidation of pyruvate via the TCA cycle rather than its reduction to ethanol by fermentation. Moreover, the results indicate that acetaldehyde may be channeled into acetate rather than ethanol, thus preventing the regeneration of NAD^+ , a pivotal product required for anaerobic metabolism.

The studies also point out that the regulatory machinery controlled by glucose was most probably the target of evolutionary pressure that directed the flow of metabolites into respiratory metabolism rather than fermentation.

RESUMO

Apesar do intenso interesse na regulação metabólica e evolução das vias produtoras de ATP, o porquê de a maioria dos microorganismos multicelulares metabolizarem glicose através de respiração, ao invés da fermentação, ainda permanece sem resposta. Um desses microorganismos é o fungo celulolítico *Trichoderma reesei* (*Hypocrea jecorina*).

Usando análise EST e microarrays de cDNA, foi estabelecido, em *T. reesei*, que a expressão dos genes que codificam as enzimas do ciclo de TCA é programada de tal modo a favorecer a oxidação de piruvato pelo ciclo de TCA, ao invés de sua redução a etanol, através da fermentação. Além disso, os resultados indicam que acetaldeído pode ser convertido a acetato, e não a etanol, prevenindo a regeneração de NAD^+ , um produto chave requerido para o metabolismo anaeróbico.

Os estudos também mostram que a maquinaria de controle regulatório por glicose, foi, provavelmente, objeto de pressão evolutiva, a qual dirigiu o fluxo metabólico à respiração, e não à fermentação.

Abreviaturas

ACO - aconitase
ACS – acetil coenzima A sintase
ACT - actina
ADH – álcool desidrogenase
ALD, ALD1, ALD2 - aldeído desidrogenase
ATP – adenosina trifosfato
CIT- citrato sintase
DHA-P – diidroxiacetona fosfato
ENO - enolase
FBA – frutose-1,6-bifosfato aldolase
FRU-1,6-P – frutose-1,6-bifosfato
FUM - fumarase
GA-3P – gliceraldeído-3-fosfato
GPM – fosfoglicerato mutase
IDH – isocitrato desidrogenase
KDH - α -cetoglutarato desidrogenase
MDH – malato desidrogenase
NADH, NAD⁺ - formas reduzida e oxidada da nicotinamida adenina dinucleotídeo
PCK – fosfoenolpiruvato carboxiquinase
PDA - diidrolipoamide S-acetiltransferase
PDC – piruvato descarboxilase
PEP - fosfoenolpiruvato
PGK – fosfoglicerato quinase
PYK - piruvato quinase
SDH –succinato desidrogenase
TCA – ciclo do ácido tricarboxílico
TDH – gliceraldeído -3-fosfato desidrogenase
TPI - triose fosfato isomerase
YGR – succinil-CoA sintetase

INTRODUÇÃO

Glicose, o mais abundante monossacarídeo na natureza, é a fonte preferencial de carbono e energia dos microrganismos eucarióticos (Vaulont *et al.*, 2000). O metabolismo da glicose ocorre pela glicólise, uma via central e universal, através da qual ocorre, na maioria das células, o principal fluxo do carbono (Johnston, 1999). No curso da evolução, a glicólise, constituída de uma série de 10 reações enzimáticas (Fig. 1), manteve-se altamente conservada. Os mecanismos que regulam o destino do piruvato, o produto final da via glicolítica, e das vias produtoras de ATP, por outro lado, estiveram sob pressão seletiva durante a evolução (Pfeiffer *et al.*, 2001).

Os microrganismos eucarióticos obtêm energia da glicose na forma de ATP através do metabolismo aeróbico (respiração) ou anaeróbico (fermentação). Nos dois casos, a glicose é convertida em duas moléculas de piruvato. O destino do piruvato depende do microrganismo e do estado metabólico da célula. Na respiração, o piruvato é completamente oxidado a CO₂ e água. Já na fermentação, que não envolve consumo de O₂, o piruvato é reduzido, com produção de NAD⁺ necessário para o funcionamento da via glicolítica. O produto da redução de piruvato no fungo *Rhizopus oryzae* e nas células de mamíferos é lactato, e em *Saccharomyces cerevisiae* é etanol e CO₂ (Fig. 2).

O metabolismo anaeróbico da glicose é, provavelmente, o mais remoto mecanismo biológico para a obtenção de energia a partir de moléculas orgânicas combustíveis, já que os organismos vivos surgiram, primeiramente, em uma atmosfera destituída de oxigênio. Com a evolução das cianobactérias, que produzem O₂ a partir da água, a atmosfera da Terra tornou possível a aerobiose e houve uma pressão seletiva sobre os organismos para o desenvolvimento do metabolismo aeróbico.

A taxa de produção de ATP (mol de ATP/unidade de tempo) e o rendimento (mol de ATP/mol de substrato) são diferentes nos metabolismos aeróbico e anaeróbico. A respiração apresenta menor taxa e maior rendimento (38 mol de ATP/mol de glicose), no entanto a fermentação apresenta maior taxa e menor rendimento (2 mol de ATP/mol de glicose).

Em um meio com reduzida disponibilidade de nutrientes, a taxa de produção de ATP, e não o rendimento, é o fator mais importante na competição entre as células pelos nutrientes disponíveis. Nestas condições, portanto, a fermentação torna-se mais vantajosa que a respiração.

Figura 1. Representação esquemática de glicólise, TCA, gliconeogênese e via alternativa do metabolismo de piruvato .

Figura 2. Representação esquemática da redução de piruvato a Etanol (I) ou Lactato (II) e a regeneração do NAD⁺ em organismos eucariotos.

O alto rendimento na produção de ATP pela respiração, no entanto, pôde ser aproveitado através da cooperação entre as células que, através da formação de agregados celulares, teriam favorecido o surgimento dos organismos multicelulares. Supõe-se, com base nisto, que a respiração esteve implicada na transição evolutiva de organismos unicelulares para organismos multicelulares (Pfeiffer *et al.*, 2001).

Um microrganismo unicelular como a levedura *S. cerevisiae* utiliza tanto a respiração quanto a fermentação (Ferea *et al.*, 1999), dependendo do estado metabólico da célula. Microrganismos multicelulares, no entanto, como os fungos filamentosos, utilizam preferencialmente a respiração (Kelly *et al.*, 1990). Apoiando esta colocação, o fungo dimórfico *Mucor racemosus* pode crescer na forma unicelular (similar à levedura) ou multicelular (micélio). A forma unicelular fermenta, enquanto a forma multicelular obtém energia através da respiração (Rogers *et al.*, 1974; Schulz *et al.*, 1974; Inderlied *et al.*, 1978).

Metabolismo Aeróbico e Anaeróbico de Glicose

S. cerevisiae é um fungo que se reproduz de forma sexuada, por meiosporos (ascósporos) e assexuada, por brotamento, processo no qual se desenvolve na superfície da célula adulta (célula mãe) uma pequena saliência (célula filha) que se transformará em uma nova célula (Alexopoulos *et al.*, 1996). Durante as duas últimas décadas, *S. cerevisiae* foi o mais importante organismo modelo para a pesquisa na área de genética molecular, fisiologia e bioquímica, devido principalmente ao fato de suas mecânicas celulares básicas de replicação, recombinação, divisão celular e metabolismo serem similares a dos eucariotos superiores (Seoighe *et al.*, 1999). Esta levedura é empregada na produção de álcool devido a sua capacidade de fermentar açúcares de arroz, trigo, cevada, cana de açúcar e outros substratos.

S. cerevisiae é capaz de satisfazer suas exigências de energia com o ATP gerado através da fermentação. Isto é claramente demonstrado através de mutantes “petite” com deficiência na função respiratória, incapazes de crescer em substratos não fermentáveis, e cujo metabolismo necessita do ciclo de Krebs e da respiração em estado funcional - ou seja, necessitam da mitocôndria - tais como etanol e glicerol (Tzagoloff, 1982). Células de levedura crescem rapidamente em um meio rico em glicose, por fermentação, com produção de etanol. Após o esgotamento da glicose, a respiração das células de levedura é restabelecida, e passam a utilizar o etanol como fonte de carbono e energia (Fig. 3). A passagem do metabolismo anaeróbico da glicose para o metabolismo aeróbico do etanol é

Figura 3. Perfil geral da passagem do metabolismo anaeróbico (fermentação) de glicose a metabolismo aeróbico (respiração) em *S. cerevisiae*.

denominada “diauxic shift”, que é o crescimento em duas fases e que acontece quando mais de uma substância é degradada como fonte de carbono e energia. A célula degrada primeiro uma substância e quando esta é esgotada do meio, degrada a segunda (DeRisi *et al.*, 1997).

A transição no modo de crescimento observada no “diauxic shift” exige que as células de levedura “sintam” a depleção de glicose, transmitam esta informação ao núcleo e à mitocôndria e respondam ao esgotamento da glicose, ativando os genes envolvidos na respiração. Assim, a célula deve equilibrar a expressão dos genes nucleares com a expressão dos genes mitocondriais e também com as exigências de crescimento.

O crescimento da levedura, em duas fases, envolve uma mudança do metabolismo anaeróbico (fermentação) para aeróbico (respiração) e é devida à expressão diferencial de genes. De fato um dos fenômenos regulatórios mais observados em microrganismos é a chamada repressão catabólica ou simplesmente repressão por glicose (Ronne, 1995).

S. cerevisiae possui um sofisticado mecanismo regulador que, em função da concentração de glicose no meio, estimula a fermentação ou a respiração (Johnston, 1999). Na presença de alta concentração de glicose (> 1 mM) (Flikweert *et al.*, 1997), *S. cerevisiae* fermenta glicose a etanol e CO₂, na presença ou ausência de oxigênio (Lagunas, 1979).

Glicose é um dos mais importantes efetores do metabolismo, podendo agir em nível de expressão gênica (repressão por glicose), estabilidade do mRNA, tradução do mRNA e estabilidade de proteínas (Carlson, 1999; Johnson, 1999). Durante o crescimento, em meio com glicose, é reprimida a transcrição de genes que codificam enzimas envolvidas no metabolismo de fontes alternativas de carbono (como galactose, sacarose e maltose), na via respiratória (como enzimas do ciclo de TCA, biossíntese de heme, transporte de elétrons e fosforilação oxidativa), sendo induzidos genes requeridos para a utilização de glicose, como os das enzimas glicolíticas e transportadores de glicose. Quando a glicose se esgota no meio, as células começam a consumir o etanol gerado durante a fermentação e a expressão dos genes correspondentes é ativada (Gancedo, 1998; Johnston, 1999).

Os processos envolvendo a repressão e a indução de conjuntos diferentes de genes estimulou a pesquisa sobre a regulação metabólica em *S. cerevisiae* e outros microrganismos eucarióticos (DeRisi *et al.*, 1997; Flores *et al.*, 2000; Zeeman *et al.*, 2000; Kuhn *et al.*, 2001).

O genoma de *S. cerevisiae* foi seqüenciado sendo mapeadas 6400 ORFs potenciais (Goffeau *et al.*, 1996). Utilizando-se estas ORFs foi realizada uma análise do programa

temporal de expressão gênica durante a passagem metabólica da fermentação a respiração, empregando-se a metodologia de DNA “microarray” (DeRisi *et al.*, 1997). Variações nos sinais de hibridação resultantes permitiram identificar e caracterizar a expressão dos genes, comparando-se o padrão de expressão da amostra inicial com o padrão de expressão de amostras obtidas antes e depois da mudança metabólica.

Muitos genes cuja função ainda era desconhecida apresentaram mudanças no padrão de expressão por um fator de 2 ou mais vezes depois do esgotamento da glicose. Os dados sobre a magnitude das mudanças de expressão ajudaram a determinar a função daqueles genes. Muitos genes que codificam enzimas do metabolismo de glicose mostraram alterações significativas no padrão de expressão. Estas mudanças são mostradas na Figura 4, onde as caixas vermelhas identificam genes cuja expressão é aumentada, e as caixas verdes representam genes cuja expressão é diminuída. Podemos observar que os genes que direcionam o fluxo de glicose a piruvato e de piruvato a etanol são mais expressos durante o crescimento anaeróbico na presença de glicose. Quando a glicose se esgota no meio, são expressos genes que mediam o crescimento aeróbico da levedura, sendo mais expressos os genes do metabolismo de etanol e do ciclo de TCA (Fig. 4).

A análise do perfil de expressão de genes com função metabólica conhecida mostra a reprogramação metabólica do padrão de expressão dos genes durante a transição de um estado de crescimento anaeróbico a aeróbico. A Figura 4 mostra o efeito da glicose sobre a expressão gênica, complementando o que havia sido determinado por estudos bioquímicos clássicos com aquelas enzimas. (DeRisi *et al.*, 1997).

Várias perguntas básicas, porém, permanecem ainda sem resposta. Por que os microrganismos eucarióticos obtêm energia da glicose preferencialmente através da respiração? Por que o “diauxic shift” não é observado em outros microrganismos eucarióticos? Que fator(es) determina(m) esta diferença e em que nível molecular a seleção teria ocorrido? As respostas a estas perguntas críticas são fundamentais não apenas para nosso conhecimento básico sobre a regulação metabólica da glicose e seus aspectos evolutivos, mas, também, para o uso potencial de microrganismos eucarióticos na engenharia metabólica e na produção de compostos úteis.

A investigação de tais questões, em microrganismos multicelulares em nível transcricional, depende da identificação e isolamento dos genes da via glicolítica e das enzimas responsáveis pelo destino do piruvato. Para o desenvolvimento desse estudo, estabelecemos, neste trabalho, um banco de dados contendo amostra significativa de

seqüências parciais de genes (“Expressed Sequence Tags” – EST) do fungo filamentoso *Trichoderma reesei*.

Utilizando estas seqüências de cDNA e a metodologia de “microarray” analisamos o perfil de expressão gênica de *T. reesei* durante o consumo de glicose. Fizemos também uma análise comparativa entre os dados obtidos com *T. reesei* e o programa temporal de expressão gênica de *S. cerevisiae*, durante a passagem metabólica de fermentação para respiração, estabelecida por DeRisi *et al.* (1997).

O fungo *T. reesei* foi escolhido para este estudo porque seu habitat natural e exigências nutricionais são muito diferentes dos de *S. cerevisiae*. Enquanto concentrações relativamente altas de açúcares prevalecem no habitat natural de *S. cerevisiae*, o habitat natural de *T. reesei* é o solo, um ambiente pobre em nutrientes (Eveleigh, 1985).

Trichoderma

O gênero *Trichoderma* compreende espécies de fungos filamentosos, mesofílicos (30 °C), que normalmente habitam o solo. Juntamente com *Neurospora*, *Penicillium* e *Aspergillus*, o gênero *Trichoderma* é um dos mais estudados entre os fungos filamentosos devido ao seu grande potencial de aplicação do ponto de vista industrial e biotecnológico (Samuels, 1996).

Nos anos 1800, o gênero *Trichoderma* foi introduzido como um novo grupo taxonômico para incluir quatro espécies de fungos, entre as quais *T. viride*. Durante a segunda guerra mundial, um programa de pesquisa do exército norte-americano, preocupado com a rápida deterioração do material que compunha as barracas de campanha do exército, em regiões tropicais, identificou o fungo *Trichoderma* “*viride*” QM6a, que produz grandes quantidades de enzimas celulolíticas extracelulares, como o responsável pelo problema. Posteriores pesquisas determinaram que se tratava de uma nova espécie de fungo do gênero *Trichoderma*, a qual foi denominada *T. reesei*, em memória a Elwyn T. Reese, um dos principais pesquisadores naquele programa. Reese, em colaboração com Mary Mandels, realizou pesquisas sobre a biossíntese, estrutura e mecanismos de degradação da celulose e outros polissacarídeos por este fungo (Reese *et al.*, 1980; Gams *et al.*, 1998).

T. reesei apresenta hifas septadas, ramificadas (de 5 a 10 µm de diâmetro), polinucleadas, de núcleo haplóide, que formam colônias brancas de rápido crescimento, formando almofadas verdes ou amarelas de filamentos de esporulação. Os filamentos férteis, conidióforos, produzem fileiras laterais de pequenas fiálides. Os mitoesporos

(conídios, geralmente de cor verde e de 3 a 5 µm de diâmetro) são produzidos sucessivamente no extremo da fiálide e agrupados em pequenas massas (Gams *et al.*, 1998). A cariotipagem, em uma linhagem selvagem QM6a e em cepas mutantes hipercelulolíticas, como QM9414, através da eletroforese de campo pulsado revela sete bandas, correspondentes a sete cromossomos. As sete bandas têm de 2,8 a 6,9 Mb. Estima-se que todo o genoma de *T. reesei*, constituído por sete cromossomos, tenha de 31 a 39 Mb (Mantyla *et al.*, 1992; Herrera-Estrella *et al.*, 1993).

No ciclo de vida dos fungos, a forma sexuada é conhecida como teleomórfica (produz meiosporos) e a forma assexuada, como anamórfica (produz mitosporos). Como estas formas têm características morfológicas diferentes, o teleomórfo e o anamórfo, de uma mesma espécie, terão nomes diferentes (Alexopoulos *et al.*, 1996). Algumas espécies do gênero *Hypocrea* (Família Hypocreales), apresentam teleomórfos que têm sido relacionados com anamorfos de *Trichoderma*; Kuhls *et al.*, (1996) determinaram que *T. reesei* e *Hypocrea jecorina*, apresentam seqüências idênticas de nucleotídeos na região do espaçador interno de transcrição do DNA ribossomal e similar padrão de fragmentos de DNA obtidos por PCR, o qual é característico de culturas da mesma espécie, estabelecendo que *T. reesei* é um anamorfo do teleomórfo *Hypocrea jecorina* e que, provavelmente, ambos fungos representariam a mesma espécie.

Por isso a classificação taxonômica de *T. reesei* (Taxonomy Browser, NCBI, USA), é:

Taxonomy ID: 51453

Super-reino: Eukaryota;

Reino Fungi;

Filo: Ascomycota

Ordem: Hypocreales

Família: Hypocreaceae

Gênero: *Hypocrea*

Espécie *Hypocrea jecorina*

Outro nome: *Trichoderma reesei* [anamorfo]

Neste trabalho, utilizaremos o nome mais comumente empregado, *Trichoderma reesei*, abreviado *T. reesei*.

O gênero *Trichoderma* é caracteristicamente reconhecido pela produção de diversos sistemas extracelulares de enzimas envolvidas na hidrólise de polissacarídeos (Béguin, 1990). Em particular, *T. reesei* se destaca por sua grande capacidade de produzir e secretar

um sistema multienzimático, induzível, de enzimas celulolíticas (Eveleigh, 1985). Este sistema extracelular consiste, em geral, de duas principais classes de enzimas: celobiohidrolases (CBH) responsáveis pela remoção de resíduos de celobiose consecutivamente a partir das extremidades da cadeia celulósica e endoglucanases (EG) que clivam, ao acaso, as ligações β -glicosídicas. As duas classes de enzimas agem de forma sinérgica, sendo necessárias para a hidrólise eficiente de celulose a oligossacarídeos solúveis (Carle-Urioste *et al.*, 1997).

Há inúmeros trabalhos a respeito do mecanismo de expressão dos genes das celulasas de *T. reesei*, alguns desenvolvidos em nosso laboratório (El-Gogary *et al.*, 1989; Abrahão-Neto *et al.*, 1995; El-Dorry *et al.*, 1996; Torigoi *et al.*, 1996; Carle-Urioste *et al.*, 1997; Carraro *et al.*, 1998).

Outras espécies do gênero *Trichoderma* têm sido empregadas como agentes de controle biológico. Estas espécies, entre as quais destaca-se o fungo *T. harzianum*, produzem quitinases, que degradam a parede celular de fungos fitopatogênicos, como *Pythium*, *Rhizoctonia*, *Alternaria*, *Phytophthora* (Chet, 1987; Samuels, 1996).

Avanços significativos sobre a biologia do gênero *Trichoderma* dependem de informações sobre a estrutura gênica. A partir destas informações podem ser realizados estudos nas áreas de genética molecular, bioquímica, biotecnologia e fisiologia celular. Estes estudos propiciariam a identificação de vias metabólicas básicas; de genes envolvidos na síntese de enzimas que degradam polissacarídeos e material xenobiótico (DDT, malathion, etc.); de genes de interesse para obtenção de plantas resistentes a fitopatógenos; e à produção de metabólitos secundários, e de substâncias para biocontrole (Samuels, 1996).

Até o momento, apenas 109 genes de *T. reesei* estão depositados na base pública de dados (GenBank, NCBI, USA), (<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wgetorg>) (até 06/02/2002). Um de nossos objetivos é construir uma base de dados de EST de *T. reesei*, incrementando significativamente o número de genes identificados.

Análise estrutural e funcional do Genoma

O dogma central da biologia molecular descreve como genes as seqüências de DNA que são transcritas a RNA mensageiro (mRNA), o qual é traduzido numa proteína funcional. Em 1953 Watson e Crick descreveram o arranjo linear dos genes e como acontece o processo de replicação do DNA. Em 1966, os grupos de Khorana e Nirenberg, independentemente, determinaram o código genético, permitindo a predição de seqüências

de proteínas a partir de seqüências de DNA. Maxam e Gilbert (1977) e Sanger *et al.* (1977) introduziram, respectivamente, técnicas de seqüenciamento químico e enzimático de DNA, conduzindo ao seqüenciamento do genoma de 40 Kb do bacteriófago Lambda por Sanger *et al.* (1982). Desde então, técnicas de seqüenciamento de DNA baseadas no método enzimático de terminação da cadeia, de Sanger, têm sido desenvolvidas e automatizadas para o seqüenciamento em larga escala (Sterky *et al.*, 2000).

O interesse pelo mapeamento e seqüenciamento do genoma de vários organismos, em especial o do homem, motivou o desenvolvimento e a implementação de bases de dados de nucleotídeos e proteínas, bem como de métodos de bioinformática para administrar e interpretar o grande volume de informação.

O projeto genoma humano (PGH), iniciado em 1990, teve como objetivos: a) determinar a seqüência completa de bases nitrogenadas que compõem o genoma (aproximadamente três bilhões); b) identificar e mapear todos os genes humanos; e c) armazenar a informação obtida em bancos de dados públicos. O PGH motivou o desenvolvimento de projetos genoma de outros organismos modelo, principalmente microrganismos, visando o desenvolvimento de tecnologia. Como existe uma ordem subjacente a toda a diversidade da vida e como todos os organismos se relacionam através de semelhanças em suas seqüências de DNA, o conhecimento adquirido a partir de genomas não-humanos freqüentemente leva a novas descobertas na biologia humana.

Em 1995, o genoma completo de dois microrganismos foi, pela primeira vez, conhecido. Fleischmann *et al.*, (1995) e Fraser *et al.*, (1995) publicaram, respectivamente, toda a seqüência de bases de *Haemophilus influenzae* (1,83 Mb) e *Mycoplasma genitalium* (0,58 Mb). No ano seguinte, Goffeau *et al.*, (1996) publicou a seqüência completa de *S. cerevisiae* (12 Mb), o primeiro eucarioto a ter seu genoma completamente seqüenciado. Desde então, 73 genomas foram seqüenciados e publicados, incluindo o genoma do primeiro patógeno de plantas, *Xylella fastidiosa* (Simpson *et al.*, 2000), seqüenciado pelo consórcio brasileiro ONSA (Organization for Nucleotide Sequencing and Analysis, SP. Brazil) (Bonalume, 1997).

Atualmente, 174 genomas de eucariotos, entre eles os de *Aspergillus nidulans*, *A. parasiticus*, *Neurospora crassa*, *Fusarium sporotrichoides*, *Magnaporthe grisea* (Yoder *et al.*, 2001) e 260 genomas de procarióticos encontram-se em andamento (Genomes OnLine Database, 06/02/2002) (<http://igweb.integratedgenomics.com/GOLD/>).

Em fevereiro de 2001, foi anunciado o esboço do genoma humano pelo PGH e pela empresa norte-americana Celera Genomics. O número de genes humanos, segundo os

cálculos de ambas as equipes de pesquisadores, situa-se em torno de 30 mil (Lander *et al.*, 2001; Venter *et al.*, 2001), confirmando a ordem de grandeza do número de genes humanos estimado por análise de ESTs (Ewing *et al.*, 2000).

A determinação da sequência nucleotídica, no entanto, consiste apenas na primeira etapa no estudo do genoma. Diversos métodos estão sendo empregados para determinar a expressão gênica, em distintas condições e tipos celulares (Staudt *et al.*, 2000).

A Genômica, ciência que estuda como os genes ou a informação genética estão organizados no genoma e de que forma esta organização determina sua função, permite a análise sistemática e simultânea de um grande número de genes e o estudo dos processos biológicos como um todo. A Genômica é constituída pela Genômica Estrutural, que determina a sequência completa do genoma de organismos, e pela Genômica Funcional, que emprega a informação da genômica estrutural para avaliar os processos biológicos de uma forma sistemática (Staudt *et al.*, 2000).

Genômica estrutural: “Expressed Sequence Tags” (ESTs)

Os seres vivos apresentam milhares de genes, dos quais apenas uma fração é expressa em cada célula individual. Então, é o programa temporal e espacial de expressão dos genes que determina os processos da vida. O desafio atual é identificar e caracterizar os genes expressos em determinadas condições fisiológicas e metabólicas nas células.

O desenvolvimento de vários “projetos genoma”, particularmente o PGH e o de levedura proporcionaram o avanço de muitas técnicas de seqüenciamento e a rápida identificação de genes. Diante do imenso tamanho do genoma de certos organismos, foi adotada uma estratégia que prioriza o seqüenciamento apenas da fração expressa do genoma. Esta estratégia é baseada no seqüenciamento parcial (200-500 bases) a partir das extremidades 3’ ou 5’ de clones tomados aleatoriamente de uma biblioteca de cDNA, constituída basicamente de coleções de genes expressos inseridos em vetores de DNA. As bibliotecas de cDNA são construídas a partir de mRNA, empregando-se como iniciadores oligonucleotídeos sintéticos poli(dT) (Sambrook *et al.*, 1989). Outra metodologia, denominada ORESTES (“Open Reading Frame Expressed Sequences Tags”) (Dias Neto *et al.*, 2000), utiliza oligonucleotídeos sintéticos aleatórios (Fig. 5). A sequência de nucleotídeo obtida é denominada “etiqueta de sequência expressa” ou EST (“Expressed Sequence Tags”) (Adams *et al.*, 1991).

Figura 5. Métodos de construção de bibliotecas de cDNA para seqüenciamento e geração de dados de EST.

Apesar da tecnologia para o completo seqüenciamento de genomas estar disponível, os custos e o tempo necessário para o desenvolvimento destes projetos são grandes, devido à complexidade e ao tamanho dos genomas. A metodologia de EST se apresenta como uma alternativa atraente, constituindo-se numa ferramenta extremamente útil na identificação de genes e em estudos de expressão gênica. Entre as desvantagens, podemos citar a necessidade de obtenção de mRNA de todos os tecidos, tipos celulares e etapas de desenvolvimento, além de não proporcionar informações sobre *íntrons* e regiões intergênicas (reguladoras ou de controle) (Adams *et al.*, 1991). A metodologia de EST tem sido utilizada nos projetos genoma de fungos (Yoder *et al.*, 2001) e também foi empregada no estudo do genoma de vários outros organismos eucarióticos, incluindo o humano (Adams *et al.*, 1992; Adams *et al.*, 1993; Hillier *et al.*, 1996; Nelson *et al.*, 1998; de Souza *et al.*, 2000) e de plantas (Sterky *et al.*, 1998; Crookshanks *et al.*, 2001). Importantes projetos de seqüenciamento, empregando esta metodologia, foram conduzidos com grande sucesso no Brasil, através do consórcio ONSA, contribuindo com mais de 700 mil ESTs de humanos, obtidas a partir de diversos tecidos tumorais e normais (Camargo *et al.*, 2001) e 300 mil ESTs de diversos tecidos de cana de açúcar (<http://sucest.lad.dcc.unicamp.br/cgi-bin/public/prod/BD/current/current.pl>). Atualmente, encontra-se em andamento o projeto de EST do genoma de *Schistosoma mansoni* (<http://verjo18.iq.usp.br/schisto/> e www.cpqrr.fiocruz.br/genoma/).

Posto que as seqüências de proteínas são mais conservadas que as seqüências de nucleotídeos (Adams *et al.*, 1991), a seqüência nucleotídica da EST deve ser convertida à seqüência de proteína, para se fazer à comparação contra uma base de dados de proteínas. Os projetos de seqüenciamento de EST de cérebro humano (Adams *et al.*, 1991; Adams *et al.*, 1995), tecidos normais e de tumores humanos (Camargo *et al.*, 2001), *Caenorhabditis elegans* (Waterston *et al.*, 1992; McCombie *et al.*, 1992) e outros têm empregado a análise de similaridade das ESTs contra a base de dados de proteínas como um método para a identificação de genes.

A obtenção de ESTs permite a comparação de seqüências entre espécies diferentes, baseando-se na observação de que genes responsáveis pela mesma função, em organismos diferentes, são conservados durante a evolução. O emprego desta metodologia permite a identificação de genes, descobrimento de novos genes, reconhecimento de *exons*, caracterização dos limites *exons/íntrons*, formação de famílias de proteínas, desenvolvimento de mapas gênicos, identificação de genes específicos de tecidos ou organismos, investigação de genes de função desconhecida e estabelecimento de padrões

de expressão gênica espacial e temporal (Adams *et al.*, 1995, Boguski *et al.*, 1995, Banfi *et al.*, 1996; Hillier *et al.*, 1996; Lluisma *et al.*, 1997). O rápido crescimento do número de seqüências de EST, obtidas de diferentes organismos, motivou o desenvolvimento de uma base pública de dados de ESTs no GeneBank (NCBI, USA) (Boguski *et al.*, 1993).

A base pública de dados dbEST (NCBI, USA) contém 10.299.149 ESTs, obtidas de mais de 179 espécies de organismos (dbEST, liberação em 01/02/2002), incluindo o homem (>3.900.000), camundongo (>2.500.000), rato (>326.000), *C. elegans* (>191.000), *Emericella nidulans* (12.993), *Neurospora crassa* (28.089), *S. cerevisiae* (3.041), *Magnaporthe grisea* (1.507), *Aspergillus niger* (2.533), *Candida albicans* (19), *Amanita muscaria* (33), *Schizosaccharomyces pombe* (8.123), entre outros.

O número de ESTs é maior que o número possível de transcritos num organismo, devido ao fato de que um determinado gene pode estar representado mais de uma vez por distintas ESTs. O projeto UniGene (NCBI, USA) (Schuler *et al.*, 1996) realiza uma classificação sistemática de ESTs, empregando métodos de montagem que permitem agrupar as seqüências de cDNA que se sobrepõem, formando contíguos (“contígs”) que provisoriamente correspondem a um único gene (Boguski *et al.*, 1995).

Uma base de dados de EST se constitui, então, em um esboço dos genes expressos de um organismo, a partir do qual podemos realizar a análise do genoma funcional (Gerhold *et al.*, 1996).

A metodologia de EST é adequada para a identificação de genes expressos em *T. reesei*, considerando-se que seu genoma ainda não foi seqüenciado e por ser necessário obter uma grande quantidade de seqüências de genes para realizar um estudo global de expressão gênica.

Genômica funcional: Análise global de expressão gênica

Apesar do processamento pós-transcricional ser importante na regulação da expressão de muitos genes, a taxa de transcrição constitui-se, *per se*, no principal mecanismo de regulação (Carulli *et al.*, 1998).

Metodologias para a análise da expressão diferencial de genes são empregadas para identificar genes envolvidos no processo de desenvolvimento e de resposta celular a uma variedade de estímulos químicos ou físicos, pesquisar o efeito molecular originado por mutações em um gene de interesse, identificar marcadores moleculares de doenças e estudar o efeito do tratamento com algumas drogas (Carulli *et al.*, 1998).

Várias técnicas de análise de expressão gênica global foram descritas (Carulli *et al.*, 1998; Kozian *et al.*, 1999). Algumas estão resumidas a seguir:

a) Análise comparativa de ESTs (Adams *et al.*, 1995): Considerando-se que, durante a preparação de bibliotecas de cDNA, a transcrição reversa de toda a população de mRNA é feita de forma uniforme, a abundância relativa das seqüências de ESTs obtidas representa a expressão gênica nas condições em estudo. O seqüenciamento aleatório de uma biblioteca de cDNA aliada a métodos de bioinformática nos permite, então, comparar a expressão gênica entre diferentes tecidos ou entre microrganismos em diferentes condições fisiológicas, identificando os genes mais e menos expressos (Carulli *et al.*, 1998).

b) Análise seriada da expressão gênica (SAGE, “Serial analysis of gene expression”, Velculescu *et al.*, 1995): SAGE é uma extensão da análise de EST. Pequenas etiquetas (9-11 nucleotídeos) de cDNA são obtidas da extremidade 3' de cada molécula de mRNA. Foi estabelecida que esse número de nucleotídeos contém suficiente informação para a identificação do transcrito (Velculescu *et al.*, 1995). Estas pequenas etiquetas são ligadas entre si e, após o fracionamento dos concatâmeros, a fração com tamanho entre 800 – 1200 pb (80 – 120 etiquetas) é inserida em um vetor adequado para seqüenciamento. Com o uso de programas computacionais, a abundância de cada etiqueta é determinada. Posteriormente, faz-se a determinação da identidade das etiquetas contra uma base de dados. O padrão de expressão para genes diferentes é refletido pela abundância relativa das etiquetas individuais. Foram estudados padrões de SAGE em humanos (Velculescu *et al.*, 1999) e em levedura (Velculescu *et al.*, 1997). Uma condição prévia para a identificação das etiquetas é a disponibilidade de um banco de dados do genoma da espécie em estudo.

c) DNA “microarray” (Schena *et al.*, 1995) (Fig. 6): As seqüências completas (genes) ou parciais (ESTs) de DNA (0,5 – 1,5 kb) são imobilizadas em lâminas de vidro. O cDNA obtido a partir do mRNA de células crescidas em diferentes condições é marcado com moléculas repórteres fluorescentes diferentes e empregado na hibridação comparativa do DNA fixado na lâmina. A intensidade de fluorescência obtida a partir das moléculas repórteres ligadas ao DNA na lâmina é empregada para determinar o nível de expressão dos genes na condição em estudo.

A metodologia de DNA “microarray” encontra-se descrita em inúmeros trabalhos entre os quais podemos mencionar Kuklin (1999); Watson *et al.*, (1998); Celis *et al.*, (2000); van Hal *et al.*, (2000); Duggan *et al.*, (1999); Brown *et al.*, (1999), e o suplemento de Nature Genetics, “The Chipping Forecast” (Janeiro, 1999).

Figura 6. Metodologia de análise de expressão gênica por cDNA microarray. O nível de expressão gênica relativa entre uma condição A e B é comparada. A intensidade de fluorescência de cada uma dos cDNA (alvos) marcados com Cy3 e Cy5 é separadamente quantificada. A razão de intensidade de fluorescência Cy5/Cy3 é uma medida da abundância relativa de cada transcrito na amostra de mRNA. O nível de expressão pode ser mostrado empregando uma escala de cor.

No estudo da expressão gênica global, pela técnica de “microarray”, os genes de um dado organismo ou parte dele são primeiramente amplificados por PCR e depositados por um sistema robótico computadorizado em lâminas de vidro de 2 x 4 cm. Cada gene representado na lâmina é denominado de elemento e se constitui em uma sonda (“probe”) de um experimento de “microarray”. O diâmetro dos elementos é de 75 - 200 μm com uma distância entre eles de, aproximadamente, 150 μm . É possível fixar, desse modo, mais de 1000 elementos de DNA por cm^2 . O conjunto de genes nas lâminas são hibridados simultaneamente com cDNA sintetizados a partir de mRNA isolado de células crescidas em duas condições metabólicas ou fisiológicas distintas.

Para fins ilustrativos, imaginemos que a técnica de “microarray”, esquematizada na Figura 6, esteja sendo utilizada para o estudo do controle da expressão gênica por glicose em um microrganismo eucariótico. Neste caso, a condição A representaria a população de mRNAs isolados em alta concentração de glicose, enquanto a condição B representaria a população de mRNAs isolados na ausência de glicose. Na preparação do cDNA, o mRNA da condição A é transcrito reversamente empregando-se “primers” de oligo-dT, dNTPs, e um dNTP adicional que será marcado com o corante fluorescente Cy3 (verde). O mRNA da condição B é tratado de forma semelhante, utilizando-se o dNTP adicional que será marcado com o corante fluorescente Cy5 (vermelho). Misturam-se então os dois cDNAs marcados, A+B, para uma hibridação simultânea dos elementos de DNA fixados na lâmina de vidro (Duggan *et al.*, 1999).

O resultado é analisado empregando-se um “scanner” confocal de 2 laserres, que excitam os corantes incorporados aos cDNAs aderidos às sondas de DNA. Um conjunto de filtros seleciona a emissão dos corantes (Cy3 e Cy5), enviando os dados a um computador que gera 2 imagens correspondentes à cada fluorescência. A análise da intensidade de fluorescência e a relação entre as fluorescências (Cy5/Cy3) dos elementos na lâmina permitem determinar se um gene específico está sendo expresso e o nível relativo de expressão dos genes nas duas condições de estudo (Khan *et al.*, 1999). Na Figura 6, o gene 1 mostrou-se mais expresso em baixa concentração de glicose (condição B), em contraposição ao gene 3, que teve uma maior expressão em alta concentração de glicose (condição A). A expressão do gene 2 permaneceu constante em ambas as condições.

Vários estudos de expressão gênica, empregando DNA “microarray”, foram desenvolvidos em células humanas (Schena *et al.*, 1996; Heller *et al.*, 1997; Loftus *et al.*, 1999; Yang *et al.*, 1999), plantas (Schena *et al.*, 1995); *S. cerevisiae* (DeRisi *et al.*, 1997; Lashkari *et al.*, 1997; Chu *et al.*, 1998); camundongos (Tanaka *et al.*, 2000), entre outros.

DNA “microarray” contendo genes ou ESTs de diferentes organismos pode ser obtido comercialmente através de empresas como Genomic Solutions Inc. (USA), ResGen (USA), PharmArray (USA), entre outras. A empresa Paradigm Genetics, Inc. (USA) (www.paragen.com) disponibiliza o DNA “microarray” do fungo patogênico de arroz *Magnaporthe grisea*.

Apesar da sensibilidade da técnica de DNA “microarray” – é possível detectar espécies de mRNA presentes em até 1:10 000 da massa de poli(A⁺) (Duggan *et al.*, 1999). Vários métodos podem ser empregados para confirmar os resultados obtidos. Entre as técnicas mais utilizadas para determinar a expressão relativa de um gene podemos mencionar: a) RT-PCR quantitativo e b) Análise por “Northern blot”, (Reue, 1998).

d) “Genechip DNA” (Lockhart *et al.*, 1996): O princípio básico do “Genechip DNA” é semelhante ao do DNA “microarray”. No “Genechip”, entretanto, os genes são representados por oligonucleotídeos (15-25 pb) fixados em lâminas de vidro. A técnica permite a preparação de lâminas de alta densidade contendo > 250 000 oligonucleotídeos /cm². A hibridação e análise dos dados são similares aos procedimentos utilizados no DNA “microarray”.

“Genechip DNA” contendo todos os genes de *S. cerevisiae* têm sido empregados em análises de expressão gênica global. A tecnologia encontra-se patenteada pela empresa Affymetrix Inc. (CA, USA), que disponibiliza comercialmente o GeneChip[®] de humano (ESTs), camundongo (ESTs), rato (ESTs), *S. cerevisiae* (todos os ORFs), *Caenorhabditis elegans* (todos os ORFs), *E. coli* (todos os ORFs), *Arabidopsis thaliana* (ESTs), *Drosophila melanogaster* (ESTs), entre outros. GeneChip[®] pode ser também empregado em análise de genotipagem e polimorfismo (Lipshutz *et al.*, 1999; Lockhart *et al.*, 2000).

Entre as técnicas acima, “Genechip DNA” e “DNA microarray” têm sido as mais utilizadas para a análise da expressão gênica global. No presente trabalho foi empregada a metodologia de “DNA microarray”.

De acordo com o que foi descrito acima, podemos estabelecer que o conhecimento e as ferramentas de pesquisa encontram-se desenvolvidas de tal forma que compete a nós planejar os modelos experimentais que possibilitem o uso adequado daquelas metodologias e a geração de novo conhecimento.

OBJETIVO

Apesar de haver grande interesse no estudo da regulação metabólica e evolução das vias produtoras de ATP, o porquê da maioria dos microrganismos multicelulares metabolizarem glicose preferencialmente através da respiração, enquanto que muitos organismos unicelulares se utilizam tanto da respiração quanto da fermentação, ainda permanece sem resposta.

O entendimento desta diferença entre organismos multicelulares e unicelulares, através da investigação do perfil de expressão gênica do fungo filamentoso *Trichoderma reesei* durante o consumo de glicose e a condução de análise comparativa com o programa temporal de expressão gênica observado em *S. cerevisiae* na transição da fermentação para respiração foi o principal objetivo deste trabalho

Para se alcançar este objetivo foi construída uma base de dados de EST de *T. reesei*, visando a obtenção do maior número de genes expressos que permitissem realizar uma análise global da expressão gênica, empregando-se a metodologia de “cDNA microarray”.

O avanço das técnicas de seqüenciamento, a disponibilidade de bases de dados e de programas computacionais de análise, comparação e identificação, bem como das técnicas de análise global de expressão gênica, geraram as ferramentas que possibilitam o estudo total integrado de um sistema, em nosso caso, metabólico. Estas técnicas foram utilizadas para se conhecer as diferenças do destino metabólico da glicose entre aqueles dois microrganismos.

Não menos importante, a realização deste trabalho permitiu o estabelecimento daquelas técnicas em nosso laboratório.

MATERIAL E MÉTODOS

Reagentes e equipamentos

Meios de cultura para bactéria e PDA para cultura de fungo foi obtido de DIFCO, USA. Os reagentes utilizados foram de grau analítico (P.A.), da marca Merck S.A. Da Sigma Chemical Company, St. Louis, MO, USA, foram adquiridos SDS (Lauril-sulfato de sódio), persulfato de amônio, glioxal, Tween-80, Agarose. Kits para sequenciamento de DNA, foram obtidos da Applied Biosystem Inc., USA; Kits para marcação de DNA ("Random primers") da Gibco-BRL. Kits para amplificação de DNA foram adquiridos da Gibco-BRL (UK) e Bioline (USA). Isótopos e membranas Hybond-N⁺ foram adquiridos da Amersham Pharmacia, International, Inglaterra. Os filmes de Raios-X de marca KODAK, X-OMAT XK-1. Os "primers" foram sintetizados sob encomenda pela Gibco-BRL. Placas de 96 poços em fundo em U e em V marca COSTAR (USA) e placas de 96 poços MultiScreen-FB marca Millipore (USA).

Para o sequenciamento automático de DNA foi utilizado aparelho modelo ABI 377 e as reações de PCR foram feitas no termociclador GeneAmp PCR System 9700, ambos da Applied Biosystem Inc., USA.

Na centrifugação de placas foi utilizada uma centrífuga marca Qiagen-USA, modelo 4K15 com rotor 09100.

Manipulação de DNA e técnicas em geral

Transferência e hibridação ("Southern" e "Northern"); marcação com material radioativo; eletroforese de DNA e RNA; e outras técnicas empregadas em biologia molecular, foram feitas segundo descrito por Ausbel *et al.*, (1990) e Sambrook *et al.*, (1989).

Cepas e condições de cultura

Foram utilizadas as seguintes cepas:

Célula *E. coli* SOLRTM (Stratagene, USA): *e14*⁻ (*McrA*⁻) Δ (*mcrCB-hsdSMR-mrr*)171 *sbcC recB recJ uvrC umuC::Tn5* (*Kan*^r) *lac gyrA96 relA1 thi-1 endA1 λ R* [*F'* *proAB lacIqZ* Δ M15] *Su*- (*non suppressing*).

Célula *E. coli* XL1-Blue MRF' (Stratagene, USA): Δ (*mcrA*)183 Δ (*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [*F'* *proAB lacIq Z* Δ M15 *Tn10* (*Tet*^r)].

Cepa de *Trichoderma reesei* QM9414 (ATCC 26921), derivada de QM6a (Montenecourt *et al.*, 1979).

A cultura de bactéria foi feita em meio LB (Sambrook *et al.*, 1989), contendo 150 µg/mL ampicilina (para seleção de clones contendo pBluescript SK(+)). Para placas de cultura foi utilizado o meio LB-Ágar contendo o mesmo antibiótico.

Em nosso laboratório, *T. reesei* é mantido em placas de ágar-dextrose de batata (PDA) à temperatura ambiente. Antes do uso, os esporos são ressuspensos em NaCl a 0,8% e contados em câmara de Neubauer.

Culturas líquidas foram feitas em meio completo (uréia a 0,3%, sulfato de amônio a 0,14%, sulfato de magnésio a 0,03%, tween 80 a 0,2%, proteose/peptona a 0,2%, cloreto de cálcio a 0,03%) suplementado com metais a 0,1% (FeSO₄ 0,5%, MnSO₄ 0,16%, ZnSO₄ 0,16%, CoCl₂ 0,2%), em tampão fosfato de potássio 0,1 M pH 6,0 (El-Gogary *et al.*, 1989), mantido a 30 °C sob agitação constante (300 rpm). O micélio coletado foi filtrado e congelado imediatamente em nitrogênio líquido.

Extração de DNA e RNA

RNA para construção da biblioteca de cDNA e DNA foram obtidos a partir de uma cultura do fungo, crescido por 18 horas, contendo glicerol a 1% como fonte de carbono.

DNA foi obtido por maceração da massa micelial em nitrogênio líquido e extraído segundo descrito por El-Gogary *et al.*, (1989).

RNA para o “microarray” foi obtido a partir de uma cultura de 0.5 L (inoculada com 10⁷ esporos/mL), que foi empregada para inocular um fermentador (MagnaFerm Fermentator, New Brunswick Scientific, N.J., USA, com capacidade de 14 L), com 10 litros de meio de cultura e glicose a uma concentração de 100 mmol/L. A cultura foi mantida a 30 °C sob agitação a 300 rpm e aeração constante. Amostras da cultura foram retiradas a intervalos, seguindo o consumo de glicose até seu esgotamento. Foi obtida e conservada a massa micelial em nitrogênio líquido. Glicose, no filtrado, foi determinada enzimaticamente empregando o Kit Sera-Pak (Bayern, USA).

RNA total foi obtido por maceração da massa micelial em nitrogênio líquido e extração segundo o protocolo de isotiocianato de guanidina/CsCl (Chirgwin *et al.*, 1979). O RNA obtido foi dissolvido em água, tratado com RNase livre de DNA (segundo indicado pelo fabricante), quantificado e estocado a -70 °C. A qualidade do RNA obtido foi estimada, após desnaturação com glioxal/DMSO e pela visualização das amostras em gel desnaturante de agarose a 1%.

Construção da biblioteca de cDNA

A biblioteca foi construída a partir de 0,5 mg do RNA total. mRNA poly(A)⁺ foi obtido empregando o Poly (A)⁺ Quick[®] mRNA isolation kit, seguindo as instruções do fabricante (Stratagene, USA).

O sistema Uni-ZAP[®] XR vector (Stratagene, USA), foi empregado na construção da biblioteca de cDNA, seguindo as instruções do fabricante. Este sistema resulta na obtenção de clones contendo insertos de cDNA clonados em forma direcionada entre os sítios de restrição *EcoRI* e *XhoI* do vetor Lambda Uni-ZAP XR.

O fago de auxílio F' foi empregado na excisão em massa da fita simples do fagomídeo pBluescript SK(+), desde o vetor Lambda Uni-ZAP XR na bactéria hospedeira *E. coli* XL1-Blue MRF'. A bactéria hospedeira *E. coli* SOLR foi infectada para converter o fagomídeo de fita simples a fagomídeo de dupla fita.

Colônias *E. coli* SOLR, brancas e individuais, contendo o inserto de cDNA clonado, foram transferidas para microplacas com fundo em U de 96 poços, contendo meio de cultura (LB/ampicilina 150 µg/mL) e foram crescidas por 18 horas a 37 °C. Em seguida adicionou-se glicerol a uma concentração final de 15%. As placas identificadas encontram-se conservadas a -70 °C.

Seqüenciamento e análise computacional

DNA plasmidial para seqüenciamento foi preparado empregando-se o “Concert[™] Rapid Plasmid Miniprep System kit” (GibcoBRL, UK). A reação de seqüenciamento empregada é baseada no método de extensão de “primer” terminada por didesoxinucleotídeos trifosfatos (ddNTPs) (Sanger *et al.*, 1977), consistindo de etapas sucessivas de desnaturação do molde (plasmídeo, contendo o inserto de cDNA), pareamento do “primer” a ser estendido e extensão do “primer” pela ação de uma DNA polimerase termoestável, utilizando-se para isso um termociclador GeneAmp, resultando numa amplificação linear dos produtos de extensão.

As reações de seqüenciamento foram realizadas empregando-se o kit de ddNTPs covalentemente ligados a agentes fluorescentes específicos para cada ddNTP (ABI Prism[®] BigDye[™] Terminator Cycle Sequencing Ready Reaction kit with AmpliTaq[®] DNA polymerase FS), nas seguintes condições:

Mistura reacional:

- 2,0 µL de suspensão de DNA molde (plasmídeo; 200 – 300 ng.);
- 2,0 µL BigDye™ premix;
- 1,0 µL de “primer” 10 µM (M13 reverso);
- 3,0 µL de tampão 2,5x (200 mmol/L Tris-Cl pH 9,0; 5 mmol/L MgCl₂);
- 2,0 µL água destilada.

Condições de extensão:

- 1 ciclo de 60 segundos a 94 °C,;
- 35 ciclos de 20 segundos a 94 °C,
- 20 segundos a 52 °C, 4 minutos 60 °C; e 4 °C.

A purificação dos produtos das reações de seqüenciamento foi realizada por precipitação com isopropanol (Sambrook *et al.*, 1989). As amostras purificadas e concentradas foram suspensas em 2 µL de tampão de aplicação (formamida deionizada/blue dextran a 5% em EDTA 25 mmol/L pH 8,0; 5:1 v/v).

As amostras são desnaturadas por 3 minutos a 95 °C e aplicadas (1,0 µL) em gel de poliacrilamida (36 x 21 x 0,02 cm, com pente dente de tubarão para 96 amostras), composto de uréia a 6 M (Sigma, USA), solução acrilamida/bisacrilamida (Long ranger FMC, USA) a 5%, persulfato de amônio (Amresco, USA) a 0,05%, TEMED (Amresco, USA) a 0,05% em tampão TBE pH 8,0 (Tris-base 89 mmol/L; ácido bórico 89 mmol/L; e EDTA 2 mmol/L).

Para a obtenção dos dados de seqüenciamento do seqüenciador automático foram empregados os módulos de programação do seqüenciador para eletroforeses desenvolvidas em gel de 36,0 cm, química de seqüenciamento BigDye™, tanto a 1200 V para 7 horas de aquisição de dados (SeqRun 36E-1200) quanto a 2400 V para 3,5 horas de aquisição de dados (SeqRun 36E-2400).

Os dados do seqüenciamento automático foram coletados pelo programa dedicado ABI Prism Data Collection (Applied Biosystem Inc., USA) em um microcomputador PowerMacintosh 7300/200 (Apple Microcomputer Inc., USA). A análise dos géis de seqüenciamento, foi realizada pelo programa Sequencing Analysis v.3.3 (Applied Biosystem Inc., USA), em um microcomputador PowerMacintosh G3 (Apple Microcomputer Inc., USA), que forneceu um perfil eletroforético para cada amostra seqüenciada.

Os perfis eletroforéticos obtidos foram analisados empregando-se ferramentas de bioinformática estabelecidas em nosso laboratório - microcomputador PC com sistema DEBIAN/GNU LINUX e pacote de programas de bioinformática phred/phrap/crossmatch e consed (Ewing *et al.*, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998) - sendo analisados de acordo a seguinte rotina:

- 1- Determinação de qualidade do inserto e identificação de seqüências de vetor: A qualidade da seqüência é dada para cada base através da análise do perfil eletroforético pelo programa phred. Perfis eletroforéticos com mínimo de 150 bases com qualidade phred ($Q \geq 20$), são aceitos; perfis eletroforéticos de baixa qualidade e correspondentes a seqüência do vetor não são aceitos, sendo retirados da base de dados.
- 2- Montagem de contíguos (“contigs”): A montagem de contíguos consiste na junção ordenada dos perfis eletroforéticos baseada na similaridade entre as seqüências e critérios de qualidade de seqüenciamento para a formação de contíguos de seqüências. As comparações das seqüências dos perfis eletroforéticos aceitos, já com qualidade determinada para cada base, e a montagem de contíguos de seqüências são realizadas pelo programa phrap/crossmatch. São atribuídos como parâmetros uma nota (“score”) mínima de 100, em um mínimo de 80 bases e penalidade de -15 a bases não idênticas, estabelecendo-se uma seqüência comum (consenso) como resultado da montagem de varias seqüências. O programa consed é utilizado para a visualização gráfica dos arquivos de montagem de contíguos e determinação de regiões das seqüências dos contíguos com qualidade baixa ou discrepâncias. As seqüências que não constituem contíguos permanecem como seqüências singletos (“singletons”).
- 3- Edição: As seqüências são convertidas em arquivos com formato “fasta” para se proceder à a remoção das regiões de baixa qualidade (Ns) e as seqüências do vetor.
- 4- NetBlast e análise de similaridade: Arquivos em formato fasta, correspondentes a cada seqüência (singletos e o consenso dos contíguos), são enviados, empregando-se o programa BlastCl3 (NCBI, USA), para comparação com a base de dados, não redundante (NR), de proteínas do GenBank (NCBI, USA). O programa BLASTX (com a matriz de substituição BLOSUM-62, e parâmetros standard) (Altschul *et al.*, 1990) (NCBI, USA), que realiza a comparação de uma seqüência de DNA contra uma base pública de seqüências de proteínas, foi empregado na análise de similaridade. Este programa traduz a seqüência de DNA (nas 6 etapas possíveis de

leitura) em sequência de aminoácidos. Após comparação com a base pública de dados de proteínas, são recebidos arquivos, um para cada sequência comparada, contendo o resultado da análise realizada pelo programa BLASTX. A similaridade com alguma sequência do banco de dados é determinada pela nota (“score”) e o “E-value”, que indica a probabilidade que a nota obtida possa apresentar-se novamente (Altschul *et al.*, 1990; Gish *et al.*, 1993). Sequências que apresentam similaridade de nota ≥ 80 , são consideradas putativamente identificadas, sequências com nota menor ou sem similaridade são consideradas não identificadas. As ESTs putativamente identificadas foram anotados em categorias e sub-categorias funcionais seguindo o esquema EGAD (TIGR, Rockville, MD, USA) (White *et al.*, 1996).

- 5- Base de dados de EST de *T. reesei*; As sequências EST obtidas (putativamente identificadas e não identificadas), foram anotadas e estão apresentadas em uma página de Internet, que agrupa as ESTs por categorias funcionais.

“Microarray” e análise computacional dos dados.

Neste trabalho foram identificados 1.151 genes e a base de dados do GenBank (NCBI, USA), dispunha de 109 genes. Dentre estes 109 foram escolhidos 23 genes para serem incluídos na análise. Os outros genes não foram considerados por serem correspondentes a proteínas ribossomais ou a sequências sem relação com o escopo de nosso trabalho.

Com base na sequência de 23 genes de *T. reesei*, foram desenhados oligonucleotídeos que foram utilizados na amplificação das sequências codificadoras dos mesmos, a partir do DNA de *T. reesei*, por reação de PCR. Os oligonucleotídeos empregados na amplificação de um fragmento de cada gene são mostrados na Tabela 1.

Para a amplificação foram utilizados: 1 ciclo de 94 °C/ 30 segundos; 30 ciclos de 94 °C/1 minuto, 60 °C/1 minuto e 72 °C/2 minutos e 72 °C/2 minutos, 4 °C, nas condições indicadas pelo fabricante do kit Taq DNA Polymerase (Gibco, BRL).

O DNA amplificado foi purificado, e tratado como indicado para todas as amostras a serem empregadas no “microarray”.

Após a análise das ESTs obtidas, foram selecionados os clones contendo o cDNA a ser empregado no “microarray” e arranjados em novas placas de 96 poços. As bactérias foram crescidas a 37 °C/18 horas e os insertos contidos no fagomídeo foram amplificados

Tabela 1. Oligonucleotídeos empregados na amplificação de genes de *T. reesei*, incluídos no "microarray".

SIGLA	IDENTIFICAÇÃO	ACCESO GENBANK (NCBI. USA)	OLIGONUCLEOTÍDEO											
PKI 1	pyruvate kinase	gi 170552 gb L07060 TRRPKI	TRPKI1-R	5'	-	CCT	ACC	TTG	TCT	GCC	TTT	GG	-	3'
			TRPKI1-F	5'	-	GGT	GTT	TTC	AGC	CTT	GAG	AC	-	3'
CREA	CreA	gi 940428 dbj D63514 TRRCREA	TRCREAR	5'	-	GTC	TGA	GAG	CCG	ACA	AGT	AC	-	3'
			TRCREAF	5'	-	GTG	CGA	AAA	GGT	GGG	AGA	AG	-	3'
XYLANES	acetyl xylan esterase	gi 1431619 emb Z69256 TRAXE1	TRXYLANES-R	5'	-	GAG	CCA	AAA	TGC	CTT	CAG	TC	-	3'
			TRXYLANES-F	5'	-	CTC	CAA	GCT	AGT	CCT	AAC	TG	-	3'
XYLAN2	endoxy lanase II	gi 455906 gb S67387 S67387 xln2	TRXYLAN2-R	5'	-	CTT	ATA	AAG	GAG	CCC	ATG	CC	-	3'
			TRXYLAN2-F	5'	-	GAG	CCT	CTA	CAG	AGC	CAA	AG	-	3'
EDGL	Endoglucanase	gi 2168675 dbj E00390 E00390	TREDGL-R	5'	-	CGC	AAA	CTT	TGT	CTT	GTC	GC	-	3'
			TREDGL-F	5'	-	CAG	ACC	AGA	GGC	AAG	TCA	AC	-	3'
GALACT	alpha-galactosidase	gi 1580817 emb Z69254 TRAGL2	TRGALACT-R	5'	-	CTT	CCA	GTA	CAA	GTC	GCA	CC	-	3'
			TRGALACT-F	5'	-	CTG	GAA	GTA	GAA	CAG	CAC	CG	-	3'
TRS1	trs1	gi 2330881 emb Y14562 TRTRS1	TRTRSI-R	5'	-	CTA	CAG	AAC	CTT	CCA	GAC	CC	-	3'
			TRTRSI-F	5'	-	CCA	TTC	GAT	TTA	GCC	GAA	GC	-	3'
SAR1	sar1	gi 1771849 emb Y08636 TRSAR1	TRSAR1-R	5'	-	CCT	TTC	CGA	TCC	CAC	TCT	AG	-	3'
			TRSAR1-F	5'	-	GAA	AGC	CAA	CAA	GTT	TCG	CC	-	3'
TEF1	translation elongation factor	gi 312886 emb Z23012 TRTEF1G	TRTEFI-R	5'	-	CAT	CCG	AAA	CTG	ACG	TTC	TG	-	3'
			TRTEFI-F	5'	-	GAG	AAG	TTC	TGA	TGG	AAC	CC	-	3'
GLUCOR	alpha-glucuronidase	gi 1419337 emb Z68706 TRAGLUCS	TRGLUCOR-R	5'	-	CCA	GTG	GGA	CAA	CCT	CAA	TG	-	3'
			TRGLUCOR-F	5'	-	GAC	ATT	GAC	GAC	CGT	GAA	GC	-	3'
EDGL IV	endo-beta-1,4-glucanase,	gi 2116582 dbj AB003694 AB003694	TREDGLIV-R	5'	-	GCA	TCT	ATA	AGA	TGG	CAC	AG	-	3'
			TREDGLIV-F	5'	-	CCC	TCG	CTC	AAT	AAC	GTA	TG	-	3'
HFB1	hydrophobin 1	gi 1085117 emb Z68124 TRHFB1	TRHFB1-R	5'	-	GAT	CTC	GAA	AGG	CCC	TTG	TC	-	3'
			TRHFB1-F	5'	-	CTG	GTA	AGG	GAA	GTT	GAC	TG	-	3'

XYLO	beta-xylosidase	gi 2791277 emb Z69257 TRBXL1	TRXYLO-R	5' - CCT CAA CCG CAC ATT GAT CC - 3'
			TRXYLO-F	5' - GGA TGC TGA GCT TGG AAG AG - 3'
HFB2	hydrophobin 2 orotate phosphoribosyl	gi 1903324 emb Y11894 TRHFB2	TRHFB2-R	5' - CTC CTC TTT CTC AAC TCT CC - 3'
			TRHFB2-F	5' - CTA TGT CCT GCT CGC AAA TC - 3'
URA5	transferase	gi 5187 emb X55879 TROPRTG	TRURA5-R	5' - GCT TGT GGG ATA CTA AGA CC - 3'
			TRURA5-F	5' - GTC TTT CTC TCC CTT CTG AC - 3'
MANN	beta-mannanase orotidine-5'-phosphate	gi 506847 gb L25310 TRRBEMA	TRMANN-R	5' - CTC AAA GAG TCT CCT TTC CG - 3'
			TRMANN-F	5' - CAG GAC TTA AGA GTT GGT GG - 3'
URA3	decarboxylase beta-D-glucoside	gi 5191 emb X55880 TRURA3G	TRURA3-R	5' - CAC GAA ATT CTC CCA ATC GC - 3'
			TRURA3-F	5' - CAA CTC CAA ACC ATC CTA CC - 3'
BGL1	glucohydrolase (bgl1)	gi 493579 gb U09580 TRU09580	TRBGL1-R	5' - GCT CAA TCT CCA AGA TAA GG - 3'
			TRBGL1-F	5' - CAC AGT CCA TAG CCG AAC TC - 3'
18	18S ribosomal RNA	gi 3021390 emb AJ004964 HJA004964	TR18-R	5' - GAA GTC GTA ACA AGG TCT CC - 3'
			TR18-F	5' - GAT ATG CTT AAG TTC AGC GG - 3'
EDGL 4	Endoglucanase IV	gi 2315273 emb Y11113 TRENDGIV	TREDGL4-R	5' - CAT CAA CGG GGT GTG GTA TC - 3'
			TREDGL4-F	5' - CCC GTA GAC TTC ATC TAG TG - 3'
XYLAN	endo-beta-1,4-xylanase I	gi 262450 gb S51973 S51973 xyn1	TRXYLAN1-R	5' - GAA GCG AGG CAA GCT TGA TG - 3'
			TRXYLAN1-F	5' - CAT CAC AAA AGA AGA GCC CC - 3'
25	25S ribosomal	gi 3163910 gb AA985385 AA985385	TR25-R	5' - CTT GTT CGC GGC TAA ATG AC - 3'
			TR25-F	5' - CCA GTT ATC CCT GTG GTA AC - 3'
ACT	Actin	gi 927566 emb X75421 TRACTIN	TRACT-R	5' -GAG CTT CGT CTC TGT CTA TC - 3'
			TRACT-F	5' -GGA ACA ATC GGA ACA AAC AG - 3'

diretamente a partir das colônias. As condições de amplificação dos insertos, empregando-se o kit BIOLASE DNA polimerase, foram as seguintes:

Mistura reacional:

- 13 µL de dNTPs 1mmol/L (Amersham Pharmacia, UK)
- 10 µL de tampão 10x NH₄
- 0,7 µL de “primer” SK 20 µM (Stratagene, USA)
- 0,7 µL de “primer” T₇ 20 µM (Stratagene, USA)
- 5 µL de MgCl₂ 50mmol/L
- Enzima BIOLASE 1 unidade
- Aproximadamente 1 µL de bactéria, contendo o DNA molde
- Água em um volume final de 100 µL.

Condições de extensão:

- 1 ciclo de 4 minutos a 95 °C
- 40 ciclos das etapas: 45 segundos a 95 °C, 45 segundos a 55 °C, 2 minutos a 72 °C
- 1 ciclo de 1 minuto a 72 e 4 °C .

Os produtos de PCR foram purificados, empregando-se placas de 96 poços do Sistema MultiScreen-FB (Millipore U.S.A.), e concentrados de acordo com as seguintes especificações da Genomic Solutions Inc (MI. USA):

- 1- Adicionar 100 µL de tampão de ligação (7M Guanidine-HCl: 200 mmol/L tampão MÊS, pH 5,6), a 100 µL do produto de PCR, em cada poço de uma placa de filtro MultiScreen-FB (MAFB NOB50), misturar rapidamente, com a pipeta, no mínimo 5 vezes.
- 2- Fixar firmemente na parte inferior da placa FB, uma placa de 96 poços com fundo em U. Centrifugar a 2500 rpm/1 minuto. Descartar o filtrado.
- 3- Adicionar 200 µL de etanol a 80% a cada poço da placa FB. Centrifugar a 2500 rpm/1 minuto. Repetir a mesma etapa mais uma vez. Descartar o filtrado.
- 4- Eliminar o resíduo de etanol por centrifugação a 2500 rpm/5 minutos. Descartar o filtrado.
- 5- Fixar firmemente na parte inferior da placa FB, uma placa de 96 poços com fundo em V. Adicionar 50 µL de tampão TE 0,1x em cada poço. Agitar a 37 °C/100 rpm/5 minutos. Centrifugar a 2500 rpm/5 minutos.
- 6- Adicionar 50 µL de tampão TE 0.1x em cada poço. Agitar a 37 °C/100 rpm/5 minutos. Centrifugar a 2500 rpm/5 minutos.

- 7- Descartar a placa FB e conservar a placa com o filtrado contendo o DNA purificado ($\cong 100 \mu\text{L}$ por poço).
- 8- Adicionar 1/10 do volume de AcNa 3M (pH 5,2) e um volume igual de isopropanol. Misturar com a micropipeta e incubar toda a noite a -20°C .
- 9- Centrifugar a 4600 rpm/30 minutos a 4°C .
- 10- Descartar o sobrenadante, por inversão da placa sobre folhas de papel toalha.
- 11- Adicionar $100 \mu\text{L}$ de etanol a 70%, homogeneizar suavemente e deixar a temperatura ambiente por 5 minutos.
- 12- Centrifugar a 4600 rpm/15 minutos a 4°C . Repetir a etapa 10.
- 13- Secar o DNA precipitado a vácuo por 10-15 minutos. Evitando secar em demasia.
- 14- Dissolver o DNA de cada poço com $6 \mu\text{L}$ de DMSO a 50%.
- 15- Realizar a eletroforeses de $1 \mu\text{L}$ de cada amostra, em gel de Agarose a 1%, para determinar a qualidade das amostras.
- 16- Consolidar 4 placas de 96 poços, em uma placa de 384 poços, com fundo em V (Thermo-Fast[®] Abgene, UK).

As amostras de cDNA e de RNA para o “microarray” foram enviadas à empresa Genomic Solutions Inc. (MI, USA), onde foram realizadas as seguintes etapas do experimento: a preparação das lâminas do “microarray” foi feita segundo o padrão do GeneTAC Flexys[®] Modular Robotic Workstations (Genomics Solutions Inc.); o cDNA (0.2 a $0.5 \mu\text{g}/\mu\text{L}$) foi impresso em duplicata em lâminas de vidro ($36,85 \times 22,26 \text{ mm}$), sendo distribuído em 32 blocos de 10×10 elementos de $100 \mu\text{m}$ de diâmetro cada um; cada bloco constituído por 48 elementos impressos em duplicata, mais um elemento controle (gene Q do fago lambda) e três posições, em volta do controle, foram deixadas em branco. O elemento controle pode ser uma sequência sintética de DNA, ou a sequência de DNA de um organismo diferente do estudado, e que é impressa na lâmina e adicionada às amostras de RNA, em igual concentração, no momento da transcrição reversa (Yang *et al.*, 2001).

A Figura 7 mostra o padrão do GeneTAC Flexys[®] Modular Robotic Workstations para a distribuição dos elementos na lâmina (A), distribuição das duplicatas em cada bloco (B), distribuição dos blocos (8×4) de elementos (C) e a qualidade do DNA impresso na lâmina foi confirmada com o corante SybrGreen (D). A análise da imagem por SybrGreen (Fig. 7D), mostra que a qualidade da impressão e do DNA empregado era adequada para a realização da análise de “microarray”.

O cDNA marcado com moléculas fluorescentes pode ser obtido de forma direta ou indireta. No primeiro processo, o cDNA é marcado no momento da transcrição reversa empregando-se um dNTP ligado à molécula fluorescente (marcação em etapa simples). No segundo processo o cDNA é sintetizado empregando-se um aminoalil-dUTP que posteriormente se liga à molécula fluorescente (marcação em duas etapas). Neste trabalho o cDNA foi marcado em duas etapas por ser um processo mais eficiente do que a marcação em etapa simples que depende da eficiência da enzima transcriptase reversa para a incorporação do dNTP ligado à molécula fluorescente. Na Figura 8 é mostrada a estrutura das moléculas fluorescentes Cy3 e Cy5, Cy5 ligada à dUTP e o aminoallyl-dUTP.

Síntese de cDNA, obtido a partir de 15 µg de RNA foi feita empregando-se o kit Clontech's Atlas Glass Fluorescent labeling Kit (www.clontech.com). O RNA colhido a 83 mmol/L de glicose foi empregado como referência e convertido em cDNA marcado com N-hydroxysuccinimide-fluoróforo verde Cy3, enquanto que o RNA dos pontos seguintes (46 mmol/L, 11 mmol/L, 1 mmol/L e 0 mmol/L), foi convertido em cDNA na presença do N-hydroxysuccinimide-fluoróforo vermelho Cy5.

Após a marcação, o cDNA foi misturado e suspenso em 100 µl de solução de hibridação. A hibridação foi feita no GeneTAC™ Hybridization Station com uma desnaturação inicial de 10 segundos a 95 °C, cDNA marcado foi adicionado a 75 °C e a hibridação foi desde 65 °C (6 horas) a 55 °C (15 horas). As lâminas foram lavadas manualmente. Após a secagem, as imagens do “microarray” foram digitalizadas no GeneTAC 2000 Microarray Analyzer (Genomics Solutions Inc., USA).

As imagens do “microarray”, recebidas de Genomics Solutions, foram analisadas num computador PC, com sistema operacional Windows NT e banco de dados SQL, empregando o programa GeneTAC Integrator v. 3.0.1 (Genomics Solutions Inc., USA).

A análise da imagem tem início com a construção de uma grade que servirá para a extração dos dados de intensidade de cada um dos elementos na lâmina. Os programas de análise de imagem têm uma rotina a ser seguida na preparação da grade, a qual deve ser construída segundo as características de construção da lâmina de “microarray”. Assim, a grade deverá apresentar 32 blocos de 8 x 4, cada um com 100 elementos de 10 x 10, sendo inserida na base de dados a relação e a localização dos elementos na lâmina. Após a construção e sobreposição da grade sobre cada uma das imagens das lâminas, é realizada a extração automática de dados de intensidade de cada elemento.

Figura 7. Perfil da lâmina de "microarray". A) Distribuição do DNA em cada bloco, B) Localização das duplicatas e gene controle em cada bloco, C) Distribuição dos 32 blocos (A-H e 1-4) de elementos na lâmina e D) Imagem SybrGreen, mostrando a impressão do DNA.

Figura 8. Estrutura de Cy3, Cy5, amino allyl-dUTP e dUTP ligado à Cy5.

Diversos programas para analisar os dados provenientes da imagem do “microarray”, podem ser obtidos comercialmente ou através de centros de pesquisa (<http://ihome.cuhk.edu.hk/~b400559/>). O tratamento dos dados foi realizado utilizando-se o programa Microsoft® Excel 2000 (Microsoft Co. USA).

Obtidos os valores de intensidade de fluorescência de cada elemento para cada um dos fluoróforos (Cy3 e Cy5), a primeira etapa de análise dos dados consistiu na normalização da intensidade dessa fluorescência para cada um dos fluoróforos, o que permitiu a correta comparação do nível de expressão de cada elemento.

Primeiramente deve-se determinar a necessidade, ou não, da subtração da fluorescência residual (“background”, Bg) do valor da intensidade de fluorescência obtido para cada elemento. Para isso foi obtida a razão Cy5/Cy3 dos dados correspondentes aos genes controle de cada lâmina com (RC-Bg) e sem subtração do valor de “background” (RC). Os resultados obtidos foram analisados estatisticamente e levados a um gráfico que mostra o desvio padrão dos dados dos elementos controle, empregando-se o programa StatView (Abacus, USA). O resultado foi considerado aceitável quando a razão da maioria dos elementos foi < 2 vezes o SD (desvio padrão).

A análise (ver resultados) mostra que se deve empregar a razão obtida dos elementos controle sem subtração do “background” (RC) e, também, decidiu-se empregar a razão do controle (RC) de cada bloco de amostras (normalização por controle local).

A rotina de normalização, aplicada aos dados de cada uma das lâminas e que forneceu os valores normalizados de cada elemento, foi a seguinte:

1. Obtenção da razão Cy5/Cy3 de cada elemento (RE);
2. A partir da razão Cy5/Cy3 do controle (RC) obteve-se a razão normalizada de cada elemento (RNE) pela divisão RE/RC.

Obtidos os valores normalizados de cada elemento e de sua correspondente duplicata (RNE₁ e RNE₂), o desvio relativo (D) de cada elemento é calculado como o valor absoluto da diferença do RNE (em duplicata), dividido pelo menor deles, empregando-se a seguinte equação:

$$D = |RNE_1 - RNE_2| / \min(RNE_1, RNE_2)$$

Na tentativa de se aproveitar o maior número possível de genes utilizou-se o critério de aceitar a média da duplicata ($\overline{RNE}_{A1}$, $\overline{RNE}_{A2}$) até desvio relativo igual a 1. Isto significa que a diferença entre a duplicata pode ser de até 100%, considerando-se que as lâminas A1 e A2 (46 mmol/L e 11 mmol/L de glicose, respectivamente) representam o estado reprimido e que as lâminas A3 e A4 (1 mmol/L e 0 mmol/L de glicose,

respectivamente) representam o estado em que não há repressão por glicose. Estabelecemos que se o desvio relativo entre a duplicata da lâmina A1 fosse maior que 1, seria determinado o desvio relativo da duplicata na lâmina A2. Com isto, pode ocorrer uma das duas situações: 1) Se a duplicata da lâmina A2 também tiver desvio relativo maior que 1, todos os elementos correspondentes a esse gene serão descartados; 2) Se a duplicata da lâmina A2 tiver desvio relativo menor do que 1, o valor de RNE_{A2} será a média entre a duplicata ($\overline{RNE}_{A2}$) e o valor de RNE_{A1} será correspondente ao valor da duplicata que esteja mais próximo ao valor de $\overline{RNE}_{A2}$. O processo está esquematizado no fluxograma abaixo para as lâminas A1 e A2 (análogo ao processo das lâminas A3 e A4):

Resumindo, o processo acima nos permite obter 3 possíveis resultados:

- 1- Obter a média da duplicata (RNE_1 e RNE_2), de cada elemento, quando o desvio relativo entre eles é menor ou igual a 100%;
- 2- Substituir o valor do elemento pela palavra “ERRO”, quando o desvio relativo entre o mesmo elemento na lâmina A1 e A2 for maior que 100%. Neste caso, o gene será descartado da base de dados;
- 3- Escolher o valor da duplicata do elemento, em A1 ou A2, que seja mais próximo ao valor da média da duplicata ($\overline{RNE}$) na lâmina A1 ou A2;

Uma vez construída uma Tabela com os dados de expressão de cada gene, procede-se à escolha de um fator de expressão diferencial, F. Como mencionado anteriormente, as lâminas A1 e A2 representam o estado reprimido enquanto as lâminas A3 e A4, representam o estado sem repressão por glicose. Assim, realizou-se a comparação em

bloco dos valores de $A3/A4$ versus $A1/A2$ e estudou-se a variação de F entre 1,5 e 3 para essa comparação, sendo escolhido o valor de 2 como o que produziu o melhor resultado.

Em seguida, utilizou-se o fluxograma abaixo onde $A1$, $A2$, $A3$ e $A4$ representam os valores obtidos de cada elemento na seleção anterior.

O processo acima nos permite obter 3 possíveis resultados:

1. Se algum elemento apresentar a inscrição “ERRO”, este é automaticamente descartado da base de dados. Isto significa que não existe relação entre os valores das duplicatas em duas lâminas consecutivas ($A1$ com $A2$ ou $A3$ com $A4$);

2. Se a inscrição for “Não Aceitar este Gene” significa que os dados do elemento não apresentam um valor maior ou menor que o estabelecido pelo fator em relação a seus valores nas lâminas A3 e A4;
3. Quando a inscrição for “Aceitar este Gene” os dados do elemento apresentam um valor maior ou menor que o fator estabelecido em relação a seus valores nas lâminas A3 e A4. Os elementos, que serão denominados “GENES” são agrupados em nova tabela, para se poder continuar a análise.

Os genes aceitos foram agrupados por grupos hierárquicos (“Hierarchical Clustering”) empregando-se o programa Gene Cluster e a visualização foi feita através do programa TreeView (Eisen *et al.*, 1998). Ambos os programas foram obtidos junto à Universidade de Stanford, EUA, (<http://genome-www5.stanford.edu/MicroArray/SMD/>) e foram descritos por Eisen *et al.*, (1998). A análise da árvore nos permite identificar os genes que são induzidos quando a glicose se esgota no meio de cultura.

“Northern blot”

Dez microgramas de RNA (empregado no “microarray”) foram desnaturados com Glyoxal/DMSO, fracionados por eletroforeses em gel de agarose a 1% e transferidos para membrana Hybond-N⁺ (Amersham Pharmacia, U.K.), e hibridados. Insetos de cDNA, amplificados por PCR, de genes de *T. reesei*, foram empregados na preparação de sondas radioativas, empregando-se o sistema de “random primers DNA labeling system” (Life Technologies, U.S.A.) na presença de (α -³²P) dATP (Amersham), tendo como controle o gene de actina (ACT) de *T. reesei* (Matheucci *et al.*, 1995). Hibridação em formamida a 43 °C/18 horas e lavagem foram feitas como descrita pelo fabricante da membrana. A membrana foi exposta a filme auto-radiográfico em diferentes intervalos de tempo.

Atividade enzimática de álcool desidrogenase (ADH)

Massa micelial obtida a 46 mmol/L, 1 mmol/L e 0mmol/L de glicose foi empregada na determinação de atividade ADH. No momento do rompimento celular, foi adicionado à massa de células o mesmo volume de tampão Tris/HCl 50 mmol/L, pH 7,5 contendo: MgCl₂ 2 mmol/L, EDTA 0.2 mmol/L, ácido aminocapróico 2 mmol/L, DTT 2 mmol/L e PMSF 1 mmol/L e também 1 volume de pérolas de vidro (0.5 mm). A mistura foi submetida a agitação com Vortex (PHOENIX AT56) com posterior centrifugação (10.000 x g/10min) para remoção dos restos celulares (“débris”) e pérolas de vidro. O sobrenadante foi utilizado para as medidas de atividade enzimática (Beutler, 1984).

A mistura de reação para medida da formação do acetaldeído foi: Tris/HCl 50 mmol/L, pH 8,8, etanol 0,7 M e NAD 3 mmol/L. A redução do NAD foi acompanhada em espectrofotômetro a 340 nm.

A mistura de reação para formação do etanol a partir do acetaldeído foi: Tris/HCl 50 mmol/L pH 7,0, acetaldeído 0,7 M e NADH 2 mmol/L. A oxidação do NADH foi acompanhada em espectrofotômetro a 340 nm.

RESULTADOS

Estabelecimento da base de dados de EST de *T. reesei*

Montagem e avaliação da biblioteca de cDNA

O RNA de células de *T. reesei*, crescidas na presença de glicerol, foi isolado pela técnica de isotiocianato de guanidina/CsCl (Chirgwin, *et al.*, 1979), e utilizado na construção da biblioteca de cDNA, usando o sistema Uni-ZAP[®] XR vector (Stratagene, USA). Este sistema foi empregado por sua eficiência na obtenção do maior número de clones recombinantes e por permitir que os insertos de cDNA sejam clonados de forma direcionada no fagomídeo pBluescript SK(+), contendo sistema de seleção por expressão de *lacZ* na presença de IPTG/X-Gal. Isso permite obter-se a seqüência dos insertos a partir da extremidade 5' ou 3' do cDNA, empregando os “primers” M13 reverso ou M13(-20). A biblioteca foi denominada TrEST-A (*Trichoderma reesei* Expressed Sequence Tags, biblioteca A).

Após a construção da biblioteca, foram selecionados aleatoriamente 6.048 clones de bactérias que foram distribuídos em 63 placas de 96 poços, contendo 75 µL de meio líquido (LB/Ampicilina 150 µg/mL) e crescidos por 18 horas, a 37 °C. No final do cultivo foram adicionados a cada poço 25 µL de glicerol a 50% e as placas foram conservadas a -70 °C.

Para iniciar o seqüenciamento de uma biblioteca de cDNA é necessário avaliar sua qualidade. Uma biblioteca de cDNA construída para os propósitos deste trabalho é considerada de boa qualidade quando mostra <2% de vetor sem inserto, <4% de seqüências de organismos contaminantes, >50% de seqüências sem similaridade na base de dados, >50% de seqüências com comprimento >300 bases (Adams *et al.*, 1995; Hillier *et al.*, 1996).

A primeira etapa para determinar a qualidade da biblioteca, consistiu em verificar o tamanho dos insertos. Os insertos de 24 clones, tomados ao acaso, foram amplificados por PCR empregando os “primers” M13-reverso e M13(-20). A Figura 9 mostra o perfil eletroforético, em gel de agarose, dos produtos de PCR dos 24 clones. O tamanho dos insertos analisados oscilou entre 0,8 e 2,4 kb, com tamanho médio de 1,0 kb. Esse tamanho é adequado para conduzir o seqüenciamento com a finalidade de construir um banco de dados de EST.

Figura 9. Eletroforese em gel de agarose, 0,8%, de produto de PCR dos insertos de 24 clones escolhidos aleatoriamente da biblioteca TrEST-A de *T. reesei*.

Outra etapa na avaliação da qualidade da biblioteca consiste em determinar a presença de seqüências contaminantes (por exemplo, seqüências de bactéria) e insertos de orientação invertida e fagomídeos sem inserto. Também é importante analisar se as seqüências identificadas são relacionadas a organismos do mesmo grupo taxonômico de *T. reesei*. Estes dados foram obtidos através do seqüenciamento, a partir da extremidade 5', de insertos de 700 clones selecionados ao acaso.

Considerando que poucas seqüências deste microrganismo encontram-se descritas na base pública de dados GenBank (NCBI, USA), decidiu-se obter as seqüências de EST a partir da extremidade 5' do cDNA, para aumentar a probabilidade de identificar seqüências codificadoras de proteínas. Isso nos permitiu obter a identidade das ESTs por análise de similaridade com seqüências de genes conhecidos e depositados em bases públicas de dados.

Os perfis eletroforéticos do gel de seqüenciamento, obtidos pelo seqüenciador automático ABI 377, foram analisados pelos programas phred/phrap/crossmatch/consed. As seqüências obtidas, geralmente contêm bases ambíguas (Ns) devido ao alongamento e superposição dos picos de leitura (Ewing *et al.*, 1998). Seguindo a rotina estabelecida (ver Materiais e Métodos), as seqüências de baixa qualidade, com <150 nucleotídeos de qualidade phred ≥ 20 , e seqüências do vetor não foram incluídas na base de dados. Por fim, antes de ser adicionado à base de dados, cada EST sofre uma edição para a remoção de regiões com bases ambíguas e seqüências de vetor.

A Tabela 2 mostra os dados da análise inicial da biblioteca. Foram obtidos 492 seqüências (70% das analisadas) com um mínimo de 150 nucleotídeos e $Q \geq 20$ (99% de confiabilidade, que a base seja correta) (Ewing *et al.*, 1998). Das 492 seqüências obtidas, 427 ESTs (86,8%) têm tamanho ≥ 300 nucleotídeos.

Após a obtenção das seqüências de EST deve-se realizar a análise de formação de contíguos (formados por 2 ou mais transcritos de um mesmo gene). Neste tipo de análise, todas as seqüências são comparadas, uma a uma, para se verificar as seqüências repetitivas que formam contíguos. A montagem dos contíguos de seqüências de EST nos permite determinar a redundância (número de vezes que um gene é representado por mais de uma EST) da biblioteca (Adams *et al.*, 1992) e estabelecer o número de contíguos e singletos ("singletons"; transcritos encontrados uma única vez). Isto permite determinar o número não redundante de ESTs, para estabelecer quantos genes únicos (UniGene) estão representados na base de dados.

Tabela 2. Análise de qualidade da biblioteca TrEST-A.

Descrição	Características	%
Título	$\sim 4,2 \times 10^9$ cfu/mL	
Clones com inserto de cDNA	24/24	
Tamanho médio do inserto	1,0 kb (0,8-2,4 kb)	
Número de clones analisados	700	100
Seqüências de vetor	7	1
Seqüências de bactéria	12	2,6
Número de seqüências ESTs obtidas	492	70
ESTs com ≥ 300 nucleotídeos	427	86,8
ESTs singletos	242	74,7
ESTs contíguos	82	25,3
ESTs com similaridade na base de dados	152	46,9
ESTs sem similaridade na base de dados	172	53,1
ESTs com similaridade a seqüências de fungos	119	78,3
ESTs com similaridade a seqüências de bactérias	12	2,6
ESTs de orientação invertida	0	0

Tabela 3. Número de ESTs de *T. reesei* com similaridade á seqüências depositadas na base pública de dados GenBank.

Organismo	Número de ESTs	%
Leveduras	66	43,42
Fungos filamentosos	53	34,86
Plantas	14	9,2
Vertebrados	6	3,9
Bactérias	4	2,6
Outros	9	6
TOTAL	152	100

Após a análise de redundância da biblioteca obtiveram-se 242 ESTs singletos (74,7%) e 250 ESTs formaram 82 contíguos (25,3%), que formam um total de 324 UniGenes.

Para se realizar a anotação da identidade putativa dos UniGenes e a determinação do seu papel no processo celular, deve-se comparar as seqüências obtidas com as seqüências depositadas nos bancos de dados. Os programas de comparação (BLAST ou FASTA) que são usados nestes bancos empregam um algoritmo e matrizes de comparação de aminoácidos que qualificam a similaridade entre duas seqüências. A similaridade entre as seqüências pode ser estabelecida pela nota e pelo “E-value”. A nota é determinada pelo algoritmo com o objetivo de produzir um alinhamento entre duas seqüências com o maior valor possível de similaridade. O valor da similaridade entre duas seqüências é determinado pela presença, de uma região alinhada de aminoácidos idênticos ou substituições conservadas, que recebem uma pontuação positiva, e de substituições não conservadas, que recebem uma pontuação negativa. Quando não existe similaridade numa região extensa o algoritmo cria uma lacuna em uma das seqüências com o objetivo de encontrar melhor alinhamento. Tanto para a criação das lacunas, como para seu prolongamento, o algoritmo estabelece uma penalidade. Então a similaridade entre as duas seqüências é determinada pela soma total das notas obtidas para cada um dos aminoácidos nas regiões alinhadas, descontadas as penalidades de abertura e prolongamento das lacunas. O “E-value” indica a probabilidade de que tal nota seja obtida novamente ao acaso por uma seqüência qualquer, com o mesmo tamanho e composição de aminoácidos (Altschul, 1991).

Para a determinação da similaridade das seqüências obtidas a seqüências conhecidas e depositadas em bancos de dados foi empregado o programa de comparação BLASTX (Altschul *et al.*, 1990). Este programa traduz a seqüência de nucleotídeos em seqüência de aminoácidos nas 6 possíveis fases de leitura e realiza uma comparação das seqüências traduzidas com uma base não-redundante de seqüências de proteínas. Em nosso caso empregamos o banco público de dados GenBank do NCBI (www.ncbi.nlm.nih.gov).

Uma nota do programa BLASTX ≥ 80 indica que a similaridade entre as duas seqüências é significativa (Pearson, 1991; Lluisma *et al.*, 1997). Um exemplo relacionado aos dados obtidos em nosso banco de dados é mostrado no caso da seqüência TrEST-A1305 que, após a análise de comparação, obteve “E-value” de 10^{-100} quando comparada com Frutose-1,6-bifosfato aldolase, isto é, a probabilidade de que o alinhamento do TrEST-A1305 à seqüência da aldolase tenha ocorrido ao acaso é igual a 10^{-100} , que é

próximo de zero. Em outras palavras, é quase zero a probabilidade que esta EST não seja similar à referida aldolase.

Na análise de similaridade dos 324 genes únicos de *T. reesei* (242 singletos e 82 contíguos), contra a base pública de dados não redundantes de proteínas do GenBank, e somente 152 seqüências (46,91%) apresentaram similaridade com proteínas de função conhecida e 53,08% das seqüências não apresentaram similaridade, mostrando o pouco conhecimento do genoma estrutural dos fungos.

A análise dos 152 genes putativamente identificados mostra que 119 genes (78,28%) apresentaram similaridade com genes de organismos do mesmo grupo taxonômico de *T. reesei* (fungos) e 33 genes (21,72%) mostraram similaridade com genes de organismos de diferentes grupos taxonômicos. A Tabela 3 mostra o número de seqüências agrupadas por organismo com os quais as ESTs de *T. reesei* apresentam similaridade. O maior número de ESTs apresenta similaridade com seqüências correspondentes de leveduras (principalmente *Saccharomyces cerevisiae*). Isto se deve ao fato de que o genoma de *S. cerevisiae* já foi seqüenciado (Goffeau *et al.*, 1996). Das seqüências, obtidas somente 1% corresponde ao vetor, 2,6% correspondem às seqüências de bactérias (diferentes de *E. coli*) e não foram identificados insertos de orientação invertida.

Considerando-se os parâmetros usados na avaliação de uma biblioteca, definidos anteriormente e estabelecidos em outros projetos de seqüenciamento (Adams *et al.*, 1995; Hillier *et al.*, 1996), a biblioteca de cDNA de *T. reesei* foi avaliada como sendo de boa qualidade e adequada para a realização de seqüenciamento em larga escala.

Seqüenciamento em larga escala

Durante o seqüenciamento em larga escala da biblioteca de TrEST-A foram analisados 4.320 clones (correspondentes a 45 placas de 96 poços). Aplicando-se os parâmetros estabelecidos para qualidade do seqüenciamento, foram obtidos 2835 seqüências com um mínimo de 150 bases e com $Q \geq 20$. Isto representa um aproveitamento de 65,6%, um valor aceitável quando comparado com o obtido durante a análise de avaliação da biblioteca (70%).

A média do número de bases por EST foi de 640 nucleotídeos, sendo que 93% das ESTs contêm mais de 300 bases. Somente 4,4% das ESTs apresentam cauda de poli(A⁺) com o mínimo de 15 bases. Considerando-se que o seqüenciamento é feito a partir da extremidade 5', a presença da cauda de poli(A⁺) no fim da seqüência indica que obteve-se

a sequência completa do inserto, o qual, provavelmente, constitui um cDNA de tamanho curto. Na análise das 2.835 sequências, só 0,8% dos insertos foram clonados em orientação invertida.

Das 2.835 ESTs, 793 sequências (28%) se mantiveram como singletos e 2.042 sequências (72%) formaram 358 contíguos. A análise mostra que foram obtidas as sequências parciais de 1.151 genes únicos (UniGenes) expressos em *T. reesei* (Tabela 4). É importante mencionar que a redundância foi incrementando com o progresso do sequenciamento. Isto está mostrado no sequenciamento da placa 45, na qual obteve-se 70 ESTs das quais só 10 (14%) constituem novas sequências. Este é o fator determinante para a decisão de se concluir a etapa de sequenciamento.

De acordo como a nota atribuída pelo programa BLASTX, os UniGenes foram agrupados como putativamente identificados (39%) (BLASTX $\geq$ 80) e não identificados (61 %) (BLASTX <80) (Tabela 4).

Anotação e construção da base de dados de EST

Podem ser atribuídas funções celulares às sequências de ESTs consideradas putativamente identificadas, baseando-se na similaridade com proteínas de função conhecida (Sterky *et al*, 1998). Nosso projeto proporcionou, então, a oportunidade de se classificar sistematicamente os genes expressos em *T. reesei* com relação a seu papel na biologia celular.

Considerando válidos muitos esquemas de anotação, temos agrupado os genes segundo a característica funcional ou o papel celular da proteína expressa, ao invés de se fazer estrita classificação bioquímica. A anotação foi determinada de acordo com o putativo ou conhecido envolvimento da proteína em determinado processo celular ou via metabólica.

A classificação, segundo a categoria do papel biológico (Fig. 10) e de acordo com o esquema EGAD (“Expressed Genome Anatomy Database”) (White *et al.*, 1996), dividiu-se em: divisão celular (1%), sinalização celular (2%), estrutura celular (1%), defesa celular (2%), síntese de proteínas (9%), síntese de RNA (2%), metabolismo (14%), não classificados (genes putativamente identificados mas cujo papel não foi determinado) (5%), desconhecidos (genes hipotéticos, de função não determinada) (3%) e sem similaridade na base de dados (ESTs não identificadas) (61%). Os genes encontrados cobrem ampla variedade de funções bioquímicas e estruturais e mostram ampla visão da biologia celular de *T. reesei*.

Tabela 4. Distribuição dos ESTs de *T. reesei*.

ESTs	Singletos	Contíguos	TOTAL
Identificados	287	173	460
Não identificados	506	185	691
TOTAL	793	358	1151

Tabela 5. Genes com maior número de ocorrências.

EST Representativa	Número de ocorrências	Gene
TrEST-A2534	85	Gliceraldeído 3-fosfato desidrogenase
TrEST-A1761	44	Não identificado
TrEST-A0940	42	Não identificado
TrEST-A1632	36	Não identificado
TrEST-A3395	32	Proteína associada à senescência putativa
TrEST-A2341	30	Proteína hipotética conservada
TrEST-A1145	29	Não identificado
TrEST-A4300	29	Não identificado
TrEST-A3569	28	Não identificado
TrEST-A4285	27	Proteína 9 do gene controlado pelo ciclo celular
TrEST-A3179	26	Não identificado
TrEST-A3477	25	Proteína de resistencia a cadmio
TrEST-A4231	23	rAsp f 4
TrEST-A2006	21	Não identificado
TrEST-A3168	21	Não identificado
TrEST-A4041	20	Proteína de transporte de cobre CTR4
TrEST-A0101	20	Não identificado
TrEST-A3879	20	Não identificado
TrEST-A4092	19	Não identificado
TrEST-A1895	18	Enolase
TrEST-A3973	17	Hidroxilação de ácidos graxos
TrEST-A2009	16	Proteína de biosínteses de tiamina NMT-1
TrEST-A2588	16	Proteína hipotética SCK13.14c
TrEST-A1864	16	Não identificado
TrEST-A2502	15	CAP20
TrEST-A2426	15	Não identificado

Figura 10. Classificação dos 1.151 ESTs transcritos únicos de *T. reesei*.

Muitos dos genes conhecidos encontram-se em grupos funcionais relacionados a genes essenciais para a viabilidade celular e que são expressos em todas as células, em processos tais como síntese de proteínas, metabolismo, síntese de ácidos nucleicos e aminoácidos (Adams *et al.*, 1995). Entre os genes identificados e que apresentam maior número de seqüências de EST podemos mencionar, entre outros, os da: enzima gliceraldeído-3-fosfato desidrogenase (85 ESTs), proteína associada à senescência putativa (“putative senescence-associated protein”; 32 ESTs), proteína 9 do gene controlado pelo ciclo celular (“clock-controlled gene-9 protein”; 27 ESTs) (Tabela 5).

Um dos objetivos deste trabalho foi estabelecer uma base pública de dados de EST de *T. reesei*. Assim, foi construída uma página de Internet, denominada *Trichoderma reesei* EST Database (<http://trichoderma.iq.usp.br/TrEST.html>), onde são mostrados os genes agrupados segundo sua categoria funcional. A Figura 11 apresenta esta página da Internet mostrando a página inicial (A), as EST que se encontram na categoria metabolismo (B) e, como exemplo, a seqüência da EST correspondente ao gene desidroorotase (C).

As seqüências dos UniGenes de *T. reesei* foram depositadas na base de dados dbEST do GenBank (NCBI, USA) e apresentam números de acesso de BM076169 a BM077297. A Tabela 6 mostra um exemplo do formato de submissão de seqüências de EST ao dbEST. O formato apresenta 4 partes que correspondem a dados da publicação do trabalho, informações sobre a biblioteca, pessoa de contato e informações de cada EST.

Cinética do destino da Glicose em *T. reesei*, determinada por análise transcricional através de “Microarray”

A análise dos genes identificados em *T. reesei* mostrou que 16 UniGenes correspondem a genes envolvidos no metabolismo da glicose, incluindo a conversão de glicose a piruvato, o ciclo dos ácidos tricarboxílicos (TCA), a via alternativa do metabolismo de piruvato (“pyruvate bypass”) e a gliconeogênese. Deve-se ressaltar que foram identificados dois UniGenes com similaridade à proteína aldeído desidrogenase, denominados ALD1 e ALD2, que não apresentam similaridade entre suas seqüências e provavelmente tratam-se de genes parálogos. Na Figura 12 são mostrados os genes identificados e sua localização no metabolismo de glicose. A Tabela 7 apresenta a similaridade, obtida através do BLASTX, de ESTs dos UniGenes presentes na Figura 12. A Tabela 7 mostra os “E-value” obtidos da análise de similaridade das ESTs contra

Figura 11. A) Página inicial (*home page*) do web site *Trichoderma reesei* EST database,

Trichoderma reesei EST Data Base - Microsoft Internet Explorer

Arquivo Editar Exibir Favoritos Ferramentas Ajuda

Endereço: <http://domybio.usp.br/tricho/>

Sequence Identification of *T. reesei* Genes

General Function

- Cell Division
- Cell Signaling
- Cell Structure
- Cell Defense
- Protein Synthesis
- RNA Synthesis
- Metabolism
- Unclassified
- No Matches

[Introduction Page](#)

Metabolism			
Clone ID	Best Match	Identification	E-Value
TrEST-A0005	emb CAB57401.1 	di hydrocorotase	8.00E-31
TrEST-A0026	sp P50094 IMH3_YEAST	inosine-5'-monophosphate dehydrogenase	3.00E-27
TrEST-A3190	sp P21594 DCOP_TRIRE	orotidine 5'-phosphate decarboxylase	6.00E-92
TrEST-A3683	ref NP_011575.1 ADE6 	phosphoribosylformylglycinamide synthase	3.00E-24
TrEST-A3972	dbj BAAB19144 	adenylsuccinate synthetase	6.00E-45
TrEST-A3512	dbj AAD53090.1 	extracellular putative dnase	4.00E-43
TrEST-A3871	dbj BAAB3495.1 	nucleoside diphosphate kinase	1.00E-64
TrEST-A2852	ref NP_012293.1 DAL1 	allantoinase	3.00E-53
TrEST-A3570	ref NP_013483.1 ADE13 	adenylsuccinate lyase	2.00E-22
TrEST-A0741	dbj AAC24957.1 	bchh	9.00E-39
TrEST-A0931	sp Q01447 ER24_FUSSO	c-14 sterol reductase	4.00E-63
TrEST-A1302	gi 2649289	acyl-coa dehydrogenase	1.00E-15
TrEST-A1975	emb CAA12224.1 	atp citrate lyase	2.00E-58
TrEST-A2142	ref NP_014353.1 	probable leukotriene a-4 hydrolase	2.00E-29
TrEST-A2385	gi 2773042	phospholipase d2	3.00E-16
TrEST-A2547	sp P14604 ECHM_RAT	enoyl-coa hydratase	4.00E-36
TrEST-A2844	sp Q09925 LCB2_SCHPO	serine palmitoyltransferase 2	4.00E-36
TrEST-A3347	dbj BAAB32278 	phospholipase d	7.00E-33
TrEST-A3385	emb CAA21095 	phosphatidylethanolamine methyltransferase	5.00E-30
TrEST-A3469	dbj AAC97606.1 	elcuri col 14alpha demethylase	5.00E-82
TrEST-A3673	sp P15368 FAS2_PENPA	fatty acid synthase	9.00E-81
TrEST-A4018	sp Q92254 ERG2_NEUCB	c-8 sterol isomerase	4.00E-31
TrEST-A4091	sp P11943 ACPM_NEUCB	acyl carrier protein	3.00E-26
TrEST-A4102	sp Q09925 LCB2_SCHPO	serine palmitoyltransferase 2	3.00E-45
TrEST-A4222	ref NP_012142.1 AYR1 	1-acyl dihydroxyacetone phosphate reductase	4.00E-24
TrEST-A4229	emb CAB55552.1 	fox2 protein	4.00E-17

Internet

Trichoderma reesei ESTs

Figura 11 B) Parte dos ESTs anotados na categoria de metabolismo.

Trichoderma reesei EST Data Base - Microsoft Internet Explorer

Arquivo Editar Exibir Favoritos Ferramentas Ajuda

Voltar - - Pesquisar Favoritos Histórico

Endereço <http://dorm/ab.iq.usp.br/tricho/> Links

T. reesei clone report

Sequence Identification of T. reesei Genes

General Function

- Cell Division
- Cell Signaling
- Cell Structure
- Cell Defense
- Protein Synthesis
- RNA Synthesis
- Metabolism**
- Unclassified
- No Matches

[Introduction Page](#)

Clone ID	TrEST-A0005
Best BLASTX Match	emb CAB57401.1
General Function	Metabolism
Function Description	dihydroorotase

The following sequence(s) were assigned to this function:

Sequence ID	Sequence
TrEST-A0005	<p>CAGAGAGGAAATAGAAAGGAGAGAAAAAGAGGAGAAATGCAAGCTCAA</p> <p>GACGGTCCAGAAAGCTGGAGCTGGCCGGCAGCGGGGACATGCAAGTCCAGC</p> <p>TGCGCCAGGGCGAGATGAGGGAGGTTGGTGGCGGACCTCGGAGCGAGGA</p> <p>GGAGTCGAGACTGCCCTTTGTTAGCCGAACTGGTCCCTCCCTTGACTGA</p> <p>AGTTGACAGAGTCCGATTAACGAGGCAAGCTCCGTCGCAATTGACGCA</p> <p>AGTCCAGCTTTCTCATGTCCCTTACCTTCATCCCTCAGTCAGCGCAGAG</p> <p>ACGATAGGAGAGGAGCGAGAGGCTGGAATCAGCGGGTGTGAAAGCTGTATTC</p> <p>CCAGGGGCGAGAGGAGGATCTGAGGATGGAATGAGCAGCGCTCAGGGCT</p> <p>TTTACATACCTTTGGCCAGATTGGAGAGGCTATGAGCTTGGTGGCGAAAC</p> <p>CGCGACCGAGGGGCTTCCGGGGTTTGGAGCCGAAAGGCTATCATTTTG</p> <p>GAAAAGGGTTTTTTTCTGCTTCTAGAAATCTTATGAGGGGTTCCCT</p> <p>ACCTTGGCTCTGGGTTTGAATTTGGCCCGGTTGGTGGTGGCAAAACA</p> <p>TATGATCTTGGCCCATATTGGATGGAAATTTGGCGCAAGTGTTCCT</p> <p>TAGAGGACAAAGGCTTGGGACTCTTTTTTTTTTGTGCTGGCGGAGG</p> <p>AGCGAGATGAGGTTGGG</p>

Concluido

Internet

Iniciador

FAPESPAP...

Trichoder...

Netcape...

RESES

Microsoft ...

5:30 PM

Figura 11. C) Exemplo da anotação dentro da categoria metabolismo, o EST correspondente a dihydroorotase.

Tabela 6. Formato de submissão de sequência EST ao dbEST GenBank (NCBI. USA).

TYPE: Pub

MEDUID:

TITLE:

Elucidation of the metabolic fate of glucose in the filamentous fungus *Trichoderma reesei* using EST analysis and cDNA microarrays

AUTHORS:

Chambergo, F. S.; Bonaccorsi, E.D.; Ferreira, A.J.S.; Ramos, A.S.P.; Ferreira Júnior, J.R.; Abrahão-Neto, J.; Farah, J.P.S. & El-Dorry, H.

JOURNAL: Journal of Biological Chemistry (www.jbc.org)

VOLUME: 277

ISSUE: 16

PAGES: 13983-13988

YEAR: 2002

STATUS: 1

||

TYPE: Lib

NAME: TrEST-A

ORGANISM: *Trichoderma reesei* (*Hypocrea jecorina*)

STRAIN: QM9414 (ATCC26921)

SEX: Asexual

TISSUE: Mycelia

STAGE: Mycelia harvested following 18 hs. growth in glycerol as a source carbon

HOST: *E. coli* SOLR cells (kanamycin resistant)

VECTOR: pBluescript SK(+)

V_TYPE: Phagemid

RE_1: *EcoRI*

RE_2: *XhoI*

DESCR:

Cloned unidirectionally, 5' end of the cDNA cloned into *EcoRI* site of pBluescript.

Primer: Oligo (dT). Average insert size: 1,2 kb; Uni-ZAP XR Vector system

-5' adaptor sequence: 5'GAATTTCGGCACGAG3'

-3' adaptor sequence: 5'CTCGAGTTTTTTTTTTTTTTTTTTT3'

||

TYPE: Cont

NAME: El-Dorry, Hamza

FAX: (55) 11-38183848

TEL: (55) 11-38183848

EMAIL: dorry@iq.usp.br

LAB: Department of Biochemistry

INST: Institute of Chemistry. University of São Paulo

ADD: Av. Prof Lineu Prestes, 748, São Paulo, SP, 05508-900, BRASIL

EST FILE

TYPE: EST

STATUS: New

CONT_NAME: El-Dorry, Hamza

CITATION: Elucidation of the metabolic fate of glucose in the filamentous fungus *Trichoderma reesei* using EST analysis and cDNA microarrays

LIBRARY: TrEST-A

EST#: TrEST-A0111

CLONE: Tr-A0111

PCR_F: Universal M13 forward primer

PCR_B: Universal M13 reverse primer

PLATE: 2

ROW: B

COLUMN: 3

SEQ_PRIMER: M13 reverse primer

P_END: 5'

HIQUAL_START: 1

HIQUAL_STOP: 651

DNA_TYPE: cDNA

PUBLIC:

PUT_ID: sorbitol utilization protein sou2

POLYA: N

COMMENT:

SEQUENCE:

CCGCCATCTGCCGTCTCACCAGTCGCTGTCAATCAAGATTCATCACGAAA
TACTCCCCATCCTTTGCATCGCCCATCATGCCTCAGCCTGTCCCCACCGC
CAACAGACTCCTTGATCTCTTCAGCTTGAAGGGCAAGGTCGTCGTCGTCA
CCGGCGCTTCCGGCCCTCGAGGCATGGGAATCGAAGCTGCCCCTGGCTGC
GCCGAGATGGGCGCTGACCTCGCCATCACCTACTCGTCTCGCAAGGAGGG
CGCGGAGAAGAACGCCGAGGAATTGACCAAGGAATACGGCGTCAAAGTCAAGGTGTA
CAAGGTCAACCAGAGCGACTACAACGATGTTGAGCGCTTTGTGAACCAGGTCGTGTCT
GACTTTGGCAAGATCGATGCCTTTATTGCCAACGCCGAGCCACAGCTAATAGCGGAG
TTGTTGACGGCAGCGCCAGCGATTGGGACCATGTCATCCAGGTCGACCTGAGCGGCAC
CGCATACTGCGCAAAGGCTGTTGGCGCGCACTTCAAGAAGCAGGGGCCACGGCTCCCTT
GTCATCACAGCTTCAATGTCCGGCCACGTCGCAAACCTATCCCCAAGAACAGACCTCAT
ACAACGTCGGCCAGGCCGGGTGCATTCATCTGGCGCGGCCTTTGCCAACGAGT

||

Tabela 7. ESTs de *Trichoderma reesei* envolvidos na glicólise, ciclo de TCA, via alternativa do metabolismo de piruvato e gliconeogênese.

NOME ^a	Número de acesso ^b	Proteína ^c	E-value ^d
TrEST-A1305	sp P36580 ALF_SCHPO	Fructose-bisphosphate aldolase	1e-100
TrEST-A4206	sp P04828 TPIS_EMENI	Triosephosphate isomerase	5e-087
TrEST-A2534	sp P17730 G3P2_TRIKO	Glyceraldehyde 3-phosphate dehydrogenase	3e-066
TrEST-A3326	sp P14228 PGK_TRIRE	Phosphoglycerate kinase	8e-039
TrEST-A1895	sp Q12560 ENO_ASPOR	Enolase	0e-000
pki1*	gi 170552	Pyruvate kinase	
TrEST-A2387	sp P20285 ODP2_NEUCR	Dihydrolipoamide acetyltransferase	7e-051
TrEST-A3950	gb AAD16178.1	Pyruvate decarboxylase	3e-064
TrEST-A2339	pir S45858	Aldehyde dehydrogenase1**	3e-064
TrEST-A0264	sp P42041 DHAL_ALTAL	Aldehyde dehydrogenase2**	3e-046
TrEST-A2353	sp P16928 ACSA_EMENI	Acetyl-Coenzyme A synthetase	1e-118
TrEST-A0857	sp P79024 CISY_CANTR	Citrate synthase	5e-037
TrEST-A3335	emb CAB62099.1	Isocitrate dehydrogenase	6e-087
TrEST-A0599	pir T49683	Alpha-ketoglutarate dehydrogenase	1e-060
TrEST-A2899	sp O42772 DHSB_MYCGR	Succinate dehydrogenase	1e-118
TrEST-A1190	ref NP_012838.1	Mitochondrial malate dehydrogenase	3e-062
TrEST-A3061	sp 013434 PPCK_CANAL	Phosphoenolpyruvate carboxykinase	4e-044

a: Um EST singleto é mostrado, quando um contígua tem sido identificado.

b: Código de identificação da proteína com maior similaridade ao EST no GenBank, segundo o BLASTX.

c: Identificação da proteína na base de dados do GenBank (NCBI. USA).

d: Probabilidade esperada, para a similaridade ao acaso pelo BLASTX.

*: Sequência previamente depositada no GenBank (Schindler *et al.*, 1993)

** : Aldehyde dehydrogenase 1 e 2 de *T. reesei*.

proteínas da base de dados do GeneBank. Os valores de “E-value” obtidos nos permite estabelecer que as seqüências obtidas codificariam as proteínas indicadas.

O gene ADH (álcool-desidrogenase) não foi identificado em *T. reesei*, mas a atividade da enzima, que catalisa a conversão reversível entre acetaldeído e etanol usando NADH / NAD⁺, foi encontrada e determinada em extratos de *T. reesei* (Beutler, 1984). A análise da Tabela 8 mostra que *T. reesei* apresenta atividade da enzima ADH em valores que podemos considerar como constitutivos.

O perfil dos UniGenes mostrados na Figura 12 indica ser possível a realização da análise transcricional do metabolismo e a determinação do destino da glicose neste microrganismo.

Por este motivo, para se avaliar o efeito da glicose sobre a transcrição e a expressão global dos UniGenes que constituem a base de dados de EST, foi utilizada a metodologia de DNA “microarray”.

Os insertos de cDNA de 1.151 clones foram amplificados por PCR empregando-se os “primers” SK e T7 cujas seqüências-alvo flanqueiam o inserto no vetor pBluescript SK+. No caso dos contíguos, foi escolhido o clone cuja seqüência encontrava-se no extremo 5’ do consenso assumindo que, embora não necessariamente, esta seqüência fosse aquela que conteria parte da região codificadora do cDNA. Foram também amplificados 23 genes utilizando-se oligonucleotídeos específicos.

O produto obtido da amplificação foi purificado empregando-se filtros MultiScreen-FB da Millipore, sendo suspenso em 6 µL de DMSO 50%, como descrito em Materiais e Métodos. A qualidade e quantidade do DNA foram avaliadas por eletroforese em gel de agarose a 1 %.

A Figura 13 mostra, como exemplo, a eletroforese de 1 µL do produto de PCR purificado. A amplificação de uma banda única e a quantidade de DNA obtido (Fig. 13), garante a qualidade do DNA a ser empregado na preparação das lâminas de “microarray”. Os 5 µL restantes do produto de PCR foram utilizados na confecção de 4 placas de 384 poços e denominadas Brazil 1 a Brazil 4.

Tabela 8. Determinação de atividade ADH em *T. reesei*

Atividade (U/mg. proteína)	Glicose (mmol/L)		
	46	1,2	0
Etanol → Acetaldeído	0,059	0,136	0,210
Acetaldeído → Etanol	0,210	0,343	0,410

Tabela 9. Valores de fluorescência obtidos para os primeiros 20 elementos da lâmina A4

Número	Elemento	Endereço	Cy3		Cy5	
			Background	Intensidade	Background	Intensidade
1	Controle	A1.a1	12488	431768	12304	497808
2	Branco	A1.a2	3448	2112	3016	1552
3	TrEST-A0003	A1.a3	3512	10536	2864	5024
4	TrEST-A0247	A1.a4	3552	2480	2856	5216
5	TrEST-A0016	A1.a5	5288	37808	3504	21264
6	TrEST-A0298	A1.a6	3880	37320	3424	50720
7	TrEST-A0034	A1.a7	3904	20816	3296	15248
8	TrEST-A0319	A1.a8	3992	15032	3008	9792
9	TrEST-A0699	A1.a9	3992	14896	3120	6400
10	TrEST-A1116	A1.a10	7368	226032	4352	101920
11	Branco	A1.b1	3416	1656	2944	5296
12	Branco	A1.b2	3456	2400	2920	752
13	TrEST-A0731	A1.b3	4184	28984	3216	20048
14	TrEST-A1173	A1.b4	6784	189248	4008	79680
15	TrEST-A0760	A1.b5	3712	10304	3000	6824
16	TrEST-A1194	A1.b6	5984	172064	4280	93256
17	TrEST-A3375	A1.b7	18744	715312	5752	150784
18	TrEST-A3680	A1.b8	4000	28608	3016	16800
19	TrEST-A3394	A1.b9	4800	84288	4432	134192
20	TrEST-A3695	A1.b10	6680	164776	4336	85232

Figura 12. Representação esquemática de glicólise, TCA, gliconeogênese e via alternativa de metabolismo de piruvato (Pyruvate bypas) apontando os genes identificados em *T. reesei*. Caixas vermelhas são genes identificados e caixas amarelas são genes ainda não identificados.

Figura 13. Eletroforese em gel de agarose 1% do produto purificado dos PCR dos cDNAs, representando uma das 4 placas de 384 poços utilizadas neste estudo. A e B linhas, 1-12 colunas da placa de 96 poços.

Para possibilitar a análise do efeito da concentração de glicose sobre o nível de expressão gênica dos transcritos obtidos de *T. reesei*, foi desenhado um experimento para se caracterizar a flutuação da expressão dos mesmos durante a cinética de consumo de glicose. Esporos de *T. reesei* foram germinados durante 8 horas em meio completo a 100 mmol/L de glicose e, posteriormente, inoculados em um fermentador contendo 10 L de meio de cultura a 100 mmol/L de glicose. A concentração de glicose foi determinada em diferentes tempos durante o crescimento da cultura e, quando chegou a um valor de 83 mmol/L, uma alíquota de meio foi retirada para a obtenção de massa micelial necessária para preparar o RNA que serviu como referência no experimento. Outras alíquotas foram retiradas ao longo do cultivo para acompanhamento do consumo de glicose, como está mostrado na Figura 14.

O RNA total isolado das amostras indicadas pelas setas verticais (em vermelho na Figura 14), foi empregado para a verificação da variação no nível de expressão dos transcritos nestas populações de RNA, em comparação com o RNA de referência (seta em verde). A qualidade e a quantidade do RNA utilizado nos experimentos foram estimadas, após desnaturação com glioxal/DMSO, pela visualização das amostras em gel desnaturante de agarose a 1%, e por dados de absorção no ultravioleta. (Figura 15).

As placas contendo as sondas de DNA e as amostras de RNA foram enviadas à empresa Genomics Solutions (MI, USA) que construiu as lâminas de “microarray” utilizadas neste trabalho. A partir do mRNA foram sintetizadas, por transcrição reversa, moléculas de fita simples de cDNA em uma reação em que além dos 4 dNTPs foi adicionado aminoalil-dUTP. Após a síntese do cDNA, os fluoróforos Cy3 ou Cy5 foram acoplados ao grupo amino do alil-dUTP.

As moléculas de cDNA provenientes dos RNAs isolados na concentração de 83 mmol/L foram marcadas com o fluoróforo Cy3 (fluorescência verde) e os outros cDNAs marcados com o fluoróforo Cy5 (fluorescência vermelha).

As sondas de DNA - totalizando 1.497 seqüências representando 1.174 genes, dos quais 1.151 foram identificados neste trabalho e 23 selecionados a partir do Genbank (alguns dos genes foram representados mais de uma vez) - foram impressas, em duplicata, em lâmina de vidro num arranjo de 2.994 elementos. Além disso, foram impressas, em cada, lâmina 32 sondas-controle (gene Q do fago lambda) distribuídas uniformemente, totalizando 3.026 elementos por lâmina.

Figura 14. Concentração de glicose e densidade celular durante o crescimento de *T. reesei*. Aliquotas da cultura foram retiradas em diferentes tempos como indicado pelas setas verticais. A seta verde indica o ponto usado como “referencia” e setas vermelhas os pontos analisados.

Figura 15. Eletroforese em gel desnatante de agarose 1%, do RNA extraído das aliquotas retiradas a distintos tempos (ver Fig. 14).

Figura 16. Imagens da fluorescência da lâmina A4. A e B) Cy3 e Cy5 respectivamente, antes da construção da grade; C e D) Imagens depois da construção da grade, para a obtenção de dados; E e F) Imagens ampliadas de um bloco.

A avaliação da expressão gênica acompanhando o consumo de glicose foi determinada através da preparação de 4 lâminas. A hibridação de cada lâmina foi realizada usando a mistura de cDNAs fluorescentes, sintetizados a partir dos RNAs isolados, como indicado a seguir: lâmina A1: 83 mmol/L glicose (Cy3)/46 mmol/L glicose (Cy5); lâmina A2: 83 mmol/L glicose (Cy3)/11 mmol/L glicose (Cy5); lâmina A3: 83 mmol/L glicose (Cy3)/1 mmol/L glicose (Cy5) e lâmina A4: 83 mmol/L glicose (Cy3)/0 mmol/L glicose (Cy5).

Obtenção de dados

A obtenção dos dados foi realizada utilizando o programa GeneTAC Integrator v. 3.0.1. A Figura 16 mostra um exemplo das imagens correspondentes à lâmina A4 (83 mmol/L glicose (Cy3)/0 mmol/L glicose (Cy5)). Como se pode observar, as imagens obtidas correspondem a cada um dos cDNAs fluorescentes hibridados com sondas de DNA (Figura 15A e 15B). Isto é, as imagens A e B correspondem às leituras diferenciais da mesma lâmina, a primeira visualizando a fluorescência promovida pelo cromóforo Cy3 e a segunda pelo Cy5.

Após a hibridação, a fluorescência dos cromóforos foi captada separadamente para cada elemento através de um “scanner”. As imagens captadas da lâmina A4, correspondentes à fluorescência do cromóforo Cy3 e à do cromóforo Cy5, são mostradas na Figura 16A e 16B, respectivamente.

A análise da imagem é iniciada com a construção e sobreposição de uma grade sobre a imagem do “microarray”. A grade serve para localizar os elementos. Um círculo é desenhado ao redor de cada elemento e deve conter todo o sinal possível. A Figura 16 (C e D) mostra, como exemplo, a sobreposição da grade em cada uma das imagens fluorescentes da lâmina A4. Nas Figuras 16E e F são mostradas ampliações de apenas um bloco das Figuras 16C e D, respectivamente, da lâmina A4, como exemplo da grade empregada na extração de dados.

Uma vez definida a grade, o programa realiza a extração automática dos dados de intensidade de todos os elementos, gerando uma tabela com valores correspondentes a cada um deles. O procedimento de análise é repetido para cada imagem fluorescente. A Tabela 9 dá um exemplo dos dados obtidos de cada imagem (Cy3 e Cy5) para os primeiros 20 elementos da lâmina A4, sendo mostrados os valores correspondentes ao background, intensidade de fluorescência de Cy3 e de Cy5, número do elemento e endereço na lâmina.

Figura 17. Imagem composta da lâmina A4. A) Imagem formada pela sobreposição das imagens fluorescentes Cy3 (Figura 16A) e Cy5 (Figura 16B); B) Ampliação de um bloco da lâmina A4.

Na Figura 17 é mostrada a sobreposição das imagens fluorescentes de Cy3 (Fig. 16A) e de Cy5 (Fig. 16B) da lâmina A4. Os genes induzidos aparecem como elementos de cor vermelha, uma vez que a amostra de referência foi marcada com Cy3, os genes reprimidos com a cor verde e os genes expressos em mesmo nível em ambas as condições aparecem como elementos de cor amarela. A Figura 17B mostra a ampliação de um bloco da lâmina 4.

Normalização de dados

A expressão relativa de cada um dos genes em estudo é dada pela razão entre os valores de fluorescência de Cy5 e Cy3. Tomando-se novamente a lâmina A4 como exemplo, podemos dizer que a razão Cy5/Cy3 representa a abundância relativa do mRNA de um dado gene a 0 mmol/L (Cy5) em relação à abundância a 83 mmol/L (Cy3) de glicose. Se este valor for igual a 1, por exemplo, a expressão do gene não é alterada pela variação na concentração de glicose. Se, por outro lado, Cy5/Cy3 for igual a dois, a expressão do gene em questão, a 0 mmol/L de glicose, seria duas vezes maior que a 83 mmol/L.

Para a obtenção destes valores, deve-se determinar, primeiro, a necessidade ou não, da subtração do “background” do valor da intensidade de fluorescência obtido de cada elemento. Para isto, a razão Cy5/Cy3 correspondente a elementos de controle em cada lâmina é obtida com e sem a subtração do valor do “background”. Se o valor do “background” afetar significativamente os valores das intensidades dos elementos ele deve ser subtraído. Em nosso experimento foram inseridos, em cada lâmina, 32 elementos-controle correspondentes ao Q-gene do fago lambda (um para cada bloco de elementos) e no momento da transcrição reversa o mRNA correspondente foi adicionado.

A Figura 18 mostra que a razão Cy5/Cy3 obtida do gene controle, em cada bloco, não varia significativamente com ou sem subtração do “background”. Assim, para a normalização, decidiu-se empregar a razão dos genes sem a subtração do “background”. Na mesma Figura podemos observar que as razões Cy5/Cy3 do gene controle são diferentes dentro de um intervalo de ± 2 SD (desvio padrão), mostrando que a intensidade do sinal dos controles não é homogênea numa mesma lâmina. Portanto, decidiu-se realizar a normalização empregando-se a razão do controle (RC) de cada bloco na lâmina (normalização por controle local). Estas diferenças são devidas a diversos fatores, entre os quais podemos mencionar as propriedades físicas específicas de cada fluoróforo, a

Descriptive Statistics

	Média	Std. Dev.	Std. Error	Count	Minimum	Maximum	# Missing
RC-Bg	1.968	.543	.096	32	1.205	3.187	1
RC	1.922	.483	.085	32	1.220	3.019	1

Descriptive Statistics

	Média	Std. Dev.	Std. Error	Count	Minimum	Maximum	# Missing
RC-Bg	1.188	.451	.080	32	.685	2.288	1
RC	1.178	.415	.073	32	.705	2.226	1

Descriptive Statistics

	Média	Std. Dev.	Std. Error	Count	Minimum	Maximum	# Missing
RC-Bg	1.368	.486	.086	32	.754	2.467	1
RC	1.359	.457	.081	32	.775	2.355	1

Descriptive Statistics

	Média	Std. Dev.	Std. Error	Count	Minimum	Maximum	# Missing
RC-Bg	1.125	.349	.062	32	.657	1.904	1
RC	1.126	.332	.059	32	.675	1.871	1

Figura 18. Análise estatístico da razão correspondente aos 32 elementos do gene controle Q-fago lambda, depositados na lâmina A1 (A), A2 (B), A3 (C) e A4 (D). No gráfico são mostradas as razões de cada elemento controle com subtração do background (RC-Bg, em vermelho) e sem subtração de background (RC). Em cada figura linhas correspondendo à média e a dois desvios padrão (SD) são apresentadas.

eficiência de incorporação, a variabilidade experimental no processamento, a impressão dos elementos de DNA e as características do “scanner” (Yang *et al.*, 1998).

Obtida a razão normalizada de cada elemento (RNE) como descrito em Materiais e Métodos, a segunda etapa da análise consistiu em determinar a relação entre as duplicatas. Considerando que as duplicatas de um mesmo elemento se encontram localizadas no mesmo bloco de uma lâmina (Fig. 14B) e visto que geralmente a concentração de glicose maior que 4 a 5 mmol/L reprime a expressão de genes controlados por glicose (Abrahão-Neto *et al.*, 1995; Flikweert *et al.*, 1997), estabelecemos um critério de análise, segundo o qual deve existir uma relação entre os elementos nas lâminas A1 e A2 (46 mmol/L e 11 mmol/L, representando o estado reprimido), bem como entre os elementos das lâminas A3 e A4 (1 mmol/L e 0 mmol/L, representando o estado em que não há repressão por glicose). Observando-se estes pontos, foi desenvolvido o critério para a tomada de decisão que está mostrado em Materiais e Métodos. Um exemplo do processo de normalização será mostrado com os dados obtidos mais adiante. Após esta análise foram descartados 117 elementos (7,8%) da base de dados, sendo aceitos 1380 elementos (92,2%). Para simplificação e melhor entendimento do texto, a partir deste ponto à denominação de elemento será trocada por gene.

Agrupamento hierárquico (“Hierarchical Clustering”) para identificação dos genes regulados por glicose

Um dos objetivos deste trabalho foi a elucidação, em nível transcricional, da expressão do conjunto de genes identificados e, especificamente, dos genes envolvidos no metabolismo e destino da glicose em *T. reesei*. Com este propósito, analisamos a expressão global dos genes durante o consumo de glicose. Como foi indicado anteriormente, os RNAs obtidos com 46 mmol/L, 11mmol/L, 1 mmol/L e 0 mmol/L de glicose foram comparados, um a um, com o RNA obtido a 83 mmol/L de glicose.

Nesta abordagem, foi estabelecido que os genes cujos dados na lâmina A3 e/ou A4 flutuassem com fator de expressão diferencial $\geq 2,0$ ou $\leq 2,0$ com relação aos respectivos dados na lâmina A1 e/ou A2, seriam considerados genes diferencialmente expressos. Por este critério foram identificados 298 genes (21,6 %) com expressão diferenciada.

Para a obtenção do padrão global de expressão gênica, os genes devem ser agrupados para possibilitar a identificação daqueles com padrão de expressão similar (Eisen *et al.*, 1998). A metodologia de agrupamento hierárquico, descrita por Eisen *et al.*, (1998), foi empregada para agrupar os dados obtidos segundo o perfil de expressão dos

genes. Os 298 genes selecionados foram agrupados hierarquicamente empregando-se o Programa Gene Cluster (com parâmetros standard), sendo o resultado visualizado com o programa TreeView.

A similaridade no padrão de expressão entre os genes é representada por uma árvore na qual o comprimento dos ramos reflete o nível de similaridade entre os genes. Para isto os valores de expressão dos genes são transformados em seu logaritmo na base 2 (para simplificação), para tratar a indução ou a repressão com magnitudes idênticas como numericamente iguais mas com sinais opostos. O algoritmo de agrupamento hierárquico calcula um dendrograma que agrupa todos os genes numa árvore, mostrada na Figura 19A.

Considerando, por exemplo, o nível de expressão determinado com os valores da lâmina A4, em que são comparados os RNA obtidos a 83 mmol/L de glicose (Cy3) e o RNA obtido a 0 mmol/L de glicose (Cy5), os valores dos genes cujo logaritmo da razão Cy5/Cy3 na base 2 é 0 (expressão do gene constante) são representados em preto; razões com logaritmo positivo são representadas em vermelho, razões com logaritmo negativo são representadas em verde. Logarítmos positivos indicam que quando a concentração de glicose é 0 mmol/L a abundância deste transcrito é maior que na condição de 83 mmol/L. Assim, os genes representados com a cor vermelha, quando a concentração de glicose é 0 mmol/L, indica que o transcrito é mais abundante em baixa concentração de glicose. De forma oposta, a representação de um gene pela cor verde significa que em 83 mmol/L de glicose o transcrito deste é mais abundante.

O perfil da expressão gênica para cada um dos 298 genes é mostrado em uma figura de agrupamento hierárquico (Fig. 19A). Avaliando-se a Figura 19A, e posto que um dos nossos objetivos era a identificação dos genes cujo nível de transcrição aumenta quando a glicose esgota-se no meio, foram selecionados os genes agrupados no ramo inferior da árvore cuja cor na lâmina A3 e A4 era vermelha e que estão detalhados na Figura 19B.

Uma diferença muito clara ao simples olhar é que o único gene do ciclo de TCA que aparece na lista de genes identificados foi o que codifica a enzima α -cetoglutarato-desidrogenase (KDH). Este fato chamou nossa atenção para uma análise mais detalhada. Para verificar e validar estes resultados foi feita análise por “Northern blot” de 18 genes que eram de nosso interesse.

Como mostrado na Figura 19C, dos 18 genes avaliados 13 foram confirmados e estão apresentados na Figura 20 com o resultado do “Northern blot” em 83 mmol/L (referência) e quando a glicose esgota-se no meio (0 mmol/L), validando os resultados

Figura 19. **A**- Agrupamento hierárquico de genes de *T. reesei* com expressão diferencial de razão direta ou inversa igual ou superior a dois. **B**- Genes induzidos em baixa concentração de glicose e **C**- Genes selecionados para validação; ✓ = confirmado por Northern blot / x = não confirmado. As colunas verticais em A e B representam o nível de expressão dos genes em uma concentração de glicose específica (46 mmol/L, 11 mmol/L, 1 mmol/L e 0 mmol/L) em relação a amostra de referência (83 mmol/L).

Figura 20. Perfil de expressão de genes reprimidos por glicose. O perfil de expressão de 14 genes cuja expressão varia mais de duas vezes nas condições analisadas. A esquerda Northern blot dos transcritos. A direita representação da fluorescência obtida nos “microarrays”. O nível de expressão é representado pela escala inferior de cor.

obtidos através do “microarray”. Citrato-sintase (CIT), apesar de não se encontrar no agrupamento mostrado na Figura 19B, foi incluído no agrupamento mostrado na Figura 20.

Para possibilitar a observação da variação no nível de expressão, na Figura 21 estão mostrados cada um dos 14 genes (os 13 genes validados por “Northern blot” mais o gene CIT) e sua duplicata em cada uma das lâminas. Podemos observar que todos os genes apresentam variação na intensidade e na cor da fluorescência proporcional aos dados obtidos no “microarray” para cada um deles. Os dados de intensidade de fluorescência de cada cromóforo, valor do background, razão Cy5/Cy3 de cada elemento (RE) e do controle local (RC) e a razão normalizada do elemento (RNE) de cada um dos 14 genes e suas duplicatas, para a lâmina A4, são mostrados na Tabela 10. Na Tabela 11 são mostrados os dados para a determinação do valor da RNE desses genes em cada lâmina a ser empregado no processo de agrupamento hierárquico.

Análise de expressão dos genes envolvidos no metabolismo de Glicose

Analisando os dados obtidos pelo “microarray” (Figura 19) foi verificado que os genes malato-desidrogenase (MDH), isocitrato-desidrogenase (IDH), e succinato-desidrogenase (SDH) não estão presentes, indicando que a abundância dos transcritos não é alterada pela concentração de glicose no meio, enquanto que os genes citrato-sintase (CIT) e α -cetoglutarato desidrogenase (KDH) têm sua expressão parcialmente reprimida por glicose, isto é, em alta concentração de glicose (83 mmol/L) o nível de transcrito é suficientemente abundante para manter a atividade do ciclo de TCA. O gene CIT, embora apresente expressão diferencial de 2 vezes, foi agrupado pelo programa de análise em um sub-grupo distinto do mostrado na Figura 19B. Entretanto, a análise por “Northern blot” mostra que ele é parcialmente reprimido por glicose, motivo pelo qual o agrupamos aos demais genes na Figura 20.

Como indicado em Objetivos, pretendíamos comparar estes dados com os obtidos com *S. cerevisiae*, já que alta concentração de glicose reprime fortemente os transcritos do ciclo de TCA nesse organismo (DeRisi *et al.*, 1997). Os resultados obtidos com “microarray” e descritos anteriormente mostram que, provavelmente, o ciclo de TCA em *T. reesei*, ao contrário de *S. cerevisiae*, é metabolicamente funcional em alta concentração de glicose. Este fato levou-nos a analisar a expressão dos genes envolvidos no metabolismo e destino da glicose.

TrEST-A3354 Prohibitin (H3.c3/H3.h1)

TrEST-A0857 Citrate synthase (C2.d5/C2.i3)

TrEST-4050 Glucoamylase (G2.d9/G2.i7)

TrEST-A0264 Aldehyde dehydrogenase2 (A1.d7/A1.i5)

TrEST-A0793 Peptide transporter ptr2-a (B2.d2/B2.h10)

TrEST-A1703 Utr2 (H3.d2/H3.h10)

TrEST-A0599 α -ketoglutarate dehydrogenase A3.d9/A3.i7

Figura 21. Imagem nas lâminas A1, A2, A3 e A4, dos 14 genes selecionados na Figura 20. Cada quadrado aponta uma das duplicatas em cada bloco.

Tabela 10. Normalização de dados de fluorescência da lâmina A4, correspondente aos 14 genes selecionados na figura 20.

Elemento	Endereço	Cy3		Cy5		RE	RC	RNE
		Intensidade	Background	Intensidade	Background			
TrEST-A0090	A4.d8	50944	5000	239985	5064	4.710763213	1.809885932	2.602795754
TrEST-A0090	A4.i6	54672	5232	278776	6040	5.099063506	1.809885932	2.81733971
TrEST-A0111	C3.d6	81968	4672	1071256	17368	13.06919774	1.2548318	10.41509924
TrEST-A0111	C3.i4	82856	5112	750256	33184	9.054938689	1.2548318	7.21605771
TrEST-A0264	A1.d7	44208	4320	606400	8864	13.7169743	1.152952512	11.8972587
TrEST-A0264	A1.i5	38848	4336	483336	13144	12.44172158	1.152952512	10.79118303
TrEST-A0595	B3.d1	27792	3792	355024	6600	12.77432355	1.1574338	11.03676387
TrEST-A0595	B3.h9	20672	3760	278720	5736	13.48297214	1.1574338	11.64902229
TrEST-A0599	A3.d9	141000	6488	591042	9688	4.19179239	1.643438551	2.550623888
TrEST-A0599	A3.i7	208120	7448	817406	10680	3.92757281	1.643438551	2.38985082
TrEST-A0793	B2.d2	39944	4392	368256	11040	9.21930703	1.204457992	7.65432011
TrEST-A0793	B2.h10	46008	4296	517104	12488	11.23943662	1.204457992	9.331530609
TrEST-A0857	C2.d5	42576	3792	85328	4408	2.004133784	1.157747327	1.731063193
TrEST-A0857	C2.i3	83960	4792	125648	6696	1.496522153	1.157747327	1.292615512
TrEST-A1703	H3.d2	313640	10920	858104	21912	2.735952047	1.045078657	2.617938879
TrEST-A1703	H3.h10	305976	9144	805488	18248	2.632520198	1.045078657	2.518968482
TrEST-A2353	E2.d7	58440	4104	185152	6512	3.168240931	0.675370569	4.691114889
TrEST-A2353	E2.i5	44384	4840	160576	8184	3.617880317	0.675370569	5.356881814
TrEST-A3037	F4.c10	215648	7744	659184	14272	3.056759163	1.102540443	2.772468966
TrEST-A3037	F4.h8	238464	6688	666912	12176	2.796698873	1.102540443	2.53659527
TrEST-A3354	H3.c3	98624	5568	252877	5528	2.564059627	1.045078657	2.453461107
TrEST-A3354	H3.h1	57144	5736	115904	6144	2.028279434	1.045078657	1.940791175
TrEST-A3388	G1.d5	53360	4624	121016	5624	2.267916042	0.6915526	3.279455592
TrEST-A3388	G1.i3	40464	5424	103200	6624	2.550415184	0.6915526	3.687955455
TrEST-A3935	D4.b8	151008	6640	1013576	13656	6.712068235	1.363019429	4.924411267
TrEST-A3935	D4.g6	169616	6904	941760	16384	5.552306386	1.363019429	4.073534292
TrEST-A4050	G2.d9	19192	4008	37616	3968	1.959983326	1.024019865	1.914009086
TrEST-A4050	G2.i7	9520	4000	29712	3984	3.121008403	1.024019865	3.047800642

RE: Razão do elemento (Cy5/Cy3); RC: Razão do controle (Cy5/Cy3); RNE: Razão Normalizada do Elemento (RE/RC).

Tabela 11. Valores normalizados correspondentes aos 14 genes selecionados na figura 20

A.- Dados brutos

Elemento	RNE ₁	RNE ₂	Lâmina-A1	RNE ₁	RNE ₂	Lâmina-A2	RNE ₁	RNE ₂	Lâmina-A3	RNE ₁	RNE ₂	Lâmina-A4
TrEST-A0090	1.157213	1.507467	1.33234	1.050986	0.924078	0.987532	2.222192	2.334969	2.278581	2.602796	2.81734	2.710068
TrEST-A0111	1.518612	0.676735	1.518612	1.875977	2.161231	2.018604	2.917529	2.54692	2.732224	10.4151	7.216058	8.815578
TrEST-A0264	0.623965	0.455724	0.539845	0.424939	0.48515	0.455045	17.67436	13.95725	15.81581	11.89726	10.79118	11.34422
TrEST-A0595	1.735411	1.112446	1.423928	1.357722	2.095902	1.726812	10.40327	11.83661	11.11994	11.03676	11.64902	11.34289
TrEST-A0599	0.866852	0.703593	0.785223	1.192083	1.30312	1.247601	1.742281	1.808993	1.775637	2.550624	2.389851	2.470237
TrEST-A0793	0.43626	0.448584	0.442422	0.513099	0.432237	0.472668	4.048135	4.601696	4.324916	7.65432	9.331531	8.492925
TrEST-A0857	0.752798	0.598456	0.675627	0.764598	0.465601	0.6151	0.988892	0.394489	0.988892	1.731063	1.292616	1.511839
TrEST-A1703	0.497054	0.642735	0.569894	1.225728	1.284716	1.255222	1.740861	1.81179	1.776326	2.617939	2.518968	2.568454
TrEST-A2353	0.908381	1.100189	1.004285	0.596656	0.873779	0.735217	3.0748	2.914269	2.994534	4.691115	5.356882	5.023998
TrEST-A3037	1.205277	1.036071	1.120674	2.046265	0.875602	0.875602	1.502546	1.487461	1.495003	2.772469	2.536595	2.654532
TrEST-A3354	0.640264	0.690403	0.665334	0.944252	1.315336	1.129794	1.797763	2.070932	1.934347	2.453461	1.940791	2.197126
TrEST-A3388	1.365475	1.244004	1.304739	3.141937	1.139059	1.139059	2.27793	2.561784	2.419857	3.279456	3.687955	3.483706
TrEST-A3935	0.293213	0.301477	0.297345	0.842094	1.436473	1.139284	6.283033	5.961117	6.122075	4.924411	4.073534	4.498973
TrEST-A4050	0.514251	0.603013	0.558632	0.466875	0.722616	0.594746	5.311202	5.809819	5.560511	1.914009	3.047801	2.480905

B.- Dados consolidados

Gene	Lâmina-A1	Lâmina-A2	Lâmina-A3	Lâmina-A4
TrEST-A0090 6-phosphogluconate dehydrogenase	1.33234	0.987532	2.278581	2.710068
TrEST-A0111 sorbitol utilization protein sou2	1.518612	2.018604	2.732224	8.815578
TrEST-A0264 aldehyde dehydrogenase-2	0.539845	0.455045	15.81581	11.34422
TrEST-A0595 subtilisin-like serine protease	1.423928	1.726812	11.11994	11.34289
TrEST-A0599 alfa-ketoglutarate dehydrogenase	0.785223	1.247601	1.775637	2.470237
TrEST-A0793 peptide transporter ptr2-a	0.442422	0.472668	4.324916	8.492925
TrEST-A0857 citrate synthase	0.675627	0.6151	0.988892	1.511839
TrEST-A1703 utr 2	0.569894	1.255222	1.776326	2.568454
TrEST-A2353 acetyl-coenzyme a synthetase	1.004285	0.735217	2.994534	5.023998
TrEST-A3037 high affinity copper transporter	1.120674	0.875602	1.495003	2.654532
TrEST-A3354 prohibitin	0.665334	1.129794	1.934347	2.197126
TrEST-A3388 glucose-6-phosphate 1-dehydrogenase	1.304739	1.139059	2.419857	3.483706
TrEST-A3935 NM	0.297345	1.139284	6.122075	4.498973
TrEST-A4050 glucoamylase	0.558632	0.594746	5.560511	2.480905

Na Figura 22 é mostrado o efeito da concentração de glicose sobre a expressão destes genes comparando-o com os resultados obtidos com *S. cerevisiae* por DeRisi *et al* (1997). A expressão dos genes envolvidos na conversão de glicose a piruvato nos dois microrganismos é similar, indicando que, em alta concentração de glicose, a via glicolítica está induzida, facilitando o fluxo de glicose a piruvato. Isto pode ser visto através do nível de expressão do transcrito de enolase (ENO, Figura 23).

Em levedura, o gene piruvato Descarboxilase (PDC) é ativado em alta concentração de glicose levando piruvato a acetaldeído. Este é o único caminho disponível uma vez que os transcritos do ciclo de TCA encontram-se fortemente reprimidos, conforme foi mostrado na Figura 22B. Posteriormente o acetaldeído é reduzido a etanol pela enzima álcool-desidrogenase (ADH), com oxidação de NADH a NAD⁺, necessário para a continuidade da atividade glicolítica. Esta é a única via disponível nessa condição devido à forte repressão da acetaldeído-desidrogenase (ALD) por glicose. Em *T. reesei* (Figura 22A), o gene PDC é induzido por glicose, entretanto o ciclo de TCA, como foi mencionado anteriormente, encontra-se metabolicamente funcional. Desta forma, o piruvato tem duas alternativas:

- 1-Ser oxidado no ciclo de TCA, para produzir, como produto final, CO₂ e H₂O, ou
- 2-Ser reduzido a acetaldeído.

Dois transcritos correspondentes à enzima acetaldeído-desidrogenase (ALD1 e ALD2) foram identificados em nosso banco de dados. A expressão de ALD1 não é afetada por glicose, enquanto que a expressão de ALD2 é fortemente reprimida. Se as duas enzimas apresentam a mesma afinidade pelo acetaldeído e sua regulação ocorresse somente em nível transcricional, o acetaldeído seria oxidado com a formação de acetato, ao invés de ser reduzido a etanol.

Analisando os dados obtidos e assumindo que a regulação só ocorre no nível transcricional, podemos afirmar que, em *T. reesei*, em alta concentração de glicose o piruvato pode ser oxidado totalmente pelo ciclo de TCA e/ou reduzido a acetato, ao invés de ser convertido a etanol como em *S. cerevisiae*.

Em *S. cerevisiae*, após o esgotamento da glicose a expressão dos genes aldeído-desidrogenase (ALD), acetil-CoA sintase (ACS) e as enzimas do ciclo de TCA são altamente ativadas. Também é altamente ativada a enzima PCK, enzima esta necessária para a entrada de intermediários do ciclo de TCA, via oxaloacetato, na gliconeogênese. A respiração é favorecida devido à ativação desse conjunto enzimático e a necessidade de glicose é satisfeita através da gliconeogênese. Comparando-se este perfil metabólico

A) *Trichoderma reesei*

B) *Saccharomyces cerevisiae*

Fig. 22. Comparação do perfil de expressão gênica de genes da glicólise, gliconeogênese, ciclo do TCA e via alternativa de piruvato (piruvate bypass) durante o esgotamento de glicose em *T. reesei* e *S. cerevisiae*. Caixas vermelhas e verdes representam genes cuja expressão aumenta ou diminui, respectivamente, em função do esgotamento de glicose no meio. Caixas brancas indicam que a expressão desses genes não altera. Caixas amarelas representam genes que não foram isolados ainda. O nível de transcrição de PCK foi determinado por Northern blot.

Figura 23. Perfil da expressão gênica, analisada por Northern blot, de 11 genes selecionados para validação dos resultados obtidos por “microarray”.

fortemente regulado durante a “diauxic shift” em *S. cerevisiae* com o perfil da expressão dos genes ortólogos em *T. reesei*, fica claro que a “diauxic shift” não ocorre já que o ciclo de TCA é metabolicamente funcional em alta concentração de glicose. A esta altura, é importante notar que tanto CIT como KDH são parcialmente reprimidos mas, provavelmente, há um nível de expressão suficiente para que o fluxo de piruvato favoreça a oxidação do piruvato pelo ciclo de TCA. Também, deve-se notar que durante o esgotamento de glicose o ciclo de TCA é mais ativo, devido à necessidade de obtenção de glicose a partir de precursores não glicosídicos. Isto é evidente pela forte expressão de ALD2, ACS, e PCK e pela maior expressão de CIT e KDH. Deste modo, podemos afirmar que a evolução afetou fortemente o destino do piruvato na presença de glicose, mas o controle das vias permaneceu na ausência deste açúcar uma vez que sua biossíntese, através de precursores não glicídicos, se torna indispensável para que vários processos vitais para a célula sejam preservados.

Os dados obtidos pela análise de “microarray”, mostrados na Figura 21A, foram validados por “Northern blot” dos RNAs isolados em glicose 83 mmol/L e em glicose 0 mmol/L, empregando-se como sondas os cDNAs de 11 genes envolvidos no metabolismo de glicose, mostrados na Figura 23. O perfil de expressão gênica é idêntico; entretanto, notam-se diferenças nos níveis de indução ou repressão obtidos pelas duas técnicas.

Discussão

Biblioteca de cDNA

O início do seqüenciamento do genoma humano motivou a formação de duas correntes de opinião entre os cientistas. A primeira sustentava que o seqüenciamento total do genoma permitiria obter as regiões gênicas, intergênicas (reguladoras e de controle) e íntrons, além de não necessitar de RNA de diversos tipos de células e de etapas distintas de desenvolvimento do organismo. A segunda justificava o seqüenciamento do cDNA, reversamente transcrito a partir do mRNA, considerando que as regiões codificadoras dos genes expressos representam o que há de mais importante no genoma e constituem somente 3% do DNA total.

Adams *et al.* (1991) demonstraram que o seqüenciamento de clones obtidos aleatoriamente de uma biblioteca de cDNA, contendo entre 150 e 400 nucleotídeos, proporciona informação suficiente para identificar e localizar o cDNA no genoma. As seqüências assim obtidas foram denominadas ESTs (“Expressed Sequence Tags”).

O seqüenciamento do genoma de um organismo precisa de uma ampla rede integrada de laboratórios com recursos suficientes que permitam cumprir os objetivos propostos. Por isso um projeto de obtenção de ESTs é uma boa opção para laboratórios de pesquisa interessados em identificar os genes expressos por um organismo.

Os genes expressos em uma determinada condição fisiológica, meio ambiental ou de desenvolvimento de uma célula ou tecido de qualquer organismo determinam a maquinaria molecular disponível para realizar a função biológica. A identificação dos genes expressos e a determinação do seu nível de expressão proporcionam um dos mais importantes indicadores para se caracterizar o estado fisiológico de um organismo (Meier-Ewert *et al.*, 1998).

Uma biblioteca de cDNA representa o mRNA da célula em estudo e pode ser normalizada ou não normalizada. Em uma biblioteca não normalizada a frequência dos transcritos representa o nível dos transcritos na célula enquanto que em uma biblioteca normalizada os transcritos abundantes e raros são representados na mesma frequência (Soares *et al.*, 1994). Por essa razão as bibliotecas normalizadas não são empregadas na determinação do nível de expressão gênica.

O seqüenciamento de uma biblioteca de cDNA não normalizada permite-nos analisar a expressão gênica, uma vez que são obtidas etiquetas que representam os genes expressos (ESTs), sendo as frequências destas etiquetas proporcionais à abundância dos

transcritos. Isso permite a identificação dos genes altamente expressos e a comparação da expressão gênica entre diferentes células, tecidos e estados fisiológicos (Carulli *et al.*, 1998).

A biblioteca não normalizada de cDNA de *T. reesei* foi preparada a partir de RNA obtido de micélio crescido em glicerol. O glicerol é uma fonte de carbono não fermentável que para ser metabolizado precisa de uma mitocôndria funcional (Hondmann *et al.*, 1991). Desta forma esperamos obter ESTs correspondentes a genes nucleares e mitocondriais.

O êxito de um projeto de seqüenciamento depende da qualidade da biblioteca de cDNA da qual as seqüências de EST são obtidas. O seqüenciamento em maior escala de uma biblioteca de cDNA (Adams *et al.*, 1995) é feito quando a biblioteca é:

- 1- Representativa, contendo todas as seqüências presentes na população inicial de poli(A)⁺ RNA, na mesma freqüência relativa;
- 2- Clonada unidirecionalmente, conhecendo-se a orientação do cDNA;
- 3- Constituída de alta porcentagem de transcritos totais;
- 4- Não contaminada.

A biblioteca de cDNA de *T. reesei* foi construída empregando-se oligonucleotídeos sintéticos oligo-d(T), clonada unidirecionalmente e seqüenciada a partir do extremo 5', que provavelmente contém a seqüência codificadora de uma proteína.

Considerando que uma biblioteca de cDNA é de boa qualidade quando mostra <2% de vetor (sem inserto), <4% de seqüência de organismos contaminantes, >50% de seqüências sem similaridade na base de dados, >50% de seqüências >300 bases (Adams *et al.*, 1995; Hillier *et al.*, 1996). A análise dos dados preliminares do seqüenciamento (Tabela 1) nos permitiu afirmar que a biblioteca de cDNA de *T. reesei* era de boa qualidade.

Base de dados de EST de *T. reesei*

Os genomas de vários organismos encontram-se em pleno seqüenciamento e só 12 projetos correspondem a fungos. Embora o seqüenciamento da maioria dos genomas de fungos seja financiado por entidades públicas, alguns são financiados pelo setor privado que restringe o acesso livre às seqüências, como é o caso do Projeto Genoma de *Aspergillus nidulans*, que apresenta acesso restrito no “site” da Cereon/Monsanto (www.cereon.com) (Yoder *et al.*, 2001), e o de *Magnaporthe grisea*, no site de Paradigm Genetics (www.paragen.com) (Hamer *et al.*, 2001). Por esta razão o número de genes de fungos depositados na base pública de dados GenBank ainda é pequeno.

O presente trabalho é o primeiro a identificar os genes do fungo filamentoso *T. reesei*, através da análise de seqüências de ESTs. *T. reesei* é um microrganismo de importância biotecnológica e é de nosso conhecimento que diversos centros de pesquisa, associados a empresas, encontram-se conduzindo projetos de seqüenciamento deste organismo em larga escala (www.fungalgenomics.ncsu.edu).

Assume-se que o tamanho médio de um gene é cerca de 2 kb e que os genes adjacentes são separados por 1000 bases de seqüência não-codificadora. Se o genoma de *T. reesei* tiver 39 Mb, podemos esperar um número aproximado de 13.000 genes. Se a média da unidade genética for maior, ou cerca de 4 kb, podem ser codificados aproximadamente 10.000 genes. De acordo com esses cálculos, o presente trabalho identificou ao redor de 10% dos genes expressos deste fungo. A estimativa do número de genes é similar à empregada para outros eucariotos de diferente complexidade, como *Drosophila melanogaster*, *Caenorhabditis elegans* (Nelson *et al.*, 1997) e fungos filamentosos (Kupfer *et al.*, 1997). Até o momento, só 109 seqüências de DNA de *T. reesei* encontram-se depositadas na base de dados. Os resultados apresentados neste trabalho representam um incremento de mais de 10 vezes no número de genes expressos identificados em *T. reesei*.

A informação mais importante obtida a partir de um projeto de seqüenciamento é a anotação funcional do produto do gene. O emprego de uma medida (“E-value” ou nota) é o método mais empregado para estabelecer o grau de similaridade entre duas seqüências (Wilson *et al.*, 2000). Uma nota para um alinhamento, proporcionada pelo programa BLASTX, maior ou igual a 80 é considerada confiável (Pearson, 1991; Lluisma *et al.*, 1997) para ser empregada na indicação de uma identidade putativa, ou função, a uma seqüência desconhecida.

A comparação dos 1.151 genes únicos de *T. reesei* com a base pública de dados do GenBank mostra que foi obtido um número significativo de genes envolvidos em distintas atividades celulares, como mostrado na Figura 10 e na base de dados de EST.

O número de ESTs sem similaridade (61%) com a base de dados é consequência da ausência de genoma de qualquer fungo filamentoso completamente seqüenciado. Tais ESTs se constituem em novos genes. Estudo para identificar as proteínas associadas com a parede celular por espectrometria de massa e análise da seqüência de aminoácidos mostra que 37% das proteínas isoladas de *T. reesei* não apresentam similaridade com seqüências na base pública de dados (Lim *et al.*, 2001). No Projeto Genoma de *Neurospora crassa* (Nelson *et al.*, 1997) foram analisadas 994 ESTs únicas e aproximadamente 56% das ESTs não apresentaram similaridade com seqüências na base de dados. No mesmo projeto, a

porcentagem de genes envolvidos no metabolismo (43,3%), síntese de proteínas (32,9%) e síntese de RNA (8,4%) foi similar aos obtidos para *T. reesei*.

A análise das ESTs nos permitiu postular que o perfil de expressão gênica obtido é complexo, com um pequeno número de transcritos altamente expressos e um grande número de transcritos de média e baixa abundância. Entre os 8 transcritos de maior ocorrência (≥ 30 ocorrências), mostrados na Tabela 5, encontram-se 5 genes não identificados. Entre os genes identificados é importante mencionar a Gliceraldeído-3-fosfato-desidrogenase como o transcrito de maior abundância (85 *ocorrências*), igual ao que ocorre em *S. cerevisiae* (Velculescu *et al.*, 1997). Esta enzima, localizada na parede celular (Lim *et al.*, 2001), apresenta similaridade com Gliceraldeído-3-fosfato-desidrogenase identificada em *T. konigii* (Sakai *et al.*, 1990) e catalisa a primeira etapa do metabolismo de moléculas de 3 átomos de carbono, que leva o gliceraldeído-3-fosfato a piruvato, sendo esta parte da via glicolítica considerada a mais antiga (Fukuchi *et al.*, 1992).

A administração adequada dos dados das seqüências obtidas é um aspecto muito importante num projeto de EST. O estabelecimento de uma base de dados de EST (dbEST) permite administrar o grande volume de seqüências produzidas. Por esta razão, durante o desenvolvimento deste trabalho, foi construída uma base de dados de EST de *T. reesei* (dbEST *T. reesei*) apresentando os genes identificados em *T. reesei*, estando disponível em uma página de Internet de acesso público (<http://trichoderma.iq.usp.br>).

As limitações de uma dbEST devem-se ao fato que ela representa um conjunto de fragmentos de genes onde existe a possibilidade de que o tamanho do inserto e a determinação das bases possam ser afetados pela errônea leitura dos perfís eletroforéticos. Erros em bases são observados inclusive em seqüências consideradas de alta qualidade. Os erros podem ocorrer por substituições (bases desiguais, incluindo as bases ambíguas Ns), deleções (a base falta na EST e está presente no mRNA), ou inserções (a base está presente na EST mas não no mRNA) (Hillier *et al.*, 1996). A obtenção de um alto número de ESTs com mais de 300 nucleotídeos garante uma boa análise de similaridade, como mostrado por Shoop *et al.* (1994) que determinaram que ESTs de 400 bases produzem um resultado de similaridade igual à ESTs de 300 bases. O número mínimo de bases requerido para submissão de ESTs a um banco de dados é de 150, sendo em nossa análise estabelecido que estas tivessem qualidade com um índice $Q \geq 20$. Isto, no momento da análise de similaridade, permitiria obter no mínimo 50 aminoácidos para realizar a comparação com outras seqüências (Adams *et al.*, 1991).

As ESTs representam aproximadamente 40% das seqüências nucleotídicas no GenBank (NCBI, USA) e encontram-se depositadas na dbEST 10.299.149 seqüências de ESTs (liberação em 01/02/2002). O TIGR Gene Index (www.tigr.org/tdb) contém as seqüências de ESTs de humano e de 10 importantes organismos-modelo (rato, camundongo, plantas e outros), tentando estabelecer seqüências consenso de contíguos de ESTs (Quackenbush *et al.*, 2000). Considerando que as ESTs obtidas representam seqüências parciais de genes transcritos, a dbEST proporciona uma base para um “Índice Gênico” compreensivo de um organismo (Gerhold *et al.*, 1996).

Na base de dados GenBank só 109 seqüências de nucleotídeos de *T. reesei* encontram-se depositadas. Dos genes já identificados, só 5 apresentam similaridade com genes na dbEST *T. reesei* e são: hydrophobin II precursor (HFBII) (sp|P79073|HYP2_TRIRE), elongation factor 1-alpha (ef-1-alpha) (sp|P34825|EF1A_TRIRE), orotidine 5-phosphate decarboxylase (sp|P21594|DCOP_TRI RE), phosphoglycerate kinase (PGK) (sp|P14228|PGK_TRIRE) e pyruvate kinase (PK) (gi|170552|gb|L07060.1|TRRPKI[170552]). Destes, só PGK (Vanhanen *et al.*, 1989) e PK (Schindler *et al.*, 1993) correspondem a genes do metabolismo de carboidratos.

O estabelecimento da base de dados das ESTs de *T. reesei* nos permitiu identificar genes envolvidos em vias metabólicas ou processos celulares específicos. Considerando que a biblioteca de cDNA foi construída a partir do RNA obtido de uma cultura crescida em glicerol, foram identificados os genes que codificam as enzimas glicerol-quinase, glicerol-3-fosfato desidrogenase e triose-fosfato isomerase, do catabolismo do glicerol. Dos 10 genes que codificam enzimas da via glicolítica foram identificados 5 e dos 8 genes que codificam enzimas do ciclo de TCA foram identificados outros 5, além dos genes PCK (piruvato-carboxiquinase) e PDA (piruvato-desidrogenase) (Fig. 12) e o gene da cadeia respiratória COX1. Um fato importante a ressaltar é a identificação de 3 genes de enzimas do “pyruvate bypass” (PDC, ALD1, ALD2, ACS) e 4 genes de enzimas da via das pentoses-fosfato (glicose-6-fosfato desidrogenase, 6-fosfogliconato desidrogenase, transcetolase e transaldolase). No metabolismo de aminoácidos podemos mencionar, como exemplo, a identificação do gene serina hidroximetiltransferase, glutamato-desidrogenase e asparatato-aminotransferase.

Estes resultados mostram, pela primeira vez, as vias metabólicas de carbono e energia neste microrganismo. Os genes de interesse na área de pesquisa de nosso laboratório estão sendo empregados como sondas para o isolamento de clones genômicos,

a partir de uma biblioteca genômica de *T. reesei* (Matheucci *et al.*, 1995) e utilizados no desenvolvimento de novas pesquisas.

Análise de expressão gênica

No meio ambiente natural *T. reesei* degrada preferencialmente celulose para obter glicose como fonte de carbono e energia (Klein *et al.*, 1998). O crescimento de *T. reesei* em meio contendo glicose como fonte de carbono, mostrado na Figura 12, é similar ao mostrado por outros fungos filamentosos (Meletiadiis *et al.*, 2001), chegando a uma fase estacionária após 35 horas de cultivo e a glicose ter se esgotado no meio. Em média, os fungos empregam 55% da glicose consumida em síntese de biomassa e 45% para o metabolismo (Maheshwari *et al.*, 2000).

A preparação de RNA total e de DNA, para possibilitar a construção do “microarray”, constituiu-se em etapa crucial do trabalho, considerando-se a qualidade, quantidade e número de amostras que foram processadas.

As imagens de cada lâmina de “microarray” preparada foram recebidas prontas para a análise da empresa Genomic Solutions. Uma variedade de programas de análise de imagens de “microarray” (ArrayVision, GeneTac, AtlasImagem, ImaGene, ScanAlyze, etc.) podem ser obtidos de empresas ou centros acadêmicos. Sendo a forma de detecção dos elementos de DNA o critério mais importante a considerar, o programa GeneTac é considerado um bom programa de análise de imagem, por apresentar um bom sistema automático de localização dos elementos de DNA, por contar com uma base de dados administrada pelo programa SQL 7.0 (Microsoft® Co., USA) e ser de utilização muito simples.

A obtenção de dados da imagem pode ser afetada por múltiplos fatores (Wang *et al.*, 2001), desde a fabricação do “microarray” até a hibridação e o processamento da imagem. Alguns dos parâmetros que podem afetar a análise da imagem são:

1-Elementos de tamanho irregular: elemento muito pequeno pode ser afetado durante o processo de incorporação dos cDNAs na hibridação; elementos de diâmetro maior podem contaminar outros ou estar contaminados;

2-Intensidade de sinal: geralmente o “background” é empregado para se determinar a intensidade de sinal de um elemento;

3-Variabilidade do background local, devido à contaminação na área do elemento,

4-Excessivo “background” local: regiões com “background” maior que a média;

Na análise das imagens, alguns destes casos foram observados de forma isolada e, tendo sido identificados, não afetaram a obtenção final dos dados. Alguns elementos foram contaminados e puderam ser observados objetos contaminantes em algumas áreas das lâminas. Em geral, o valor do “background” foi baixo.

Os dados de intensidade relativa de fluorescência obtidos de cada uma das imagens (com Cy3 ou Cy5) devem ser normalizados. Este processo permite a correção de diferenças na eficiência da marcação dos cDNAs com os fluoróforos, e de diferenças na quantidade de RNA empregado na síntese do cDNA. A metodologia empregada para normalização de dados de “microarray”, consistiu na utilização de controles inseridos na lâmina e na preparação do cDNA fluorescente (Quackenbush, 2001). A análise dos dados obtidos do gene-controle em cada lâmina, como mostrado na Figura 18, mostra que, por não apresentarem intensidade média aceitável entre eles, foi necessário realizar um processo de normalização empregando-se os dados do elemento-controle de cada bloco de elementos (normalização local). A diferença da razão entre os elementos-controle deve-se, provavelmente, a problemas durante o processo de lavagem ou secagem das lâminas. O baixo número de genes descartados na análise de duplicata mostra que existe uma relação confiável entre os valores obtidos para cada elemento em cada uma das lâminas.

Assim, os dados obtidos correspondem aos genes depositados na lâmina e que foram agrupados para estabelecer um padrão similar de expressão. Para estabelecer este padrão de expressão, selecionamos todos os genes que apresentavam uma expressão diferencial de 2,0 vezes. Este fator foi utilizado considerando que variações no nível de expressão entre 1,5 e 2 são consideradas como variações na expressão de um gene (Brazma *et al.*, 2000). O fator de expressão diferencial é um fator empírico definido pelo pesquisador e não existe um critério definido para seu estabelecimento. Por exemplo, o grupo de Schena estudando genes de *Arabidopsis thaliana* utiliza um fator de expressão diferencial de 5 (Schena *et al.*, 1995), enquanto que, estudando a expressão de genes de células T humanas, utiliza um fator de 2 (Schena *et al.*, 1996). Com este critério, procuramos identificar os genes que apresentam um padrão de expressão comum, influenciado pela quantidade de glicose presente no meio. O resultado obtido mostra que só 298 genes (26%) da base de dados da EST de *T. reesei* apresentam expressão diferencial influenciada pela quantidade de glicose, por um fator de dois ou mais.

Para identificar os genes que apresentam um padrão de expressão similar, os dados podem ser tratados através de uma serie de métodos estatísticos como mostrado por Sherlock (2000) e Quackenbush (2001), pelo qual se adota uma descrição matemática de

similaridade que mede a diferença de expressão entre um par de genes. A metodologia de agrupamento hierárquico, descrita por Eisen *et al.* (1998), foi empregada por ser considerada mais simples.

Os resultados obtidos utilizando o programa TreeView, mostrados na Figura 19, nos permitem identificar os genes com padrão de expressão similar. Posto que o interesse maior deste trabalho era a identificação dos genes reprimidos em alta concentração de glicose (83 mM) e expressos em baixa concentração de glicose (1 mM e 0 mM), selecionamos o agrupamento de genes mostrado na Figura 19B.

No processo de validação por “Northern blot”, somente 13 genes (Figura 19C) foram validados. A validação dos dados obtidos no “microarray” é muito importante devido a fatores como a abundância do mRNA, o tamanho dos insertos e pequenos extremos 5’ do cDNA impresso que podem afetar a detecção e o nível de expressão determinado para um gene (Yang *et al.*, 1999). Este processo de validação busca confirmar os resultados obtidos pelo “microarray”, como mostrado por Schena *et al.* (1996) que avaliaram a expressão de 1056 genes de linfócitos T humanos em resposta a “shock” térmico e forbol-éster obtendo, respectivamente, 17 e 6 genes com expressão diferencial, confirmando por “Northern blot” 14 e 5 genes, respectivamente. Yang *et al.* (1999) realizaram a avaliação de 48 genes, mostrando que o “microarray” não havia detectado 3 genes que na análise por “Northern blot” se mostraram mais expressos e, ainda, que de 21 genes com expressão diferencial no “microarray” somente 17 foram confirmados por “Northern blot”.

Dentre os 14 genes com expressão diferencial podemos mencionar o gene da proibitina (“Prohibitin”), proteína que é considerada parte de um complexo de proteínas que agem de forma similar a “chaperones”, protegendo as proteínas mitocondriais da proteólise (Nijtmans *et al.*, 2000); uma glicoamilase (“glucoamylase”), enzima envolvida na degradação de amido; um transportador de peptídios (“Peptide transporter”) envolvido no transporte de aminoácidos e dipeptídios e de uma proteína envolvida na utilização de sorbitol (“sorbitol utilization protein”). Os genes serina-protease da família da subtilisina (“subtilisin like-serine protease”), glicose-6-fosfato 1-desidrogenase (“Glucose-6-phosphate 1-dehydrogenase”), transportador de cobre de alta afinidade (“high affinity copper transporter”), glicoamilase (“glucoamylase”) e 6-fosfogliconato desidrogenase (“6-Phosphogluconate dehydrogenase”) não são induzidos em *S. cerevisiae* quando a glicose se esgota no meio.

Em *T. reesei* existe repressão parcial dos dois primeiros genes da via das pentoses-fosfato (glicose-6-fosfato 1-desidrogenase e 6-fosfogliconato desidrogenase) em alta concentração de glicose, enquanto que em *S. cerevisiae* estes genes não são influenciados pela presença de glicose.

É importante mencionar que entre os genes identificados e validados encontramos 2 genes do ciclo de TCA (CIT e KDH) e 2 genes do “pyruvate bypass” (ALD2 e ACS). Estudos sobre repressão por glicose, em *S. cerevisiae*, têm mostrado que todos os genes do ciclo de TCA e os genes ALD e ACS são reprimidos por glicose (DeRisi *et al.*, 1997; Gancedo, 1998; Johnston, 1999). A análise da Figura 22A mostra que os transcritos dos genes do metabolismo de glicose, em *T. reesei*, não são alterados e só alguns genes diminuem ligeiramente a expressão com o esgotamento da glicose no meio. A comparação das Figuras 21A e 21B mostra algumas diferenças e similaridades no padrão de expressão de genes envolvidos no metabolismo de glicose discutidas a seguir.

Dos genes identificados na via glicolítica de *T. reesei* (FBA, TPI, TDH, PGK, ENO e PYK – este último amplificado do genoma e inserido no “microarray”), 3 deles - FBA, TDH e PGK, apresentaram expressão constitutiva, enquanto que outros 3 - TPI, ENO e PYK, apresentaram maior expressão quando a concentração de glicose no meio era 83 mmol/L.

O gene que codifica a enzima enolase (ENO) foi altamente expresso em meio rico em glicose. Portanto, como ocorre em *S. cerevisiae*, a regulação dos transcritos de genes da via glicolítica na presença de glicose permite o fluxo de metabólitos para o piruvato. O fato de que duas enzimas (glicose-6-fosfato 1-desidrogenase e 6-fosfogliconato desidrogenase), envolvidas nas primeiras etapas da via das pentoses-fosfato, serem parcialmente reprimidas na presença de glicose (Figura 20) também favorece o fluxo de metabólitos nesse sentido.

O gene PDC, tanto em *T. reesei* como em *S. cerevisiae*, é induzido na presença de glicose de forma que o piruvato produzido pela glicólise pode ser convertido a acetaldeído.

Em *T. reesei*, identificamos dois genes parálogos de aldeído-desidrogenase: o transcrito de ALD1, que não é reprimido por glicose, e o transcrito de ALD2, que é fortemente reprimido na presença de glicose. Se ambas as enzimas tivessem a mesma especificidade, então acetaldeído seria convertido em acetato em *T. reesei*. Deve-se ressaltar que foi detectada atividade enzimática de ADH em baixo nível (Tabela 8) que poderia conduzir a uma pequena produção de etanol.

Em *S. cerevisiae*, o acetaldeído formado é reduzido a etanol com participação de NADH, em uma reação catalisada pela enzima álcool-desidrogenase (ADH); o piruvato

não é convertido a acetato devido à forte repressão da enzima aldeído-desidrogenase (ALD) por glicose. Este passo é essencial para o metabolismo anaeróbico porque gera NAD^+ que é requerido para a continuidade da glicólise.

A ativação do gene que codifica a enzima acetil-coenzima A sintase (ACS) após o esgotamento da glicose, em ambos organismos, permite a entrada de acetato no ciclo de TCA. Nesta condição metabólica, será ativado o gene que codifica a enzima fosfoenolpiruvato-carboxiquinase (PCK), o que permite o fluxo de intermediários do ciclo de TCA, via oxaloacetato (como combustível), para a gliconeogênese.

Em *T. reesei* o gene PDA dirige o metabolismo do piruvato através do ciclo de TCA em que só CIT e KDH encontram-se parcialmente reprimidos, enquanto que IDH, SDH e MDH não são afetados na presença de glicose. Isto permite a completa oxidação do piruvato e mostra que em *T. reesei* a atividade mitocondrial é completamente funcional na presença de alta concentração de glicose. Tal característica se contrapõe ao que acontece em *S. cerevisiae* cujos transcritos do ciclo de TCA são reprimidos em alta concentração de glicose e a única via livre para o piruvato formado durante a glicólise é o piruvato “bypass”, com a formação de etanol.

A partir destes resultados podemos estabelecer que existe uma diferença notável com relação ao destino do piruvato entre os dois microrganismos, e que é evidente pela análise do padrão de expressão dos genes do ciclo de TCA após a depleção de glicose no meio. Enquanto em *S. cerevisiae* a alta concentração de glicose reprime fortemente os genes que codificam enzimas do ciclo de TCA, a expressão dos transcritos correspondentes em *T. reesei* é claramente diferente. Assim, a transição de uma condição rica em glicose a outra pobre em glicose origina grandes mudanças no padrão de expressão gênica em *S. cerevisiae*, o que não ocorre em *T. reesei*.

Portanto, em *T. reesei*, num meio rico em glicose, o piruvato formado pela via glicolítica é metabolizado por dois caminhos catabólicos: oxidação completa pelo ciclo de TCA e/ou a redução a acetaldeído através da via de “pyruvate bypass”. Em *S. cerevisiae*, o piruvato é exclusivamente levado a acetaldeído, em resultado da forte repressão dos transcritos de TCA por glicose.

A segunda diferença possível é o destino do acetaldeído formado pela descarboxilação do piruvato pela enzima piruvato Descarboxilase (PDC), que é ativada em ambos os microrganismos por glicose. Em *S. cerevisiae*, a repressão de ALD e a expressão de ADH levam o acetaldeído à formação de etanol, enquanto que em *T. reesei*, devido à

presença dos parálogos ALD1 e ALD2 e visto que ALD1 não é reprimido por glicose, é muito provável que acetaldeído seja convertido a acetato.

Gombert *et al.* (2001) realizaram a análise do fluxo metabólico em *S. cerevisiae*, mostrando que na respiração o fluxo principal dos metabólitos se realiza através da via das pentoses-fosfato e glicólise. Em condições fermentativas, o fluxo principal se realiza através da glicólise e o piruvato forma etanol como maior produto e pequenas quantidades de glicerol e acetato que são excretados ao meio. Durante o crescimento respiratório, em meio limitado de glicose, a via direta oxidativa é predominante na conversão de piruvato a acetyl-CoA. No entanto, um pequeno fluxo de piruvato, através do “pyruvate bypass”, é importante para proporcionar acetyl-CoA citossólico para síntese de lipídios (Flikweert *et al.*, 1997). Moller *et al.* (2002) realizaram a análise de fluxo metabólico de *S. kluyveri*, estabelecendo diferenças com relação a *S. cerevisiae*. *S. kluyveri*, em condições fermentativas, produz menos etanol devido ao fato do ciclo de TCA não ser reprimido significativamente na presença de glicose. Estes dados, comparados com os obtidos neste trabalho, mostrariam uma adaptação metabólica de *T. reesei* que, em condições respiratórias, tem a via das pentoses-fosfato parcialmente reprimida, desviando os substratos para serem metabolizados majoritariamente através da glicólise.

S. cerevisiae é uma levedura excepcional devido à sua capacidade de crescer em condições estritamente anaeróbicas (Lagunas, 1979). Em muitas leveduras não convencionais (leveduras não *Saccharomyces*, por exemplo: *Kluyveromyces*, *Cândida*, *Schizosaccharomyces*, *Pichia*, *Yarrowia*, etc.), o metabolismo de glicose é predominantemente aeróbico (Flores *et al.*, 2000). A regulação transcricional de genes do metabolismo de glicose (PDC, ADH, ACS) é similar tanto em *S. cerevisiae* como *K. lactis*, quando crescidas em meio com 50 mmol/L de glicose. Mas a diferença entre as duas leveduras provavelmente deve-se ao alto fluxo metabólico através da enzima PDC e/ou baixa atividade PDC em *K. lactis* (Zeeman *et al.*, 2000).

O perfil de expressão gênica descrito neste trabalho para *T. reesei* é possivelmente adotado por outros microrganismos multicelulares para a obtenção de energia através da respiração, ao invés da fermentação, na presença de meio rico em glicose. Em *Aspergillus nidulans*, por exemplo, o gene de citocromo C (cytochrome c) (*cycA*) também não é reprimido por glicose (Raitt *et al.*, 1994).

A engenharia metabólica de células eucarióticas representa um grande desafio para se produzir compostos úteis. O poder deste procedimento é evidente a partir da recente publicação na qual uma microalga fotossintética foi convertida num organismo

heterotrófico por introdução de um gene humano que codifica uma proteína transportadora de glicose (Zaslavskaja *et al.*, 2001). Sabendo que *T. reesei* é capaz de hidrolisar celulose a glicose (Béguin, 1990), o perfil de expressão gênica descrito neste trabalho constitui-se em informação importante para a possível conversão, por engenharia metabólica, deste microrganismo, preferencialmente aeróbio, em um microrganismo anaeróbio - um procedimento pertinente para a produção de etanol combustível a partir de biomassa celulósica. Em apoio a esta colocação, gostaríamos de mencionar que, recentemente, foi mostrado que a deleção do gene citocromo c (*cycC*) de *A. nidulans*, um microrganismo estritamente aeróbio, permite a obtenção de mutantes com capacidade fermentativa (Bradshaw *et al.*, 2001).

Outros procedimentos de alteração gênica de organismos encontram-se descritos na literatura. Blom *et al.* (2000) induziram a alta expressão do regulador transcricional Hap4p, que participa no controle transcricional de genes envolvidos no metabolismo respiratório, em leveduras crescidas em meio com 150 mmol/L de glicose. Esta supraexpressão motivou um metabolismo aeróbico e diminuição na produção de etanol. Porro *et al.* (1999) deletaram o único gene PDC de *K. lactis* e introduziram o gene L-lactato-desidrogenase (LDH) de bovino nesta levedura obtendo, assim, uma cepa mutante que produz lactato a partir de piruvato.

As células adaptam a expressão do genoma para se adequarem a mudanças no meio ambiente e no programa de crescimento e desenvolvimento celular. Mudanças no meio ambiente promovem rápida resposta das células, através da alteração no padrão global de expressão gênica, adaptando o organismo às novas condições ambientais (Gasch *et al.*, 2000; Kuhn *et al.*, 2001). Ferea *et al.* (1999) mostraram que depois de manter uma cultura de *S. cerevisiae* por 250 gerações em meio limitado de glicose, são obtidas leveduras mutantes com alteração na capacidade fermentativa, o que sugere a aparição, nas cepas evoluídas, de um pequeno grupo de mutações que favorecem a adaptação ao metabolismo aeróbico.

Em *Aspergillus* sp. a repressão por glicose afeta três grupos de genes: a) sistemas enzimáticos envolvidos na degradação de polissacarídeos (celulose, pectina, xylose); b) enzimas da gliconeogênese e do ciclo do glioxalato; e c) enzimas relacionadas à produção de metabólitos secundários (por exemplo penicilina, em *A. nidulans*) (Flipphi *et al.*, 2001). Em *T. reesei*, o sistema celulolítico é reprimido por glicose e induzido por celulose (El-Gogary *et al.*, 1989; Abrahão-Neto *et al.*, 1995).

A repressão por glicose de aproximadamente 10-15% dos genes de levedura é realizada por Mig1p e Mig2p, duas proteínas do tipo “zinc-finger” que se ligam ao DNA reprimindo a transcrição (Lutftyya *et al.*, 1998). A proteína *creA* de *Aspergillus* sp. é similar a Mig1 e tem homologia a *cre1* de *T. reesei* (Strauss *et al.*, 1995). Portanto, se *Aspergillus* sp. e *T. reesei* apresentam uma proteína similar a Mig1p e são organismos aeróbicos, então o mecanismo de regulação do metabolismo por glicose é diferente do mecanismo de *S. cerevisiae* (Ruijter *et al.*, 1997).

Finalmente, é aceito que, no curso da evolução, os organismos multicelulares aeróbicos surgiram após os unicelulares fermentadores depois do incremento de oxigênio na atmosfera, e que as mitocôndrias evoluíram a partir de bactérias aeróbicas (Zaslavskaja *et al.*, 2001). Não surpreende, portanto, que a alteração molecular da transcrição, em resposta à glicose, que determina a troca de fermentação por respiração, seja observada em genes que codificam enzimas mitocondriais. Uma vez que os efetores específicos para repressão por glicose são conservados em levedura e fungos filamentosos (Margulis, 1981; Ronne, 1995; Carraro *et al.*, 1998), e que os componentes da maquinaria de transcrição (proteínas reguladoras, fatores gerais e RNA polimerases) foram conservados durante a evolução em eucariotos diversos, desde leveduras até mamíferos (Schena, 1989), então a alteração provavelmente ocorreu nas regiões promotoras, impedindo a ligação do repressor com glicose nas seqüências reguladoras.

CONCLUSÕES

A identificação dos genes expressos e a análise do padrão de expressão gênica, nos permitem compreender e caracterizar o programa temporal e espacial de expressão gênica que determinam os processos da vida e as diferenças entre os seres vivos

A obtenção de ESTs nos permite identificar o maior número possível de genes representados numa biblioteca de cDNA, sendo necessário o estabelecimento de um banco de dados para administrar a informação obtida. O grande número de genes identificados, permite realizar análises em grande escala para caracterizar o perfil de expressão dos genes em determinadas condições.

T. reesei e *S. cerevisiae*, dois microorganismos que evoluíram em meios ambientes diferentes e adaptaram a expressão do seu genoma para adequar-se a câmbios no meio ambiente e no programa de crescimento e desenvolvimento celular, foram o tema de estudo neste trabalho. *T. reesei* organismo modelo de estudo de nosso laboratório, um fungo de grande interesse em biotecnologia e geneticamente pouco caracterizado, foi crescido em meio com glicose e o perfil de expressão gênica deste foi comparado com os dados já obtidos para *S. cerevisiae*, um organismo modelo eucariótico.

Para este estudo ser factível, foi construída uma biblioteca de cDNA e obtido o maior número de ESTs. A informação dos genes obtidos está disponível através de uma pagina de internet *Trichoderma reesei* EST database (<http://trichoderma.iq.usp.br/TrEST.html>).

A boa qualidade da biblioteca TrEST-A nos permite identificar mais de 10% dos genes expressos neste fungo, incrementando em mais de 10 vezes o número de genes depositados na base de dados GenBank (NCBI. USA).

O estabelecimento da base de dados EST de *T. reesei*, que permite administrar a informação obtida, possibilitará realizar a análise e identificação de genes que participam em distintas atividades metabólicas e aprofundar o conhecimento do metabolismo de este organismo.

O alto número de genes (61 %) sem similaridade na base de dados, que se constituem em novos genes e a identificação de genes envolvidos em distintas atividades celulares, demonstra a complexidade do genoma e evidencia a falta de estudos de genômica estrutural em fungos.

A identificação de um número significativo de genes, entre eles os do metabolismo de glicose, nos permitiu realizar a análise comparativa através da metodologia de DNA microarray. Sendo os resultados apresentados uma descrição definitiva da base molecular que dirige o produto final da via glicolítica, o piruvato, para o metabolismo aeróbio ou anaeróbio.

No caso de *T. reesei* fica claro que a regulação transcricional dos genes da glicólise envolvidos nos passos de frutose-1,6-bisfosfato a piruvato é similar à de *S. cerevisiae*, entretanto a oxidação de piruvato pelo TCA não é drasticamente reprimida, no nível transcricional, por glicose como na levedura. Na verdade, os transcritos que fazem parte do TCA e são parcialmente reprimidos por glicose são citrato sintase e alfa-cetoglutarato desidrogenase.

Estes resultados permitem compreender porque alguns microorganismos eucarióticos, como *T. reesei*, utilizam glicose através do metabolismo aeróbio. Na presença de alta concentração de glicose, a expressão dos genes de enzimas do ciclo de TCA permite o fluxo de piruvato através do ciclo até sua completa oxidação a CO₂ e H₂O. O piruvato também pode se dirigir a acetaldeído e, em seguida, a acetato pela rota de piruvato. A conversão de acetaldeído a acetato, e não a etanol, como em *S. cerevisiae*, não permite a regeneração de NAD⁺, requerido para o metabolismo anaeróbico.

A regulação da transcrição de genes pela glicose é, então, diferente em *S. cerevisiae* e *T. reesei*, em relação a genes críticos que controlam a direção do fluxo dos metabólitos. Enquanto a expressão de genes envolvidos no ciclo de TCA está reprimida fortemente em *S. cerevisiae* na presença de glicose, estes genes estão ativos em *T. reesei*. Assim, em contraste com *S. cerevisiae*, o metabolismo aeróbio prevalece em *T. reesei*.

O perfil de expressão gênica descrita em *T. reesei* é, possivelmente, adotado por outros microorganismos multicelulares para obter energia através da respiração, ao invés da fermentação. A informação obtida é importante na eventual tentativa de converter, por engenharia metabólica, este microorganismo, preferencialmente aeróbio, em um microorganismo anaeróbio; um procedimento pertinente para a produção de etanol a partir de biomassa celulósica.

O mecanismo de repressão por glicose é conservado em leveduras e fungos filamentosos. É razoável supor, portanto, que o metabolismo diferenciado de *T. reesei* e *S. cerevisiae*, em alta concentração de glicose, esteja associado, ao nível molecular, com as regiões promotoras dos genes. A avaliação desta hipótese está em andamento.

REFERÊNCIAS

- [1] Abrahao-Neto,J., Rossini,C.H., El Gogary,S., Henrique-Silva,F., Crivellaro,O., & El Dorry,H. (1995) Mitochondrial functions mediate cellulase gene expression in *Trichoderma reesei*. *Biochemistry*, **34**, 10456-10462.
- [2] Adams,M.D., Kelley,J.M., Gocayne,J.D., Dubnick,M., Polymeropoulos,M.H., Xiao,H., Merrill,C.R., Wu,A., Olde,B., Moreno,R.F., & . (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, **252**, 1651-1656.
- [3] Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M., Utterback,T.R., Nagle,J.W., Fields,C., & Venter,J.C. (1992) Sequence identification of 2,375 human brain genes. *Nature*, **355**, 632-634.
- [4] Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C., & Venter,J.C. (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat. Genet.*, **4**, 373-380.
- [5] Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., & . (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*, **377**, 3-174.
- [6] Alexopoulos,C.J., Mims,C.W., & Blackwell,M. (1996) *Introductory Mycology*, Fourth edition edn.
- [7] Altschul,S.F., Gish,W., Miller,W., Myers,E.W., & Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403-410.
- [8] Altschul,S.F. (1991) Amino acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.* , **219**, 555-565.
- [9] Ausbel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., & Smith,J.S.K. (1990) *Short protocols in Molecular Biology*, 2 edc.
- [10] Banfi,S., Borsani,G., Rossi,E., Bernard,L., Guffanti,A., Rubboli,F., Marchitello,A., Giglio,S., Coluccia,E., Zollo,M., Zuffardi,O., & Ballabio,A. (1996) Identification and mapping of human cDNAs homologous to *Drosophila* mutant genes through EST database searching. *Nat. Genet.*, **13**, 167-174.
- [11] Beutler,H.O. (1984) Ethanol. In *Methods of enzymatic Analysis* (Ed.Bermeyer,H.U., ed), pp. 598-606. Verlag Chemie GMBH.
- [12] Béguin,P. (1990) Molecular biology of cellulose degradation. *Annual Review of Microbiology*, **44**, 219-248.
- [13] Blom,J., De Mattos,M.J., & Grivell,L.A. (2000) Redirection of the respiro-fermentative flux distribution in *Saccharomyces cerevisiae* by overexpression of the transcription factor Hap4p. *Appl. Environ. Microbiol*, **66**, 1970-1973.

- [14] Boguski,M.S., Lowe,T.M., & Tolstoshev,C.M. (1993) dbEST--database for "expressed sequence tags". *Nat. Genet.*, **4**, 332-333.
- [15] Boguski,M.S. & Schuler,G.D. (1995) ESTablishing a human transcript map. *Nat. Genet.*, **10**, 369-371.
- [16] Bonalume,R. (1997) Brazil to sequence 'first plant pathogen'. *Nature*, **389**, 654.
- [17] Bradshaw,R.E., Bird,D.M., Brown,S., Gardiner,R.E., & Hirst,P. (2001) Cytochrome c is not essential for viability of the fungus *Aspergillus nidulans*. *Mol. Genet. Genomics*, **266**, 48-55.
- [18] Brazma,A. & Vilo,J. (2000) Gene expression data analysis. *FEBS Lett.*, **480**, 17-24.
- [19] Brown,P.O. & Botstein,D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat. Genet.*, **21**, 33-37.
- [20] Camargo,A.A., Samaia,H.P., Dias-Neto,E., Simao,D.F., Migotto,I.A., Briones,M.R., Costa,F.F., Nagai,M.A., Verjovski-Almeida,S., Zago,M.A., Andrade,L.E., Carrer,H., El Dorry,H.F., Espreafico,E.M., Habr-Gama,A., *et al.* (2001) The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc. Natl. Acad. Sci. U. S. A*, **98**, 12103-12108.
- [21] Carle-Urioste,J.C., Escobar-Vera,J., El Gogary,S., Henrique-Silva,F., Torigoi,E., Crivellaro,O., Herrera-Estrella,A., & El Dorry,H. (1997) Cellulase induction in *Trichoderma reesei* by cellulose requires its own basal expression. *J. Biol. Chem.*, **272**, 10169-10174.
- [22] Carlson,M. (1999) Glucose repression in yeast. *Curr. Opin. Microbiol.*, **2**, 202-207.
- [23] Carraro,D.M., Ferreira,J., Jr., Schumacher,R., Pereira,G.G., Hollenberg,C.P., & El Dorry,H. (1998) A region of the cellobiohydrolase I promoter from the filamentous fungus *Trichoderma reesei* mediates glucose repression in *Saccharomyces cerevisiae*, dependent on mitochondrial activity. *Biochem. Biophys. Res. Commun.*, **253**, 407-414.
- [24] Carulli,J.P., Artinger,M., Swain,P.M., Root,C.D., Chee,L., Tulig,C., Guerin,J., Osborne,M., Stein,G., Lian,J., & Lomedico,P.T. (1998) High throughput analysis of differential gene expression. *J. Cell Biochem. Suppl*, **30-31**, 286-296.
- [25] Celis,J.E., Kruhoffer,M., Gromova,I., Frederiksen,C., Ostergaard,M., Thykjaer,T., Gromov,P., Yu,J., Palsdottir,H., Magnusson,N., & Orntoft,T.F. (2000) Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.*, **480**, 2-16.
- [26] Chet,I. (1987) *Trichoderma*, application, mode of action, and potential as a biocontrol agent of soilborne plant pathogenic fungi. In ***Innovative approaches to Plant Disease Control*** (I.Chet, ed), pp. 137-160.
- [27] Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J., & Rutter,W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*, **18**, 5294-5299.

- [28] Chu,S., DeRisi,J., Eisen,M., Mulholland,J., Botstein,D., Brown,P.O., & Herskowitz,I. (1998) The transcriptional program of sporulation in budding yeast. *Science*, **282**, 699-705.
- [29] Crookshanks,M., Emmersen,J., Welinder,K.G., & Nielsen,K.L. (2001) The potato tuber transcriptome: analysis of 6077 expressed sequence tags. *FEBS Lett.*, **506**, 123-126.
- [30] de Souza,S.J., Camargo,A.A., Briones,M.R., Costa,F.F., Nagai,M.A., Verjovski-Almeida,S., Zago,M.A., Andrade,L.E., Carrer,H., El Dorry,H.F., Espreadico,E.M., Habr-Gama,A., Giannella-Neto,D., Goldman,G.H., Gruber,A., Hackel,C., *et al* (2000) Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A*, **97**, 12690-12693.
- [31] DeRisi,J.L., Iyer,V.R., & Brown,P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680-686.
- [32] Dias,N.E., Garcia,C.R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da,S.W., Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J., & Simpson,A.J. (2000) Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A*, **97**, 3491-3496.
- [33] Duggan,D.J., Bittner,M., Chen,Y., Meltzer,P., & Trent,J.M. (1999) Expression profiling using cDNA microarrays. *Nat. Genet.*, **21**, 10-14.
- [34] Eisen,M.B., Spellman,P.T., Brown,P.O., & Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A*, **95**, 14863-14868.
- [35] El-Dorry,H., Crivellaro,O., Leite,A., Abrahao-Neto,J., Henrique-Silva,F., Escobar-Vera,J., Matheucci,J.E., Carle-Urioste,J.C., Pereira,G.G., Carraro-Pereira,D.M., & el Gogary,S. (1996) Transcriptional control of the cellulase genes in *Trichoderma reesei*. *Braz. J. Med. Biol. Res.*, **29**, 905-909.
- [36] El-Gogary,S., Leite,A., Crivellaro,O., Eveleigh,D.E., & El Dorry,H. (1989) Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*. *Proc. Natl. Acad. Sci. U. S. A*, **86**, 6138-6141.
- [37] Eveleigh,D.E. (1985) *Biology of Industrial Microorganisms*.
- [38] Ewing,B. & Green,P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186-194.
- [39] Ewing,B., Hillier,L., Wendl,M.C., & Green,P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, **8**, 175-185.
- [40] Ewing,B. & Green,P. (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.*, **25**, 232-234.

- [41] Ferea,T.L., Botstein,D., Brown,P.O., & Rosenzweig,R.F. (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. U. S. A*, **96**, 9721-9726.
- [42] Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J.F., Dougherty,B.A., Merrick,J.M., & . (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496-512.
- [43] Flikweert,M.T., van Dijken,J.P., & Pronk,J.T. (1997) Metabolic responses of pyruvate decarboxylase-negative *Saccharomyces cerevisiae* to glucose excess. *Appl. Environ. Microbiol.*, **63**, 3399-3404.
- [44] Flipphi,M., Mathieu,M., Cirpus,I., Panozzo,C., & Felenbok,B. (2001) Regulation of the aldehyde dehydrogenase gene (*aldA*) and its role in the control of the coinducer level necessary for induction of the ethanol utilization pathway in *Aspergillus nidulans*. *J. Biol. Chem.*, **276**, 6950-6958.
- [45] Flores,C.L., Rodriguez,C., Petit,T., & Gancedo,C. (2000) Carbohydrate and energy-yielding metabolism in non-conventional yeasts. *FEMS Microbiol. Rev.*, **24**, 507-529.
- [46] Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., Bult,C.J., Kerlavage,A.R., Sutton,G., Kelley,J.M., & . (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397-403.
- [47] Fukuchi,S. & Otsuka,J. (1992) Evolution of metabolic pathways by chance assembly of enzyme proteins generated from sense and antisense strands of pre-existing genes. *J. Theor. Biol.*, **158**, 271-291.
- [48] Gams,W. & Bissett,J. (1998) Morphology and identification of *Trichoderma*. In *Trichoderma & Gliocladium* (Christian P Kubicek & Gary E Harman, eds), pp. 3-34.
- [49] Gancedo,J.M. (1998) Yeast carbon catabolite repression. *Microbiol. Mol. Biol. Rev.*, **62**, 334-361.
- [50] Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D., & Brown,P.O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, **11**, 4241-4257.
- [51] Gerhold,D. & Caskey,C.T. (1996) It's the genes! EST access to human genome content. *Bioessays*, **18**, 973-981.
- [52] Gish,W. & States,D.J. (1993) Identification of protein coding regions by database similarity search. *Nat. Genet.*, **3**, 266-272.
- [53] Goffeau,A., Barrell,B.G., Bussey,H., Davis,R.W., Dujon,B., Feldmann,H., Galibert,F., Hoheisel,J.D., Jacq,C., Johnston,M., Louis,E.J., Mewes,H.W., Murakami,Y., Philippsen,P., Tettelin,H., & Oliver,S.G. (1996) Life with 6000 genes. *Science*, **274**, 546, 563-546, 567.

- [54] Gombert,A.K., Moreira dos,S.M., Christensen,B., & Nielsen,J. (2001) Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *J. Bacteriol.*, **183**, 1441-1451.
- [55] Gordon,D., Abajian,C., & Green,P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.*, **8**, 195-202.
- [56] Hamer,L., Adachi,K., Montenegro-Chamorro,M.V., Tanzer,M.M., Mahanty,S.K., Lo,C., Tarpey,R.W., Skalhunes,A.R., Heiniger,R.W., Frank,S.A., Darveaux,B.A., Lampe,D.J., Slater,T.M., Ramamurthy,L., DeZwaan,T.M., Nelson,G.H., Shuster,J.R., Woessner,J., & Hamer,J.E. (2001) Gene discovery and gene function assignment in filamentous fungi. *Proc. Natl. Acad. Sci. U. S. A*, **98**, 5110-5115.
- [57] Heller,R.A., Schena,M., Chai,A., Shalon,D., Bedilion,T., Gilmore,J., Woolley,D.E., & Davis,R.W. (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. U. S. A*, **94**, 2150-2155.
- [58] Herrera-Estrella,A., Goldman,G.H., Van Montagu,M., & Geremia,R.A. (1993) Electrophoretic karyotype and gene assignment to resolved chromosomes of *Trichoderma* spp. *Mol. Microbiol.*, **7**, 515-521.
- [59] Hillier,L.D., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Marra,M., & . (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.*, **6**, 807-828.
- [60] Hondmann,D.H., Busink,R., Witteveen,C.F., & Visser,J. (1991) Glycerol catabolism in *Aspergillus nidulans*. *J. Gen. Microbiol.*, **137** (Pt 3), 629-636.
- [61] Inderlied,C.B. & Sypherd,P.S. (1978) Glucose metabolism and dimorphism in *Mucor*. *J. Bacteriol.*, **133**, 1282-1286.
- [62] Johnston,M. (1999) Feasting, fasting and fermenting. Glucose sensing in yeast and other cells. *Trends Genet.*, **15**, 29-33.
- [63] Kelly,J.M., Drysdale,M.R., Sealy-Lewis,H.M., Jones,I.G., & Lockington,R.A. (1990) Alcohol dehydrogenase III in *Aspergillus nidulans* is anaerobically induced and post-transcriptionally regulated. *Mol. Gen. Genet.*, **222**, 323-328.
- [64] Khan,J., Saal,L.H., Bittner,M.L., Chen,Y., Trent,J.M., & Meltzer,P.S. (1999) Expression profiling in cancer using cDNA microarrays. *Electrophoresis*, **20**, 223-229.
- [65] Klein,D. & Eveleigh,D.E. (1998). Ecology of *Trichoderma*. *Trichoderma & Gliocladium*. 57-74. Christian P Kubicek and Gary E Harman. Edit.
- [66] Kozian,D.H. & Kirschbaum,B.J. (1999) Comparative gene-expression analysis. *Trends Biotechnol.*, **17**, 73-78.

- [67] Kuhls,K., Lieckfeldt,E., Samuels,G.J., Kovacs,W., Meyer,W., Petrini,O., Gams,W., Borner,T., & Kubicek,C.P. (1996) Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. *Proc. Natl. Acad. Sci. U. S. A*, **93**, 7755-7760.
- [68] Kuhn,K.M., DeRisi,J.L., Brown,P.O., & Sarnow,P. (2001) Global and specific translational regulation in the genomic response of *Saccharomyces cerevisiae* to a rapid transfer from a fermentable to a nonfermentable carbon source. *Mol. Cell Biol.*, **21**, 916-927.
- [69] Kuklin,A. Microarray processing Technology. BioDiscovery Inc. 103, 1-4. 1999. Biodiscovery News. Technical Report
- [70] Kupfer,D.M., Reece,C.A., Clifton,S.W., Roe,B.A., & Prade,R.A. (1997) Multicellular ascomycetous fungal genomes contain more than 8000 genes. *Fungal. Genet. Biol.*, **21**, 364-372.
- [71] Lagunas,R. (1979) Energetic irrelevance of aerobiosis for *S. cerevisiae* growing on sugars. *Mol. Cell Biochem.* , **27**, 139-146.
- [72] Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., Funke,R., Gage,D., Harris,K., Heaford,A., Howland,J., Kann,L., Lehoczy,J., LeVine,R., McEwan,P., *et al* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
- [73] Lashkari,D.A., DeRisi,J.L., McCusker,J.H., Namath,A.F., Gentile,C., Hwang,S.Y., Brown,P.O., & Davis,R.W. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. U. S. A*, **94**, 13057-13062.
- [74] Lim,D., Hains,P., Walsh,B., Bergquist,P., & Nevalainen,H. (2001) Proteins associated with the cell envelope of *Trichoderma reesei*: a proteomic approach. *Proteomics.*, **1**, 899-909.
- [75] Lipshutz,R.J., Fodor,S.P., Gingeras,T.R., & Lockhart,D.J. (1999) High density synthetic oligonucleotide arrays. *Nat. Genet.*, **21**, 20-24.
- [76] Lluisma,A.O. & Ragan,M.A. (1997) Expressed sequence tags (ESTs) from the marine red alga *Gracilaria gracilis*. *J. Applied Phycology*, **9**, 287-296.
- [77] Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H., & Brown,E.L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675-1680.
- [78] Lockhart,D.J. & Winzeler,E.A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405**, 827-836.
- [79] Loftus,S.K., Chen,Y., Gooden,G., Ryan,J.F., Birznieks,G., Hilliard,M., Baxeavanis,A.D., Bittner,M., Meltzer,P., Trent,J., & Pavan,W. (1999) Informatic selection of a neural crest-melanocyte cDNA set for microarray analysis. *Proc. Natl. Acad. Sci. U. S. A*, **96**, 9277-9280.

- [80] Lutfiyya,L.L., Iyer,V.R., DeRisi,J., DeVit,M.J., Brown,P.O., & Johnston,M. (1998) Characterization of three related glucose repressors and genes they regulate in *Saccharomyces cerevisiae*. *Genetics*, **150**, 1377-1391.
- [81] Maheshwari,R., Bharadwaj,G., & Bhat,M.K. (2000) Thermophilic fungi: their physiology and enzymes. *Microbiol. Mol. Biol. Rev.*, **64**, 461-488.
- [82] Mantyla,A.L., Rossi,K.H., Vanhanen,S.A., Penttila,M.E., Suominen,P.L., & Nevalainen,K.M. (1992) Electrophoretic karyotyping of wild-type and mutant *Trichoderma longibrachiatum* (reesei) strains. *Curr. Genet.*, **21**, 471-477.
- [83] Margulis,L. (1981) *Symbiosis in Cell Evolution*. Freeman, San Francisco.
- [84] Matheucci,E., Jr., Henrique-Silva,F., el Gogary,S., Rossini,C.H., Leite,A., Vera,J.E., Urioste,J.C., Crivellaro,O., & El Dorry,H. (1995) Structure, organization and promoter expression of the actin-encoding gene in *Trichoderma reesei*. *Gene*, **161**, 103-106.
- [85] Maxam,A.M. & Gilbert,W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **74**, 560-564.
- [86] McCombie,W.R., Adams,M.D., Kelley,J.M., FitzGerald,M.G., Utterback,T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J.C., & Fields,C. (1992) *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nat. Genet.*, **1**, 124-131.
- [87] Meier-Ewert,S., Lange,J., Gerst,H., Herwig,R., Schmitt,A., Freund,J., Elge,T., Mott,R., Herrmann,B., & Lehrach,H. (1998) Comparative gene expression profiling by oligonucleotide fingerprinting. *Nucleic Acids Res.*, **26**, 2216-2223.
- [88] Meletiadiis,J., Meis,J.F., Mouton,J.W., & Verweij,P.E. (2001) Analysis of growth characteristics of filamentous fungi in different nutrient media. *J. Clin. Microbiol.*, **39**, 478-484.
- [89] Moller,K., Christensen,B., Forster,J., Piskur,J., Nielsen,J., & Olsson,L. (2002) Aerobic glucose metabolism of *Saccharomyces kluyveri*: Growth, metabolite production, and quantification of metabolic fluxes. *Biotechnol. Bioeng.*, **77**, 186-193.
- [90] Montenecourt,B.S. & Eveleigh,D.E. (1979) Selective screening for the isolation of high yielding cellulase mutants of *T.reesei*. *Adv. Chem. Ser.*, **181**, 289-301.
- [91] Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L., Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E., Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K., Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S., Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Natvig,D.O., & . (1997) Expressed sequences from conidial, mycelial, and sexual stages of *Neurospora crassa*. *Fungal. Genet. Biol.*, **21**, 348-363.
- [92] Nelson,P.S., Ng,W.L., Schummer,M., True,L.D., Liu,A.Y., Bumgarner,R.E., Ferguson,C., Dimak,A., & Hood,L. (1998) An expressed-sequence-tag database of the human prostate: sequence analysis of 1168 cDNA clones. *Genomics*, **47**, 12-25.

- [93] Nijtmans,L.G., de Jong,L., Artal,S.M., Coates,P.J., Berden,J.A., Back,J.W., Muijsers,A.O., van der,S.H., & Grivell,L.A. (2000) Prohibitins act as a membrane-bound chaperone for the stabilization of mitochondrial proteins. *EMBO J.*, **19**, 2444-2451.
- [94] Pearson,W.R. (1991) Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics*, **11**, 635-650.
- [95] Pfeiffer,T., Schuster,S., & Bonhoeffer,S. (2001) Cooperation and competition in the evolution of ATP-producing pathways. *Science*, **292**, 504-507.
- [96] Porro,D., Bianchi,M.M., Brambilla,L., Menghini,R., Bolzani,D., Carrera,V., Lievense,J., Liu,C.L., Ranzi,B.M., Frontali,L., & Alberghina,L. (1999) Replacement of a metabolic pathway for large-scale production of lactic acid from engineered yeasts. *Appl. Environ. Microbiol.*, **65**, 4211-4215.
- [97] Quackenbush,J., Liang,F., Holt,I., Pertea,G., & Upton,J. (2000) The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Res.*, **28**, 141-145.
- [98] Quackenbush,J. (2001) Computational analysis of microarray data. *Nat. Rev. Genet.*, **2**, 418-427.
- [99] Raitt,D.C., Bradshaw,R.E., & Pillar,T.M. (1994) Cloning and characterisation of the cytochrome c gene of *Aspergillus nidulans*. *Mol. Gen. Genet.*, **242**, 17-22.
- [100] Reese,E.T. & Mandels,M. (1980) Stability of the cellulase of *Trichoderma reesei* under use conditions. *Biotechnol. Bioeng.*, **22**, 323-335.
- [101] Reue,K. (1998) mRNA quantitation techniques: considerations for experimental design and application. *J. Nutr.*, **128**, 2038-2044.
- [102] Rogers,P.J., Clark-Walker,G.D., & Stewart,P.R. (1974) Effects of oxygen and glucose on energy metabolism and dimorphism of *Mucor genevensis* grown in continuous culture: reversibility of yeast- mycelium conversion. *J. Bacteriol.*, **119**, 282-293.
- [103] Ronne,H. (1995) Glucose repression in fungi. *Trends Genet.*, **11**, 12-17.
- [104] Ruijter,G.J. & Visser,J. (1997) Carbon repression in *Aspergilli*. *FEMS Microbiol Lett.*, **151**, 103-114.
- [105] Sakai,K., Hasumi,K., & Endo,A. (1990) Two glyceraldehyde-3-phosphate dehydrogenase isozymes from the koningic acid (heptelidic acid) producer *Trichoderma koningii*. *Eur. J. Biochem.*, **193**, 195-202.
- [106] Sambrook,J., Fritsch,E.F., & Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour, NY.
- [107] Samuels,G.J. (1996) *Trichoderma*: a review of biology and systematics of the genus. *Mycology Research*, **100**, 923-935.

- [108] Sanger,F., Nicklen,S., & Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A*, **74**, 5463-5467.
- [109] Sanger,F., Coulson,A.R., Hong,G.F., Hill,D.F., & Petersen,G.B. (1982) Nucleotide sequence of bacteriophage lambda DNA. *J. Mol. Biol.*, **162**, 729-773.
- [110] Schena,M. (1989) The evolutionary conservation of eukaryotic gene transcription. *Experientia*, **45**, 972-983.
- [111] Schena,M., Shalon,D., Davis,R.W., & Brown,P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467-470.
- [112] Schena,M., Shalon,D., Heller,R., Chai,A., Brown,P.O., & Davis,R.W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. U. S. A*, **93**, 10614-10619.
- [113] Schindler,M., Mach,R.L., Vollenhofer,S.K., Hodits,R., Gruber,F., Visser,J., De Graaff,L., & Kubicek,C.P. (1993) Characterization of the pyruvate kinase-encoding gene (pki1) of *Trichoderma reesei*. *Gene*, **130**, 271-275.
- [114] Schuler,G.D., Boguski,M.S., Stewart,E.A., Stein,L.D., Gyapay,G., Rice,K., White,R.E., Rodriguez-Tome,P., Aggarwal,A., Bajorek,E., Bentolila,S., Birren,B.B., Butler,A., Castle,A.B., Chiannilkulchai,N., Chu,A., Clee,C., Cowles,S., Day,P.J., Dibling,T., Drouot,N., Dunham,I., Duprat,S., East,C., Hudson,T.J., & . (1996) A gene map of the human genome. *Science*, **274**, 540-546.
- [115] Schulz,B.E., Kraepelin,G., & Hinkelmann,W. (1974) Factors affecting dimorphism in *Mycotypha* (Mucorales): a correlation with the fermentation-respiration equilibrium. *J. Gen. Microbiol.*, **82**, 1-13.
- [116] Seoighe,C. & Wolfe,K.H. (1999) Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.*, **2**, 548-554.
- [117] Sherlock,G. (2000) Analysis of large-scale gene expression data. *Curr. Opin. Immunol.*, **12**, 201-205.
- [118] Shoop,E., Carlis,J., & Retzel,E. Automating and Streamlining Inference of Function of ESTs within a Data Analysis System. virtual library at Michigan State University.1-11.1994.
- [119] Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., Bueno,M.R., Camargo,A.A., Camargo,L.E., *et al.*. (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature*, **406**, 151-157.
- [120] Soares,M.B., Bonaldo,M.F., Jelene,P., Su,L., Lawton,L., & Efstratiadis,A. (1994) Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. U. S. A*, **91**, 9228-9232.

- [121] Staudt,L.M. & Brown,P.O. (2000) Genomic views of the immune system*. *Annu. Rev. Immunol.*, **18**, 829-859.
- [122] Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A., Holmberg,A., Amini,B., Bhalerao,R., Larsson,M., Villarroel,R., Van Montagu,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W., Gustafsson,P., Uhlen,M., Sundberg,B., & Lundeberg,J. (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5, 692 expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A*, **95**, 13330-13335.
- [123] Sterky,F. & Lundeberg,J. (2000) Sequence analysis of genes and genomes. *J. Biotechnol.*, **76**, 1-31.
- [124] Strauss,J., Mach,R.L., Zeilinger,S., Hartler,G., Stoffler,G., Wolschek,M., & Kubicek,C.P. (1995) Cre1, the carbon catabolite repressor protein from *Trichoderma reesei*. *FEBS Lett.*, **376**, 103-107.
- [125] Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H., III, Becker,K.G., & Ko,M.S. (2000) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc. Natl. Acad. Sci. U. S. A*, **97**, 9127-9132.
- [126] Torigoi,E., Henrique-Silva,F., Escobar-Vera,J., Carle-Urioste,J.C., Crivellaro,O., El Dorry,H., & el Gogary,S. (1996) Mutants of *Trichoderma reesei* are defective in cellulose induction, but not basal expression of cellulase-encoding genes. *Gene*, **173**, 199-203.
- [127] Tzagoloff,A. Mitochondria. (1982) Plenum Press, New York, NY.
- [128] Van Hal,N.L., Vorst,O., van Houwelingen,A.M., Kok,E.J., Peijnenburg,A., Aharoni,A., van Tunen,A.J., & Keijer,J. (2000) The application of DNA microarrays in gene expression analysis. *J. Biotechnol.*, **78**, 271-280.
- [129] Vanhanen,S., Penttila,M., Lehtovaara,P., & Knowles,J. (1989) Isolation and characterization of the 3-phosphoglycerate kinase gene (pgk) from the filamentous fungus *Trichoderma reesei*. *Curr. Genet.*, **15**, 181-186.
- [130] Vaulont,S., Vasseur-Cognet,M., & Kahn,A. (2000) Glucose regulation of gene transcription. *J. Biol. Chem.*, **275**, 31555-31558.
- [131] Velculescu,V.E., Zhang,L., Vogelstein,B., & Kinzler,K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484-487.
- [132] Velculescu,V.E., Zhang,L., Zhou,W., Vogelstein,J., Basrai,M.A., Bassett,D.E., Jr., Hieter,P., Vogelstein,B., & Kinzler,K.W. (1997) Characterization of the yeast transcriptome. *Cell*, **88**, 243-251.
- [133] Velculescu,V.E., Madden,S.L., Zhang,L., Lash,A.E., Yu,J., Rago,C., Lal,A., Wang,C.J., Beaudry,G.A., Ciriello,K.M., Cook,B.P., Dufault,M.R., Ferguson,A.T., Gao,Y., He,T.C., Hermeking,H., Hiraldo,S.K., Hwang,P.M., Lopez,M.A., Luderer,H.F., Mathews,B., Petroziello,J.M., Polyak,K., Zawel,L., Kinzler,K.W., & . (1999) Analysis of human transcriptomes. *Nat. Genet.*, **23**, 387-388.

- [134] Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A., Gocayne,J.D., Amanatides,P., Ballew,R.M., Huson,D.H., Wortman,J.R., Zhang,Q., Kodira,C.D., Zheng,X.H., *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304-1351.
- [135] Wang,X., Ghosh,S., & Guo,S.W. (2001) Quantitative quality control in microarray image processing and data acquisition. *Nucleic Acids Res.*, **29**, E75-E75.
- [136] Waterston,R., Martin,C., Craxton,M., Huynh,C., Coulson,A., Hillier,L., Durbin,R., Green,P., Shownkeen,R., Halloran,N., & . (1992) A survey of expressed genes in *Caenorhabditis elegans*. *Nat. Genet.*, **1**, 114-123.
- [137] Watson,A., Mazumder,A., Stewart,M., & Balasubramanian,S. (1998) Technology for microarray analysis of gene expression. *Curr. Opin. Biotechnol.*, **9**, 609-614.
- [138] White,O. & Kerlavage,A.R. (1996) TDB: new databases for biological discovery. *Methods Enzymol.*, **266**, 27-40.
- [139] Wilson,C.A., Kreichman,J., & Gerstein,M. (2000) Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. *J. Mol. Biol.*, **297**, 233-249.
- [140] Yang,G.P., Ross,D.T., Kuang,W.W., Brown,P.O., & Weigel,R.J. (1999) Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res.*, **27**, 1517-1523.
- [141] Yang,Y.H., Dudoit,S., Luu,P., & Speed,T. (1998) Normalization for cDNA Microarray Data.. Unpublished Work
- [142] Yoder,O.C. & Turgeon,B.G. (2001) Fungal genomics and pathogenicity. *Curr. Opin. Plant Biol.*, **4**, 315-321.
- [143] Zaslavskaja,L.A., Lippmeier,J.C., Shih,C., Ehrhardt,D., Grossman,A.R., & Apt,K.E. (2001) Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science*, **292**, 2073-2075.
- [144] Zeeman,A.M., Kuyper,M., Pronk,J.T., van Dijken,J.P., & Steensma,H.Y. (2000) Regulation of pyruvate metabolism in chemostat cultures of *Kluyveromyces lactis* CBS 2359. *Yeast*, **16**, 611-620.

Curriculum vitae

Felipe Santiago Chambergo Alcalde.

Nascido em Chiclayo (Departamento de Lambayeque, Perú), em 04 de Setembro de 1965

Educação

Nível Primário

Colégio Particular Francisco Bolognesi

Chiclayo, Lambayeque. Perú

Início 1973 – Conclusão 1977

Nível Secundário

Colégio Nacional San José

Chiclayo, Lambayeque. Peru.

Início 1978 – Conclusão 1982

Nível Superior

Universidad Nacional Pedro Ruiz Gallo de Lambayeque

Licenciado em Ciências Biológicas, especialidad Microbiología-Parasitología

Lambayeque, Perú

Início 1983 – Conclusão 1991

Ocupação

Estudante de Doutorado

Área Bioquímica

Departamento de Bioquímica, Instituto de Química

Universidade de São Paulo, SP.

Bolsista: CNPq: 1998-1999

FAPESP: 2000 – 2002.

Publicação

Chambergo, F.S., Bonaccorsi, E., Ferreira, A, Ramos, A, Ferreira Jr, R., Abrahão-Neto, J., Farah, J. & El-Dorry, H. (2002) Elucidation of the metabolic fate of glucose in the filamentous fungus *Trichoderma reesei* using EST analysis and cDNA microarrays. *Journal of Biological Chemistry*, 277(16):13983 – 13988.