

Decifrando o genoma

Newton Portilho Carneiro¹
Andréa Almeida Carneiro²
Cláudia Teixeira Guimarães³
Edilson Paiva⁴

Resumo - Os genes, unidades biológicas que determinam as características de um organismo, estão contidos em moléculas de DNA presentes no núcleo das células. Portanto, o estudo do DNA é de fundamental importância para o entendimento de como as características de um organismo são formadas. Com o grande avanço alcançado pelas técnicas de Biologia molecular nas últimas décadas, hoje, somos capazes de isolar e seqüenciar moléculas de DNA de maneira rotineira. A partir da otimização das técnicas de seqüenciamento, surgiram programas com o objetivo de seqüenciar genomas inteiros. Atualmente, estão depositados em bancos de dados mais de 3 milhões de seqüências de DNA e, grande parte delas ainda não tem sua função conhecida. O processo de caracterização da função gênica envolve um volume muito maior de pesquisas e conhecimentos do que as etapas de isolamento e seqüenciamento. A função de um gene pode ser desvendada por meio de técnicas de genética direta ou reversa. Descrevem-se algumas das metodologias usadas para a caracterização da função gênica, bem como o potencial dos programas de seqüenciamento de genomas.

Palavras-chave: Genoma; Seqüenciamento; cDNA; Microarrays; Transposons; T-DNA.

INTRODUÇÃO

A estrutura tridimensional do DNA, definida por Watson e Crick em 1953, é similar para todos os seres vivos. Basicamente, uma molécula de DNA consiste de duas cadeias de nucleotídeos mantidas juntas por pontes de hidrogênio. Existem quatro nucleotídeos diferentes, cada um contendo uma desoxirribose, um grupo fosfato, e uma base púrica (adenina e guanina) ou pirimídica (timina e citosina). Na cadeia polinucleotídica, a desoxirribose e o fosfato estão sempre unidos por uma ligação fosfodiéster. Portanto, esta parte da molécula é muito regular. Em contraste, a ordem e a quantidade das bases púricas e pirimídicas, que formam uma cadeia de DNA, são altamente irregulares, variando de um organismo para outro. É exatamente

essa variação das bases que constitui o DNA, que produz as diferenças existentes entre os diversos genes.

SEQÜENCIAMENTO DE FRAGMENTOS DE DNA

Nos anos 70, foram desenvolvidos protocolos que permitiram o seqüenciamento das bases que compõem qualquer fragmento de DNA purificado. No entanto, as técnicas de seqüenciamento têm-se tornado cada vez mais avançadas e automatizadas, o que tem viabilizado os programas de cooperação internacional com o objetivo de seqüenciar os genes expressos, *Expressed Sequencing Tags* (ESTs), ou mesmo genomas inteiros, como são os casos do genoma humano, da bactéria *Shigella* e de *Arabidopsis*. Com isso, a

quantidade de seqüências de DNA, atualmente disponíveis, é tão grande que são necessários bancos de dados e programas de computador específicos para armazenar e analisar todas elas. Os projetos de seqüenciamentos, tanto de genomas completos quanto apenas de ESTs, apresentam vantagens e desvantagens. No caso do seqüenciamento de todo o genoma, são geradas informações sobre regiões que correspondem a genes e aquelas que, embora não sejam transcritas, podem possuir funções importantes na regulação gênica e manutenção da integridade do genoma. Nesse caso, o projeto envolve muito mais trabalho, uma vez que as regiões não-transcritas podem atingir 90% do genoma. Assim, mesmo considerando todas as informações obtidas com o seqüenciamento

¹Biólogo, Ph.D., Pesq. EMBRAPA Milho e Sorgo, Caixa Postal 151, CEP 35701-970 Sete Lagoas-MG. E-mail: newtonc@cnpm.br

²Bióloga, Ph.D., Pesq. CNPq/EMBRAPA Milho e Sorgo, Caixa Postal 151, CEP 35701-970 Sete Lagoas-MG. E-mail: andreac@cnpm.br

³Eng^a Agr^a, D.Sc., Pesq. EMBRAPA Milho e Sorgo, Caixa Postal 151, CEP 35701-970 Sete Lagoas-MG. E-mail: claudia@cnpm.br

⁴Eng^a Agr^a, Ph.D., Pesq. EMBRAPA Milho e Sorgo, Caixa Postal 151, CEP 35701-970 Sete Lagoas-MG. E-mail: edilson@cnpm.br

de todo o genoma, o projeto tem sido viável apenas para organismos com genomas pequenos. Aqueles organismos cujos genomas são grandes, como é o caso da maioria das plantas cultivadas, a melhor opção é seqüenciar apenas os ESTs e tentar obter o máximo de informações com relação às suas funções. Um sumário do número de seqüências de DNA ou ESTs disponíveis em bancos de dados de domínio público é mostrado no Quadro 1.

Com o seqüenciamento massivo de DNAs, o isolamento de um gene passou a não ser mais um problema. O grande desafio agora é saber qual a função desses genes seqüenciados. Várias seqüências de DNA, inicialmente incorporadas em bancos de dados, tiveram suas funções determinadas a partir de estudos detalhados de Biologia molecular. Atualmente, com o grande número de seqüências já depositado nos bancos de dados, muitas vezes, a função de uma seqüência gênica pode ser deduzida em comparações por homologia com aquelas já existentes (Fig. 1). Entretanto, grande parte das seqüências de DNA, incorporadas aos bancos de dados, não apresenta homologia significativa com as seqüências, cujas funções já são conhecidas. Assim, torna-se necessário um grande investimento em pesquisas para desvendar a função dos vários genes responsáveis pelas funções vitais de um organismo.

A função de um gene pode ser estudada através da genética direta ou da genética reversa. Genética direta é um processo que envolve inicialmente a análise do fenótipo de um organismo e, em seguida, o isolamento e a identificação do gene responsável pela característica estudada. Já a genética reversa, como o próprio nome indica, é o processo inverso à genética direta, inicia-se mutando um gene, introduzindo-o no organismo de interesse e estudando o fenótipo resultante (Fig. 2). Os processos de genética direta foram muito usados na década de 80. Com o desenvolvimento dos projetos de genomas e a disponibilidade de genes com seqüências caracterizadas, a genética reversa tomou um grande impulso.

QUADRO 1 - Sumário do número de seqüências de diferentes organismos depositados em bancos de dados

Organismo	Nome científico	ESTs seqüenciados
Homem	<i>Homo sapiens</i>	1.631.323
Camundongo	<i>Mus musculus + domesticus</i>	786.324
Rato	<i>Rattus sp.</i>	128.948
Mosca-das-frutas	<i>Drosophila melanogaster</i>	86.121
Nematóide	<i>Caenorhabditis elegans</i>	72.567
Arroz	<i>Oryza sativa</i>	47.141
<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	45.752
Milho	<i>Zea mays</i>	41.848
Tomate	<i>Lycopersicon esculentum</i>	43.436
Soja	<i>Glycine max</i>	27.158
Levedura	<i>Saccharomyces cerevisiae</i>	3.041
<i>Pinus</i>	<i>Pinus taeda</i>	7.554
Canola	<i>Brassica napus</i>	1.702
Algodão	<i>Gossypium hirsutum</i>	8.025
Total ⁽¹⁾	-	3.203.327

FONTE: List... (1999).

(1) Inclui o resultado do seqüenciamento de todos os organismos depositados no banco de dados.

Milho	ATG GGT AAG GAG AAG TCC CAC ATC AAC ATT GTG GTT ATT GGC
Arroz	ATG GGT AAG GAG AAG ACG CAC ATC AAC ATT GTG GTC ATT GGC
Trigo	ATG GGT AAA GAG AAG TTT CAC ATC AAC ATT GTG GTC ATT GGC
Rato	ATG GGC AAG GAG AAG ACA CAC ATC AAC ATT GTG GTC ATT GGC
Homem	ATG GGC AAG GAG AAG ACC CAC ATC AAC ATC GTG GTC ATC GGC
Galinha	ATG GGC AAG GAG AAG ATC CAT ATT AAC ATT GTG GTC ATT GGC

Figura 1 - Alinhamento da seqüência do gene fator de alongamento 1A de vários organismos

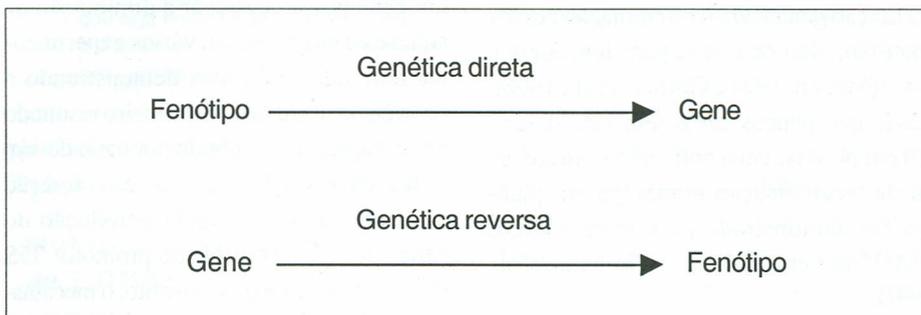


Figura 2 - Genética direta e reversa

NOTA: Genética direta é a identificação do gene a partir de um fenótipo. Genética reversa é a identificação de um fenótipo a partir de um gene.

TÉCNICAS DE GENÉTICA REVERSA USADAS PARA MANIPULAR GENES E PRODUZIR ORGANISMOS MUTANTES

As técnicas de recombinação homóloga, anti-senso e co-supressão podem ser usadas para alterar um gene conhecido, fazendo com que ele se torne não-funcional. As plantas transgênicas são importantes ferramentas científicas para auxiliar na identificação da função gênica.

Recombinação homóloga

A recombinação homóloga é a alteração de uma pequena parte da seqüência de um gene de interesse que, geralmente, é feita com a incorporação de um gene marcador e a reintrodução desse gene mutado no organismo de origem. Uma vez dentro do organismo de interesse, o gene mutado tem a capacidade de substituir o gene nativo pela recombinação homóloga criando um organismo mutante. Um gene marcador pode ser o gene *bar* de seleção para o herbicida PPT. As plantas contendo o gene modificado com o marcador são selecionadas na presença do herbicida e, após ficar demonstrada a incorporação do gene modificado no genoma da planta transformada, essa é autofecundada. Por estudos moleculares é possível identificar a planta que tenha apenas o alelo modificado em homozigose (Fig. 3, p.63). A planta homozigota para o gene modificado de interesse pode ou não demonstrar um fenótipo aparente. No entanto, aquelas cujos fenótipos são evidentes podem constituir uma forma rápida de correlacionar o fenótipo com o gene modificado. Métodos de interrupção da função gênica via recombinação homóloga têm sido descritos para leveduras e ratos (Melton, 1994 e Goffeau et al., 1996), sendo que poucos casos têm sido descritos em plantas. Uma aplicação com sucesso da recombinação homóloga em plantas foi demonstrada para o gene AGL MADS-box em *Arabidopsis* (Kempim et al., 1997).

Anti-senso

A metodologia de anti-senso envolve

a introdução na célula de moléculas de RNA ou de DNA, construídas artificialmente, que sejam complementares (anti-senso) ao RNA mensageiro (mRNA) do gene de interesse. Uma das explicações para o RNA anti-senso causar mutação no organismo transformado é o fato de que estas moléculas de RNA ou de DNA artificiais se ligam ao mRNA celular inativando-o (Fig. 4, p.63). Usando esta tecnologia o anti-senso do gene da *chalcone sintase* (*chs*), responsável pela pigmentação das flores, foi introduzido em plantas de petúnia e de tabaco, resultando em plantas com alteração na pigmentação das suas flores (fenótipo mutante) (Krol et al., 1988). Sheehy et al. (1988) também utilizaram a tecnologia do anti-senso para inibir a produção de enzimas responsáveis pelo desenvolvimento do fruto do tomate, produzindo assim um tomate mutante com o amadurecimento retardado.

Co-supressão

Na técnica de co-supressão, um gene de interesse é engenheirado, por meio de técnicas de Biologia molecular, para superexpressar uma proteína de interesse. Uma vez que a célula possui uma maquinaria minuciosamente ajustada, qualquer alteração nos níveis de expressão de uma proteína produz uma grande confusão nos mecanismos de regulação celular. Como conseqüência, na maioria das vezes, a superexpressão de uma proteína na célula faz com que a produção dessa proteína seja desligada e, ao contrário do que seria esperado, o organismo mutante não produz a proteína de interesse. Embora seja difícil compreender como a superexpressão de um gene possa ocasionar a diminuição da síntese de sua proteína, vários experimentos têm sido realizados demonstrando a ocorrência desse fato. O primeiro resultado de co-supressão foi obtido por meio de um estudo envolvendo a variação da coloração das flores de petúnia pela introdução do gene *chs* sob o controle do promotor 35S (Napoli et al., 1990). No entanto, o mecanismo que impede a expressão de genes endógenos por meio da co-supressão ainda não é totalmente entendido.

Mutação por inserção de transposon

Transposons são elementos de DNA que têm a capacidade de sair de uma região do genoma e incorporar em outra. Sendo o movimento dos transposons um processo aleatório, quando estes elementos "pulam" em um genoma podem inserir no meio de um gene tornando-o inativo. Em plantas, uma série de transposons tem sido usada como ferramentas, tanto na genética direta quanto na reversa, para isolamento e caracterização de genes ou fenótipos. O princípio da técnica está descrito na Figura 5, p.63. O primeiro gene de planta clonado por transposons via genética direta foi o gene "bronze" de milho que codifica UDP-glucose:flavonóide 3-O-glucosiltransferase, uma enzima da via metabólica das antocianinas (Fedoroff et al., 1984).

A genética reversa usando mutantes por inserção de transposons foi descrita inicialmente em *Drosophila melanogaster* (Ballinger & Benzer, 1989 e O'Hare, 1990). Essa metodologia baseia-se em uma reação de Polimerase Chain Reaction (PCR), utilizando um *primer* complementar ao final dos transposons e outro complementar ao final do gene (Rosenzweig et al., 1983 e Emmons & Yesner, 1984). Desta forma, os produtos de PCR só serão obtidos se um transposon estiver inserido no gene de interesse. Genes que tiveram sua expressão alterada pela inserção do transposon serão recuperados por meio da amplificação do DNA extraído do indivíduo com fenótipo mutante, utilizando um par de *primer* específico (Zwaal et al., 1993). Esse processo tem sido utilizado na identificação de genes em *Caenorhabditis elegans* (Rushforth et al., 1993 e Plasterk, 1993) e milho (Bensen et al., 1995).

A família dos transposons *Mutator* (*Mu*) apresenta alta taxa de mutação e tem alto grau de conservação nas extremidades das seqüências invertidas (Walbot, 1992). Essas duas características são bastante interessantes, pois ajudam a selecionar alelos que contenham os elementos *Mu*, cuja seqüência é previamente conhecida. Inserções dos elementos *Mu* podem ser identificadas pela amplificação por PCR.

O gene mutado pode então ser recuperado e propagado em sementes F2. Outros aspectos importantes desse processo é que o número de transposições e o número de cópias do transposon *Mu* na planta são fatores essenciais para a análise de um menor número de plantas. A tecnologia de caracterização de funções gênicas através do *Mu* foi utilizada pela primeira vez em milho pela Pioneer Hi-Bred Co., e foi denominada *Trait Utility System* (TUSC) (Fig. 6). Bensen et al. (1995) utilizaram a técnica do TUSC para caracterizar o mutante *Anther ear1* (*An1*), cujo produto gênico está envolvido na síntese do primeiro intermediário tetracíclico na via da biossíntese de giberelina (GA). A mutação *An1* resultou em um fenótipo responsivo a GA, caracterizado por uma altura reduzida de planta, atraso na maturidade e o desenvolvimento de flores perfeitas em espigas normalmente pestiladas.

Um projeto financiado pela *National Science Foundation* (NSF), coordenado pela Dra. Virginia Walbot da Universidade de Stanford, Estados Unidos, tem por objetivo demonstrar a funcionalidade de genes de milho, usando dois métodos complementares, o seqüenciamento de DNA genômico flanqueando as inserções do transposon *Mu* e a identificação e caracterização dos indivíduos mutantes contendo esses transposons. Um banco de mutantes está sendo criado, utilizando-se plantas transformadas com os elementos *Mu* contendo o plasmídeo Bluescript (Fig. 7). A nova inserção contendo o transposon poderá ser seletivamente clonada em *E. coli*, gerando uma biblioteca de mutação insercional para análise de DNA. Cada planta F2 será autofecundada e as sementes serão estocadas no *Maize Genetics Cooperative Stock Center*, um órgão especialmente criado para armazenar os estoques de sementes mutadas. Os usuários poderão utilizar a técnica de PCR para seleção de uma coleção de plasmídeos que foi inserida em genes de interesse. A estimativa é de que 50 mil ESTs, flanqueados pelo transposon *Mu*, sejam completamente seqüenciados durante o primeiro ano do projeto. Sendo que o objetivo final é se-

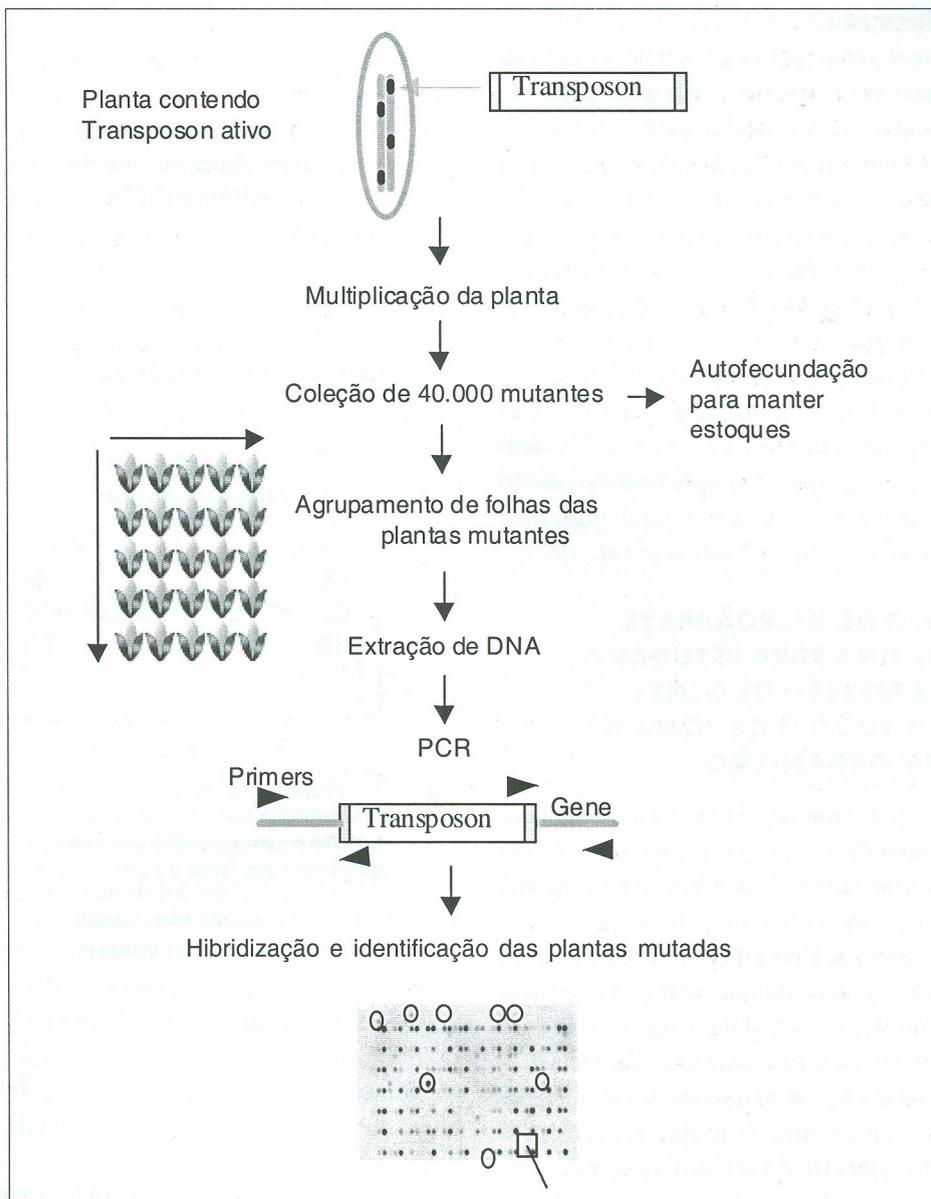


Figura 6 - Trait Utility System in Corn (TUSC)

NOTA: Plantas contendo transposons ativos são multiplicadas para a montagem de uma biblioteca de mutantes. Conhecendo a frequência com que o transposon se multiplica e insere em regiões codantes, pode-se calcular teoricamente o número necessário de plantas, para ter um transposon em cada gene. Cada planta mutante é autofecundada e as sementes estocadas. Uma reação de PCR é feita com DNA extraído de folhas de grupos de plantas dessa biblioteca de mutantes, usando um primer do transposon e um primer do gene (a seqüência de ambos é conhecida). A hibridização é feita para auxiliar no escrutínio de um grande número de plantas. Os grupos de plantas cujo sinal foi positivo são subdivididos até que seja encontrada a planta que contenha o transposon inserido no gene. Para aumentar a chance de encontrar a planta mutante, testa-se uma série de primers de um único gene, simultaneamente.

qüenciar cerca de 150 mil segmentos de DNA genômico.

Mutação por inserção de T-DNA

O T-DNA é um segmento de DNA presente no plasmídeo Ti (DNA não-cromossômico) de *Agrobacterium tumefaciens*,

uma bactéria de solo que causa tumores, quando infecta plantas. O tumor é causado pela capacidade da bactéria em transferir seu T-DNA para as células vegetais. O T-DNA contém genes que codificam hormônios e aminoácidos necessários para a sobrevivência da agrobactéria dentro das plantas. Cientistas descobriram que podiam

alterar este fragmento de DNA substituindo os genes nativos por qualquer gene de interesse, e que esses novos genes continuariam sendo transferidos para as plantas pelo sistema mediado pela bactéria. A partir dessa descoberta, o T-DNA passou a ser usado como uma ferramenta nas genéticas direta e reversa de maneira semelhante aos transposons. O T-DNA insere-se aleatoriamente no meio de genes tornando-os não-funcionais e produzindo um organismo mutante (Fig. 8). A função do gene *Agamous* de *Arabidopsis* foi caracterizada como sendo um regulador transcricional necessário para o desenvolvimento floral, utilizando essa metodologia (Yanofsky et al., 1990).

USO DE MICROARRAYS DE DNA PARA ESTUDAR A EXPRESSÃO DE GENES EM TODO O GENOMA DE UM ORGANISMO

O *microarray* é uma metodologia utilizada para comparar a expressão de um grande número de genes simultaneamente. Este sistema baseia-se na impressão de milhares de clones de cDNAs em placas de vidro através de um robô e, subsequentemente, de uma hibridização com duas sondas marcadas com fluorescências diferentes, em geral uma que emite uma cor vermelha e outra verde. As sondas são conjuntos de cDNAs gerados a partir de células ou tecidos em duas situações diferentes que se deseja comparar (resistência e suscetibilidade ao alumínio, por exemplo). Os resultados são produzidos sob forma de diferentes intensidades de fluorescência que são captadas por microscopia de fluorescência a laser, em função dos diferentes níveis de expressão de cada gene. A imagem dos pontos fluorescentes é processada por computadores e programas específicos, sendo gerada uma grande quantidade de informação simultaneamente (Fig. 9, p.63).

Um robô é capaz de colocar 16 amostras em 48 placas de vidro, lavar e secar a sonda para o próximo grupo de cDNA em 70 segundos. Assim, cerca de 10 mil amostras podem ser impressas em aproximadamente 12 horas (Schena et al., 1995). Vários sis-

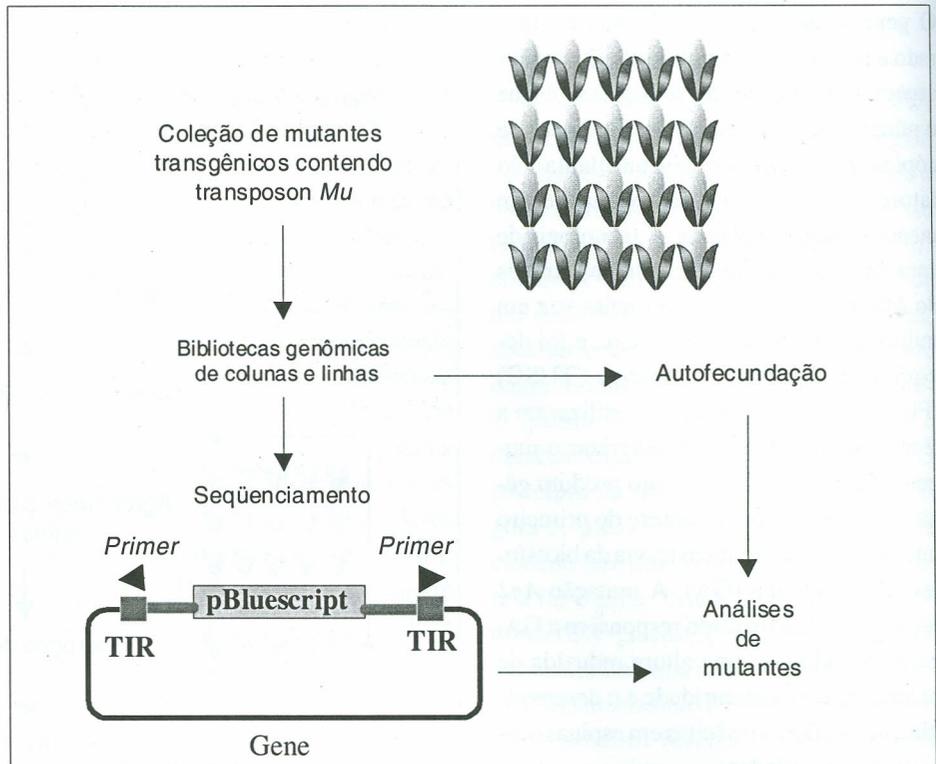


Figura 7 - Projeto NSF coordenado pela Dra. Virginia Walbot (University of Stanford)

NOTA: O transposon contendo uma marca de seleção para resistência à ampicilina em bactéria é transferido para o milho por transformação. As plantas são multiplicadas e as bibliotecas genômicas são feitas a partir de colunas e fileiras de plantas que contêm esses transposons engenheirados inseridos aleatoriamente no genoma. Os plasmídeos originados da biblioteca genômica contendo esses cassetes são seqüenciados e as plantas contendo os transposons inseridos nos genes de interesse são identificadas.

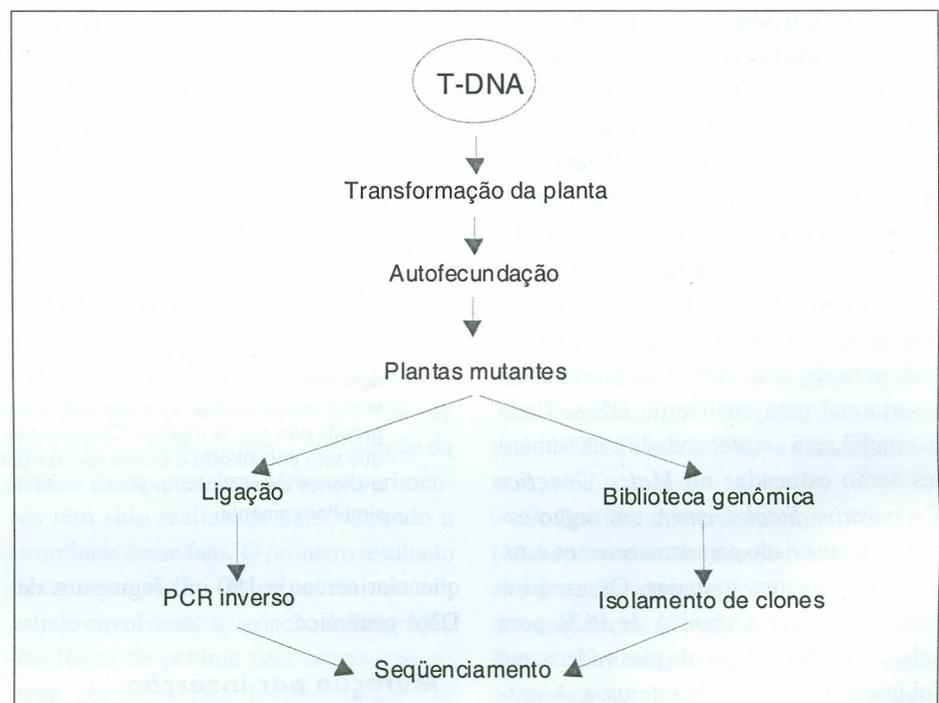


Figura 8 - T-DNA na produção de uma biblioteca de mutantes

NOTA: Plantas de *Arabidopsis* são transformadas por agrobactéria contendo o T-DNA. Essas plantas são autofecundadas e os mutantes analisados. O T-DNA contendo uma marca de seleção é usado para criar bibliotecas genômicas. Os plasmídeos provenientes dessa biblioteca são seqüenciados ou mesmo usados em reação de PCR com *primers* específicos do gene e do T-DNA.

temas robotizados mais rápidos e eficientes para a impressão e processamento dos cDNAs nas placas de vidro, além de metodologias cada vez mais sensíveis e precisas para a detecção e análise dos resultados, têm sido desenvolvidos e disponibilizados para os pesquisadores.

A tecnologia de *microarrays* não fornece apenas informações sobre a função de genes anônimos, o que favorecerá bastante os processos de genética reversa, mas também constitui uma ferramenta indispensável para estudos globais de expressão gênica, com grandes aplicações nos estudos de Biologia molecular e Fisiologia vegetal.

PROTEOMICS - ESTUDO DO PERFIL PROTÉICO DE UM ORGANISMO

Cientistas têm utilizado o termo *proteomics* para descrever as análises do perfil protéico de um organismo. O comportamento de muitas proteínas em um proteoma pode ser monitorado por meio de géis de eletroforese em duas dimensões. Nesse processo, proteínas provenientes de células ou tecidos em duas condições fisiológicas diferentes (por exemplo, plantas resistentes e sensíveis a uma determinada doença ou condição de estresse) são extraídas, aplicadas em géis e comparadas qualitativa e quantitativamente. As proteínas presentes em

apenas um dos géis ou que tenham a sua quantidade alterada são fortes candidatas para atuarem no controle do processo em estudo. A partir da identificação das proteínas de interesse pode-se isolar sua sequência gênica e caracterizá-la, utilizando as técnicas de Biologia molecular.

LOCALIZAÇÃO CELULAR DE UM mRNA OU PROTEÍNA PARA INFERIR SOBRE SUA FUNÇÃO

Os genes são transcritos em moléculas mRNA que são traduzidas em proteínas. Essas são transportadas para locais específicos dentro da célula de um determinado tecido ou organismo. Enquanto alguns genes codificam para proteínas que são expressas em todas as células de um indivíduo durante toda a sua vida (genes constitutivos), outros codificam para proteínas que são expressas apenas em algumas células, ou em um determinado período de desenvolvimento, ou em resposta a algum estímulo externo (genes tecidos-específicos). A localização de uma proteína ou de seu mRNA nas células ou tecidos onde são produzidos é uma importante fonte de informação que pode auxiliar na determinação da função de um gene ou proteína. Um dos princípios dessa metodologia é a identificação do mRNA ou da proteína de interesse por meio de sondas

específicas. As sondas, após se ligarem à proteína ou ao mRNA em estudo, são detectadas com o auxílio de técnicas de microscopia. Uma vez conhecida a localização da molécula dentro de um organismo, podem ser programados experimentos mais refinados para a definição de sua função. Técnicas de hibridizações *in situ* têm sido amplamente utilizadas para localizar genes em cromossomos, tecidos e células de diversos organismos.

INTERAÇÃO PROTEÍNA-PROTEÍNA

A caracterização da função gênica pode ser auxiliada por meio de estudos de interação proteína-proteína *in vitro* e *in vivo*. A determinação de que uma proteína, cuja função desconhecida interage com uma que é conhecida, pode sugerir que as duas proteínas fazem parte do mesmo processo ou podem estar localizadas no mesmo compartimento celular. Estes testes podem ser usados para verificar a existência de interação entre duas proteínas conhecidas, identificar tanto as novas proteínas que interagem com uma proteína em estudo quanto as regiões dentro de uma proteína que são importantes no processo de reconhecimento e da interação. Existe um grande número de alternativas e variações para estudar as interações entre proteínas (Fig. 10). Apesar de bastante informativas,

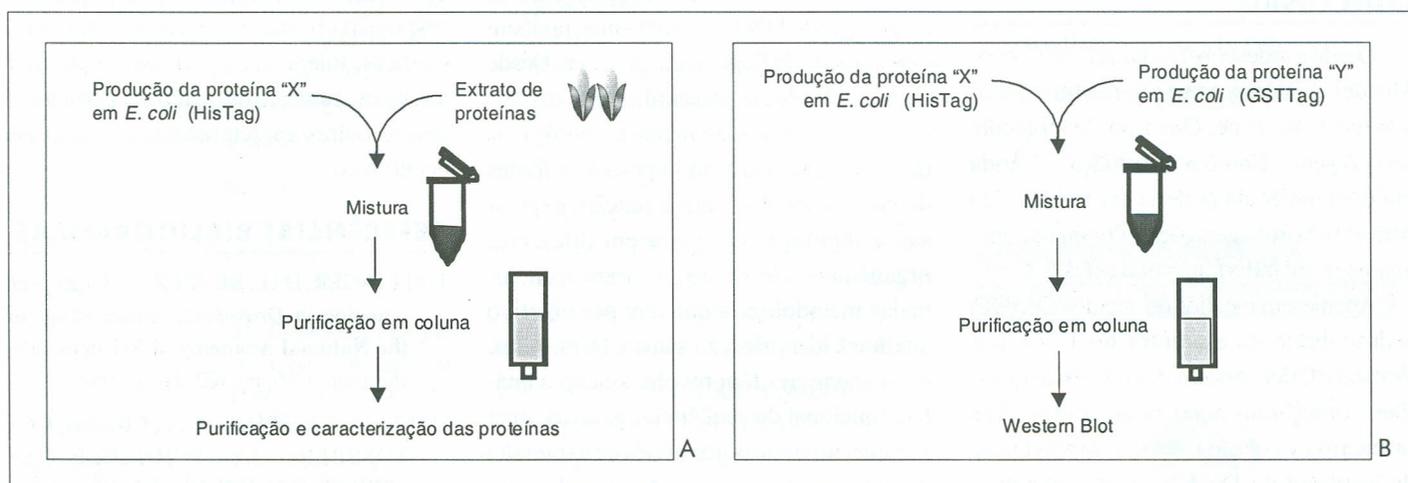


Figura 10 - Identificação de interação proteína-proteína (*in vitro*)

NOTA: A - A proteína de interesse é clonada em um vetor de expressão em bactéria que permite a sua purificação em uma coluna de afinidade. Um extrato de proteínas produzido nas células vegetais é também submetido à coluna que já contém fixada a proteína de interesse. A coluna é lavada e as proteínas que interagem com a proteína de interesse podem ser sequenciadas ou usadas para produzir anticorpos que serão utilizados em uma biblioteca de expressão de cDNA; B - Proteínas "X" e "Y" produzidas em *E. coli* são misturadas e passadas através de uma coluna de níquel. Se a proteína "Y" interagir com a "X", ambas ficarão retidas na coluna. O processo pode ser visualizado por anticorpos ou raios X, caso a proteína "Y" esteja marcada com radioatividade.

as reações *in vitro* podem mascarar os resultados reais, devido à impossibilidade de reproduzir *in vitro* todas as condições existentes nas células *in vivo*. Por esse motivo foram desenvolvidos protocolos que estudam a interação entre proteínas *in vivo*.

Uma maneira efetiva para detecção de interação *in vivo* é utilizar o sistema híbrido de levedura. Esse processo constitui na construção de duas proteínas de fusão por Engenharia genética. Uma delas gera um híbrido entre seqüências para um domínio que se liga ao DNA do fator de transcrição Gal4 (aminoácidos 1-147) (Keegan et al., 1986) e a proteína de interesse. Um segundo plasmídeo de expressão contém uma seqüência que ativa o fator de transcrição Gal4 (aminoácidos 768-881) (Ma & Ptashne, 1987), fundida com cDNAs. Estes são provenientes de bibliotecas de onde possam existir genes candidatos. Desta forma, se as duas proteínas expressas na levedura são capazes de interagir, o complexo resultante ativará a transcrição dos promotores contendo sítios de ligação para o Gal4, gerando uma colônia azul em meio contendo um substrato apropriado (X-GAL). O sucesso dessa metodologia foi primeiramente demonstrado pela interação Gal4-Gal80 (Ma & Ptashne, 1988), sendo mais tarde generalizado por Fields & Song (1989) (Fig. 11).

CONCLUSÃO

Desde a redescoberta das leis de Gregor Mendel, cientistas começaram a questionar a natureza do gene. Que tipo de molécula seria o gene? Como a informação contida em uma molécula poderia ser transmitida para as próximas gerações? Por que e como apareceriam indivíduos mutantes?

Apenas em meados do século XX, com a descoberta da estrutura do DNA por Watson e Crick, essas e muitas outras questões começaram a ser respondidas. Nas primeiras três décadas após a compreensão da estrutura do DNA, o conhecimento a respeito de sua biologia cresceu fantásticamente. Dentro desse período ficou entendida a natureza química dos genes, como a informação genética era armazenada,

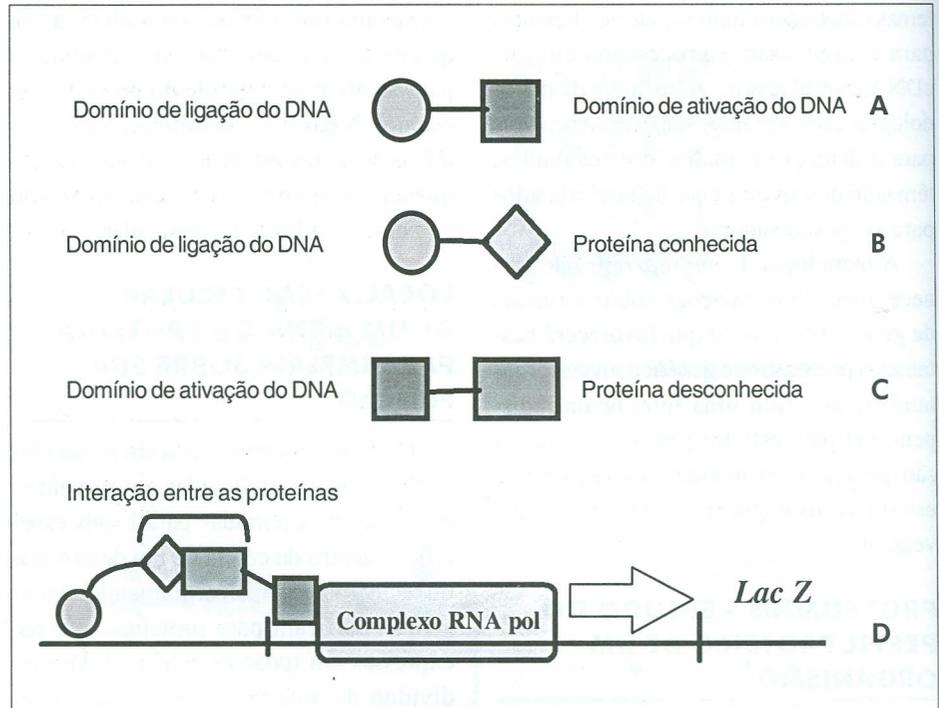


Figura 11 - Interação proteína-proteína *in vivo*

NOTA: A - O ativador de transcrição GAL4 possui dois domínios: um que se liga ao DNA e outro que ativa a transcrição; B e C - Os dois domínios foram separados: o primeiro pode ser fundido à proteína de interesse (B) e, o segundo, à proteína desconhecida (C); D - As colônias de levedura contendo os dois plasmídeos, cujas proteínas interagem serão azuis em meio contendo X-GAL (substrato para a enzima), devido à reconstituição do fator de transcrição GAL4 e à ativação do gene da beta-galactosidase (*lac Z*).

como as células respondiam a essa informação e como ela era transmitida de uma geração para outra.

A partir dos anos 70, cientistas começaram a aprender como manipular a molécula de DNA utilizando as técnicas de Biologia molecular. Inaugurou-se a era da tecnologia do DNA recombinante, também denominada de Engenharia genética. Desde então, a Biologia molecular tem experimentado grandes avanços tecnológicos que têm culminado em importantes fontes de conhecimento sobre a função, expressão e regulação de genes em diferentes organismos. Neste artigo foram mencionadas metodologias que têm por objetivo auxiliar a identificação gênica. Dentre elas, os *microarrays* têm revolucionado a análise funcional de seqüências gênicas, uma vez que, diferenças nos níveis de expressão de milhares de genes, podem ser detectadas simultaneamente. Desta forma, vários genes ainda desconhecidos poderão ter suas funções biológicas desvendadas utilizando esse processo. As metodologias

de análise gênica, tanto individual quanto em larga escala, fornecerão um acesso a informações sem precedentes para todas as áreas da Biologia. Dentre elas, a agropecuária será amplamente beneficiada, uma vez que existe uma grande necessidade de identificar e estudar genes que são responsáveis por conferir resistência a doenças, tolerância a estresses bióticos e abióticos, aumento da qualidade nutricional dentre outras características de interesse econômico.

REFERÊNCIAS BIBLIOGRÁFICAS

- BALLINGER, D.G.; BENZER, S. Target gene mutation in *Drosophila*. **Proceedings of the National Academy of Sciences**, Washington, v.86, p.9402-9406, 1989.
- BENSEN, R.J.; JOHAL, G.S.; CRANE, V.C.; TOSSBERG, J.T.; SCHNABLE, P.S.; MEELEY, R.B.; BRIGGS, S.P. Cloning and characterization of the maize An1 gene. **Plant Cell**, Rockville, v.7, p.75-84, 1995.
- EMMONS, S.; YESNER, L. High frequency excision of transposon element Tc1 in the

- nematode *Caenorhabditis elegans* is limited to somatic cells. **Cell**, Cambridge, v.36, p.599-605, 1984.
- FEDEROFF, N.V.; FURTEK, D.B.; NELSON, O.E. Cloning of the bronze locus in maize by a simple and generalized procedure using the transposable element Activator (Ac). **Proceedings of the National Academy of Sciences**, Washington, v.81, p. 3825-3829, 1984.
- FIELDS, S.; SONG, O. A novel genetic system to detect protein-protein interactions. **Nature**, London, v.340, p.245-246, 1989.
- GOFFEAU, A.; BARRELL, B. G.; BUSSEY, H.; DAVIS, R. W.; DUJON, B.; FELMANN, H.; GALIBERT, F.; HOHENSEL, J. D.; JACQ, C.; JOHNSTON, M.; LOUIS, E.J.; MEWES, H.W.; MURAKAMI, Y.; PHILLIPSEN, P.; TETTELIN, H.; OLIVER, S.G. Life with 6000 genes. **Science**, New York, v.274, p. 546-567, 1996.
- KEEGAN, L.; GILL, G.; PTASHNE, M. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. **Science**, New York, v.231, p. 699-704, 1986.
- KROL, A.R. van der; MOL, J.N.; STUITJE, A.R. Antisense genes in plants: an overview. **Gene**, Amsterdam, v.72, p. 45-50, 1988.
- KEMPIM, S.A.; LILJEGREN, S.J.; BLOCK, L.M.; ROUNSLEY, S.D.; YANOFSKY, M.F. Target disruption in *Arabidopsis*. **Nature**, London, v. 389, p. 802-803, 1997.
- LIST OF ESTs (Express Sequences Tags). Disponível site **National Center for Biotechnology Information**. URL: <http://www.ncbi.nlm.nih.gov/dbEST/dbEST-summary.html> Consultado em nov. 1999.
- MA, J.; PTASHNE, M. Converting a eukaryotic transcriptional inhibitor into an activator. **Cell**, Cambridge, v. 55, p. 433-446, 1988.
- MA, J.; PTASHNE, M. Detection analysis of GAL4 defines two transcriptional activating segments. **Cell**, v.48, p. 847-853, 1987.
- MELTON, D.W. Gene targeting in the Mouse. **Bioessays**, Cambridge, v.16, p.633-638, 1984.
- NAPOLI, C.; LEMIEUX, C.; JORGENSEN, R. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. **Plant Cell**, Rockville, v.2, p. 279-289, 1990.
- O'HARE, K. Searching for a needle in haystacks via the polymerase chain reaction. **Theoretical and Applied Genetics**, New York, v.6, p.202-203, 1990.
- PLASTERK, R.H.A. Reverse genetics of *Caenorhabditis elegans*. **Bioessays**, Cambridge, v.14, p. 629-633, 1993.
- ROSENZWEIG, B.; LIAO, L.W.; HIRSH, D. Target sequences for the *C. elegans* transposable element Tc1. **Nucleic Acids Research**, London, v.11, p. 7137-7140, 1983.
- RUSHFORTH, A.M.; SAARRI, B.; ANDERSON, P. Site-selected insertion of the transposon Tc1 into *Caenorhabditis elegans* myosin light chain gene. **Molecular and Cellular Biology**, Washington, v.13, p. 902-910, 1993.
- SCHENA, M.; SHALON, D.; DAVIS, R.W.; BROWN, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science**, New York, v. 270, p. 467-470, 1995.
- SHEEHY, E.R.; KRAMER, M.; HIATT, W.R. Reduction of polygalacturonase activity in tomato fruit by antisense RNA. **Proceedings of the National Academy of Sciences**, Washington, v.85, p. 8805-8809, 1988.
- WALBOT, V. Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 43, p. 49-82, 1992.
- YANOFSKY, M.F.; MA, H.; BOWMAN, J.L.; DREWS, G. N.; FELDMANN, K. A.; MEYEROWITZ, E.M. The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. **Nature**, London, v. 346, p. 35-39, 1990.
- ZWAAL, R.R.; BROEKS, A.; MEURS, J. van; GROENEN, J.T.; PLASTERK, R.H. Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. **Proceedings of the National Academy of Sciences**, Washington, v.90, p. 7431-7435, 1993.