

Genoma milho: o estado da arte

Claudia Teixeira Guimarães

Embrapa Milho e Sorgo, CP 151, Sete Lagoas, MG 35701-970 claudia@cnpms.embrapa.br

Introdução

A biotecnologia vegetal teve início com o aparecimento das técnicas de cultura de tecidos, objetivando a propagação de plantas. Essa tecnologia tem sido amplamente utilizada na multiplicação de plantas em larga escala, no manejo e conservação de germoplasma, na transferência de genes eliminando barreiras de incompatibilidade genética, na obtenção de plantas haplóides, e na eliminação de doenças através do cultivo "in vitro" de meristemas. Em meados dos anos 80, a tecnologia de marcadores moleculares possibilitou o mapeamento e a identificação de regiões genômicas que controlam características de importância agrônômica, e a sua rápida transferência para outros genótipos. Além disso, os marcadores moleculares têm sido utilizados na caracterização de genótipos, no estudo da diversidade genética, na análise da pureza de sementes e em processos de melhoramento assistido. O advento da tecnologia de DNA recombinante, a possibilidade de se transformar plantas e o grande volume de informações de seqüências de DNA têm disponibilizados o acesso a um novo e variado conjunto de genes e de tecnologias, possibilitando a transferência dos mesmos além dos limites permitidos pelo melhoramento tradicional. Assim, o desenvolvimento de cultivares cada vez mais produtivas e adaptadas às mais diversas condições de cultivo, pode ser dramaticamente acelerado com a utilização de técnicas de mapeamento, manipulação gênica e transformação.

As primeiras plantas transgênicas obtidas por engenharia genética começaram a ser liberadas no campo em meados de 1990. Atualmente, já foram autorizados mais de 25 mil testes de campo no mundo, metade deles nos Estados Unidos, no Canadá e na Europa. Na América Latina, o maior número de liberações ocorreu na Argentina e México. O Brasil promoveu 800 liberações para teste de plantas transgênicas a partir de 1996, tendo sua legislação de biossegurança aprovada em 1995. A comercialização de plantas transgênicas começou no final da década de 1990. Atualmente, soja, milho e canola geneticamente modificados já têm participação relevante na agricultura dos Estados Unidos, Canadá e Argentina, sendo que a área cultivada deve atingir 40 milhões de hectares no início do próximo século. Tais plantas têm como características a resistência a insetos, vírus, herbicidas e uma melhor qualidade nutricional. Uma lista contendo os eventos de milho transgênico produzidos pelas principais empresas produtoras de milho, e seus respectivos genes de resistência e tolerância estão disponíveis na página da Comissão Técnica Nacional de Biossegurança - CTNBio (http://www.mct.gov.br/ctnbiotec/tab_milho.htm).

A Organização do Genoma do Milho

O genoma do milho é comparável ao tamanho do genoma humano, girando em torno de 2.500 Mb, no entanto, sua complexidade tende a ser superior devido à abundância de elementos altamente repetitivos. Grande parte do genoma do milho é composto por seqüências repetitivas, sendo que 65 a 70% é representado por retrotransposons, cujos elementos transponíveis são mediados por RNA (Bennetzen, 1996). Tais retrotransposons

representam poucos membros com elevada identidade de seqüência entre si, onde cada classe é altamente amplificada, variando de 10 a 30.000 cópias, distribuídas ao longo de todo o genoma do milho (SanMiguel et al., 1996). Elementos transponíveis de seqüências repetidas e invertidas, DNA satélite, repetições específicas de centrômero e telômero representam de 15 a 20% da porção que não correspondem aos genes. Os transposons de repetições invertidas possuem de dezenas a centenas de cópias e estão intimamente associados aos genes. Finalmente, é estimado que os genes representem no máximo 15% do genoma, sendo estes compostos por pequenos introns, possuindo uma região transcrita em média de 4 a 5 Kb e, os genes ativos apresentam padrões sub-metilados quando comparados com as seqüências repetitivas.

Filogenia

O milho é uma das espécies cultivadas mais importante economicamente, sendo também uma das mais bem estudadas, considerando a sua complexidade genômica. O mapeamento comparativo tem revelado uma extensiva conservação no conteúdo e na ordem dos genes entre os genomas de gramíneas como milho, sorgo, arroz, trigo, cevada e cana-de-açúcar. No entanto, rearranjos genéticos como inversões, duplicações e transposições são freqüentemente observados entre as gramíneas, sendo sugerido que, pelo menos 15.000 rearranjos cromossômicos diferenciaram o arroz do milho (Tikhonov et al., 1999; Dubcovsky et al., 2001). Violações na colinearidade de genes entre linhagens de milho também tem sido apresentadas (Fu & Dooner, 2002). Tais revelações indicam que o seqüenciamento de BACs (biblioteca artificial de bactéria) e do genoma do milho pode aumentar significativamente o conhecimento sobre a organização e a função gênica em milho, além de servir como modelo para as espécies de gramíneas de genoma mais complexo com centeio e cana-de-açúcar.

Mapas Genéticos

A disponibilidade de um mapa de ligação densamente saturado com marcadores e com elevado poder de resolução genética é de fundamental importância para a integração de mapas genéticos e físicos, e para o mapeamento de genes de interesse. O primeiro mapa genético de milho baseado em marcadores moleculares foi publicado por Helentjaris et al. (1986), sendo composto por 116 locos RFLPs (polimorfismo de comprimento de fragmentos de restrição). A partir de então, uma série de mapas genéticos cada vez mais saturados pelos diferentes tipos de marcadores moleculares, utilizando diferentes progenitores e populações de mapeamento, têm sido publicados, objetivando principalmente a identificação de caracteres de locos quantitativos (QTLs). O mapa UMC 98, publicado por Davis et al. (1999), além de ser altamente saturado com 1736 marcadores RFLPs e SSRs, 962 locos foram seqüenciados e identificados como genes estruturais e fatores regulatórios responsáveis por vários processos metabólicos importantes. Dentre os marcadores, 273 são sondas heterólogas, originadas de arroz, trigo, aveia e tripsacum, e grande parte dos mesmos são comuns a vários mapas genéticos de gramíneas, incluindo o mapa de arroz do Projeto Japonês do Genoma de Arroz (Harushima et al., 1998). Considerando-se todas as características, o referido mapa genético constitui uma poderosa fonte de informações para estudos da organização e evolução de genes e genomas de gramíneas, clonagem de genes e avaliação de características quantitativas. Além de possibilitar a integração dos conhecimentos genéticos, bioquímicos, funcionais e fenotípicos.

Apesar das grandes vantagens do mapa UMC 98, algumas regiões de distância física reduzida e grandes distâncias genéticas, e regiões telomérica foram pouco saturadas com

marcadores. Provavelmente, o fato tenha ocorrido devido ao tamanho reduzido da população de mapeamento, onde pouco recombinantes seriam esperados. Para aumentar a resolução dos mapas genéticos, uma população constituída de 304 linhagens recombinantes endogâmicas (RILs) foi obtida a partir de quatro ciclos de intercruzamentos ao acaso após a geração F₂. Os progenitores foram as linhagens B73 e Mo17, que representam os dois maiores grupos heteróticos temperado, e a população foi denominada IBM ("Intermated B73 x Mo17") (Lee et al., 2002). As recombinações adicionais ocasionadas pelo intercruzamento resultou em uma expansão três vezes a distância genética, que associada ao mapeamento de mais de 1850 marcadores mapeados, aumentando significativamente a resolução e o potencial de utilização do mapa genético. Todas as informações relacionadas com a população IBM e com os locos mapeados estão disponíveis no *Maize Genome Database* (www.agron.missouri.edu) e no *Maize Mapping Project* (www.maizemap.org). Adicionalmente, encontram-se disponíveis, para qualquer pesquisador, um kit contendo DNA de 94 linhagens da população IBM e sementes dessas linhagens, possibilitando o mapeamento de genes, clones ou fenótipos de interesse.

Mapas Físicos

A primeira etapa da construção de um mapa físico é a obtenção de bibliotecas genômicas em BACs. Em milho, a linhagem selecionada para a construção das bibliotecas de BACs publicamente disponíveis foi a B73, principalmente pelo fato de ser um dos progenitores da população IBM, o que possibilita a conexão entre mapas genéticos e físicos. As bibliotecas foram feitas utilizando três enzimas de restrição diferentes (*EcoRI*, *HindIII* e *MboI*) para assegurar a máxima cobertura do genoma, representando uma cobertura de 27 vezes o genoma do milho (Coe et al., 2002).

A organização das bibliotecas de BACs requer um processo de genotipagem dos clones e sua posterior análise usando o programa Fingerprinting Contigs (FPC), que gera automaticamente a montagem dos *contigs* (Soderlund et al., 2000). Para aumentar a precisão e a rapidez com que os *contigs* são montados, é importante que as bibliotecas de BACs sejam genotipadas com vários tipos de marcadores moleculares, principalmente aquele que já possuem posições definidas nos mapas genéticos de milho. Assim, a genotipagem das BACs tem sido realizada com marcadores RFLPs que ancoram os *bins* nos mapas de milho, coleções de ESTs, sondas genômicas e de cDNA de sorgo, AFLPs, marcadores baseados em PCR e os MITEs (miniaturas de elementos transponíveis de repetições invertidas).

Seqüenciamento Genômico e de Genes

O seqüenciamento de genes pela iniciativa pública e privada vem causando grande impacto na comunidade científica, e para a cultura do milho, tal realidade não tem sido diferente. O grande número de seqüências de genes expressos ou ESTs (*Expressed Sequences Tagged*) disponíveis para vários organismos tem dado suporte para a chamada "genética reversa" onde, a partir das informações sobre a seqüência de um gene, procura-se então identificar a característica controlada por ele. Assim, várias metodologias têm sido desenvolvidas e utilizadas para a identificação das funções gênicas, também conhecida como genética funcional. As seqüências dos genes expressos são depositadas em bancos de dados de domínio público, como o NCBI (<http://www.ncbi.nlm.nih.gov/dbEST/>) ou o *Maize Genome Database* (http://nucleus.agron.missouri.edu/cgi-bin/est_summary.pl).

O Projeto de Seqüenciamento Genômico de Milho, em função da sua complexidade, têm

sido direcionado às regiões ricas em genes e com reduzido número de cópias, objetivando o seqüenciamento de aproximadamente 50.000 genes, representando de 10 a 15% do genoma. Um dos pontos importantes do projeto genoma será a possibilidade de associar as informações das seqüências gênicas com os mapas genéticos e físicos. Com isso, o conhecimento sobre o conteúdo e a organização gênica no genoma do milho poderão ser significativamente ampliados, contornando uma séria limitação aos avanços nas pesquisas básicas e aplicadas nessa importante cultura. O seqüenciamento dos genes e a elucidação das suas funções terão impactos na compreensão das bases moleculares de caracteres agronomicamente importantes, auxiliando os processos de melhoramento genético de milho e de outras espécies de gramíneas (Bennetzen et al., 2001).

Estratégias para Identificação de Função Gênica em Milho

Transposons

Transposons são elementos de DNA que têm a capacidade de sair de uma região do genoma e se incorporar em outra. Como o movimento dos transposons é um processo aleatório, os mesmos podem se inserir na região codificadora de um gene tornando-o inativo. Essa capacidade dos transposons em causar mutações fenotípicas tem sido explorada na identificação genes associados com fenótipos de importância agrônômica. O transposon *Mu* (*Mutator*) foi utilizado pela primeira vez em milho pela Pioneer Hi-Bred Co., e a tecnologia foi denominada *Trait Utility System* (TUSC). Atualmente, um projeto usando transposon *Mu* foi financiado pela *National Science Foundation* (NSF) sob a coordenação da Dra. Virginia Walbot e com a participação de várias universidades americanas. O referido projeto tem como objetivo demonstrar a funcionalidade de genes de milho usando o seqüenciamento do DNA genômico flanqueando as inserções do transposon *Mu* e caracterizar os indivíduos mutantes contendo esses transposons. Para uma ampla utilização do material genético resultante desse projeto, foi criado um banco de mutantes onde são armazenados os estoques de sementes mutadas, o *Maize Genetics Cooperative Stock Center*. Os usuários poderão utilizar a técnica de PCR para seleção de uma coleção de plasmídeos contendo genes de interesse flanqueados pelo transposon *Mu*.

Microarrays

O *microarray* é uma metodologia utilizada para comparar a expressão de um grande número de genes, simultaneamente. Milhares de clones de cDNAs, contendo um grupos de genes que se deseja quantificar os níveis de expressão, são roboticamente impressos em placas de vidro e, então, hibridizados com duas sondas marcadas com fluorescências diferentes. As sondas são conjuntos de cDNAs gerados a partir de células ou tecidos em duas situações diferentes que se desejam comparar, por exemplo resistência e suscetibilidade ao alumínio. Os resultados são produzidos sob forma de diferentes intensidades de fluorescência que são captadas por microscopia de fluorescência à 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. Vários sistemas robotizados, cada vez mais rápidos e eficientes, têm sido desenvolvidos para a impressão e o processamento dos cDNA nas placas de vidro, permitindo uma detecção e análise cada vez mais sensível e precisa dos genes diferencialmente expressos dentro de um genoma ou de um grupo de genes selecionados. Assim, a tecnologia de microarrays apresentam-se como uma ferramenta indispensável para estudos globais de expressão gênica, com grandes aplicações nos estudos de biologia molecular e fisiologia vegetal.

Novas Perspectivas para o Milho

Milho para produção de fármacos, nutraceuticos e biopolímeros

Além das melhorias nas qualidades agrônômicas como, resistência a pragas e doenças, tolerância a estresses abióticos e a herbicidas, e o aumento da qualidade protéica do grão, a transgenia tem trazido várias possibilidades interessantes para o milho. O milho têm sido utilizado para produção de fármacos, nutraceuticos e polímeros biodegradáveis. O endosperma constitui 80% do grão de milho e tem despertado especial interesse devido à sua importância como fonte de energia e proteínas na alimentação humana e animal. Além do grão do milho apresentar em média 9% de proteína na matéria seca, este apresenta uma baixa umidade, em torno de 14%, constituindo uma forma eficiente de armazenamento de proteínas, anticorpos e outros peptídeos de interesses à saúde humana e animal como, hormônio de crescimento, insulina, vitaminas e vacinas.

Milho tolerante ao ataque de insetos

No milho geneticamente engenheirado com a toxina Bt, ela é produzida pela própria planta onde a lagarta se alimenta. A toxina do Bt não é tóxica para humanos, gado, insetos benéficos e outros animais, mas na lagarta do milho a toxina destrói tecidos do intestino, causando a morte das mesmas. As lavouras de milho Bt têm a capacidade de reduzir a aplicação de inseticidas químicos, reduzindo consequentemente os custos de produção e os riscos de contaminação ambiental. Na natureza o baculovírus pode infectar larvas de insetos-praga do milho, utilizando-as para se multiplicar e espalhar. No entanto, as larvas infectadas não morrem imediatamente e continuam a se alimentar da planta. Assim, alternativas têm sido estudadas para alterar geneticamente o baculovírus com gene que codifica um hormônio que diminui o apetite das larvas, matando-as por anorexia. A superexpressão desse gene no baculovírus causa um desequilíbrio hormonal que leva a morte da larva por falta de apetite e perda de água. A vantagem do vírus é que, fora do inseto ele é destruído pela luz ultravioleta, tornando-o inofensivo ao homem, outros animais e plantas.

A maisina é um composto da via metabólica que origina o cabelo da espiga (estilo-estigma) e que tem sido responsável por promover uma resistência natural a insetos, pelo fato de se ligar aos aminoácidos no intestino do inseto, impedindo a sua utilização e matando a larva do inseto. Os genes que regulam mais da metade da maisina produzida já foram identificados, e a superexpressão dos mesmos têm sido testada em plantas transgênicas e seu efeito nas larvas. O aumento das concentrações de maisina no cabelo da espiga do milho poderia reduzir o uso de inseticida em cerca de 85%, além da possibilidade de aplicação desses genes em alfalfa, algodão, amendoim e tomate.

Uso de marcadores moleculares no melhoramento de milho

Os avanços na área de biotecnologia têm possibilitado a utilização efetiva de marcadores moleculares no melhoramento genético de plantas, na caracterização de molecular de cultivares e linhagens, na determinação de pureza genética de sementes, no mapeamento e identificação de genes e QTLs associados às características de interesse. O objetivo geral na Embrapa Milho e Sorgo é utilizar essa tecnologia como auxílio efetivo na solução de problemas prioritários para a cultura do milho e do

sorgo nas condições tropicais do Brasil. Assim, vários projetos vêm sendo desenvolvidos visando acelerar os programas de retrocruzamento para conversão de linhagens elite, suporte aos processos de proteção intelectual das linhagens elites, avaliações de pureza genética em lotes de sementes básicas fornecidas pela Embrapa e a identificação de marcadores associados com características de importância agrônômica que possam ser utilizados tanto nos programas de melhoramento genético, quanto na clonagem de genes e no estudos dos mecanismos envolvidos. Dentre as principais características estudadas, podem ser citadas a melhoria da qualidade nutricional dos grãos, tolerância a estresses bióticos e abióticos visando a maior adaptação das cultivares aos solos ácidos, deficiência hídrica e nutricional, ataque de pragas e resistência a doenças.

Prospecção de genes em outras gramíneas

A Embrapa Milho e Sorgo vem procurando identificar genes e seqüências regulatórias na espécies cultivadas de maior importância econômica como o milho e sorgo, além de outras gramíneas como o milheto, tripsacum e braquiária. Considerando o alto grau de sintenia entre os genomas das gramíneas, o estudos dos genes responsáveis pelos mecanismos de tolerância e seus elementos regulatórios podem fornecer informações importantes sobre a conservação desses genes durante a evolução das espécies, além da identificação de novos alelos. Tais alelos poderão ser utilizados tanto em programas de melhoramento genético convencional quanto na produção e avaliação de plantas transgênicas.

Referências Bibliográficas

- Bennetzen, J.L. (1996) The contributions of retroelements to plant genome organization, function and evolution. *Trends Microbiol.* 4: 347-353.
- Bennetzen, J.L.; Chandler, V.L.; Schnable, P. (2001) National science foundation-sponsored workshop report: maize genome sequencing project. *Plant Physiol.* 127: 1527-1578.
- Coe, E.; Cone, K.; McMullen, M.; Chen, S.-S.; Davis, G.; Gardiner, J.; Lisum, E.; Polacco, M.; Paterson, A.; Sanchez-Villeda, H.; Soderlund, C.; Wing, R. (2002) Access to the maize genome: an integrated physical and genetic map. *Plant Physiol.* 128: 9-12.
- Davis, G.L.; McMullen, M.D.; Baysdorfer, C.; Musket, T.; Grant, D.; Staebell, M.; Xu, G.; Polacco, M.; Koster, L.; Melia-Hancock, S.; Hochins, K.; Chao, S.; Coe, E.H. (1999) A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. *Genetics*, 152: 1137-1172.
- Dubcovsky, J.; Ramakrishna, W.; SanMiguel, P.J.; Busso, C.S.; Yan, L.; Shiloff, B.A.; Bennetzen, J.L. (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosome. *Plant Physiol.* 125: 1342-1353.
- Fu, H.; Dooner, H.K. (2002) Intraspecific violation of genetic colinearity and its implication in maize. *Proc. Natl. Acad. Sci. USA* 99:
- Harushima, Y.; Yano, M.; Shomura, A.; Sato, M.; Shimano, T. et al. (1998) A high-density rice genetic linkage map with 2275 markers using a single F₂ population. *Genetics* 148: 479-494.
- Helentjaris, T.; Slocum, M.; Wright, S.; Schaefer, A.; Nienhuis, J. (1986) Construction of genetic linkage map in maize and tomato using restriction fragment length polymorphisms. *Theor. Appl. Genet.* 72: 761-769.
- Lee, M.; Sharopova, D.; Beavis, W.D.; Grant, D.; Katt, M.; Blair, D.; Hallauer, A. (2002)

Expanding the genetic map of maize with the inbred B73 x Mo17 (IBM) population. *Plant Mol. Biol.* 48: 453-461.

SanMiguel, P.J.; Tikhonov, A.P.; Jin, Y.K.; Motchoulskaia, N.; Zakharov, D.; Melake-Berhan, A.; Springer, P.S.; Edwards, K.J.; Lee, M.; Avramova, Z.; Bennetzen, J.L. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, 274: 765-768.

Soderlund, C.; Humphray, S.; Dunham, A.; French, L. (2000) Contigs built with fingerprints, markers and FPC v4.7. *Genome Res.* 10: 1772-1787.

Tikhonov, A.P.; Dubcovsky, J.; SanMiguel, P.J.; Nakajima, Y.; Gorenstein, N.M.; Bennetzen, J.L.; Avramova, Z. (1999) Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* 96: 7409-7414.

Palavras-chave: genoma, milho, mapeamento, seqüenciamento