

A Importância da Pesquisa Genômica e o Sequenciamento de DNA

Márcio Elias Ferreira¹

Carlos Rodrigues Borges Neto²

O estudo sistemático de genomas completos iniciou-se com a proposta de utilizar a tecnologia de DNA para expandir o conceito básico de mapeamento genético proposto por A. H. Sturtevant no começo do século XX. Assim, através da construção de mapas genéticos completos dos cromossomos iniciou-se uma caminhada com passos cada vez mais largos na direção de genes individuais e, por fim, de todo o repertório gênico de uma espécie. O primeiro grande sucesso do uso desta estratégia veio em 1983 quando o gene que causa a doença de Huntington foi mapeado no cromossomo 4 humano. Foi a primeira vez que um gene de grande importância para a saúde humana foi localizado precisamente em um dos 23 pares de cromossomos. Pouco tempo depois observou-se uma revolução na genética médica quando os genes que causam mais de 1.000 doenças humanas foram mapeados em sítios cromossômicos específicos. O sequenciamento do primeiro genoma completo de bactéria (*Haemophilus influenza*) foi completado ainda em 1995. De lá para cá um grande número de genomas procariotos foi seqüenciado, e alguns genomas eucariotos decifrados (levedura, arroz, homem, drosófila, *arabidopsis*, *C. elegans*).

O sequenciamento de genomas completos permitiu observar, com grande surpresa, que o número de genes distintos necessários para o desenvolvimento de um organismo complexo como o ser humano (< 40.000 genes) não é muito maior do que o de um genoma de um eucarioto como a planta *arabidopsis* (~ 25.000 genes) e, possivelmente, inferior ao de outras plantas como o arroz (> 40.000 genes). A tecnologia empregada na análise genômica hoje disponível sofreu dramático progresso em pouco tempo. No final da década de 70, por exemplo, uma tese de doutorado completa seria devotada ao sequenciamento de um gene com alguns milhares de bases de DNA. No final dos anos 90, seqüenciadores automáticos de DNA são capazes de seqüenciar meio milhão de pares de base por dia. Mapas genéticos que levariam anos para ser construídos com marcadores moleculares nos anos 80, hoje podem ser construídos em poucos dias.

As informações até o momento apreendidas dos projetos genoma são inúmeras. Por exemplo, comparações de seqüências gênicas revelam que proteínas altamente similares são codificadas nos genomas de organismos tão distantes evolucionariamente quanto a levedura e os

¹ PhD. Genética Vegetal – Pesquisador III, Embrapa Recursos Genéticos e Biotecnologia.
E-mail : ferreira@cenargen.embrapa.br

² Dr. Produção Vegetal – Técnico de Nível Superior III, Embrapa Recursos Genéticos e Biotecnologia.
E-mail : Borges@cenargen.embrapa.br

humanos. Está claro também que a evolução da vida no Planeta foi muito conservativa, isto é, a partir do momento em que as células nucleadas surgiram há mais de 1,5 bilhão de anos atrás, a grande maioria das proteínas então desenvolvidas foram perpetuadas em uma infinidade de células descendentes, algumas vezes com apenas pequenas mudanças. O que temos observado é que a identificação de famílias gênicas produziu uma enorme sinergia na pesquisa, uma vez que a função de um membro de uma família protéica pode ser deduzida a partir daquela dos seus familiares.

A análise de seqüência revolucionou o estudo de várias áreas de pesquisa. É comum hoje a estimativa de árvores filogenéticas relacionando organismos com base na similaridade dos seus genes e não da sua morfologia. Algumas perspectivas se apresentam para os próximos anos. Por exemplo, acredita-se que a genética vegetal será afetada em profundidade em curto prazo. Uma vez que os genomas de uma mono e uma dicotiledônea estão finalizados (arroz e arábida), o próximo passo será a tarefa de compreensão do espectro de variação genética no complemento gênico da espécie e sua relação com características de interesse econômico. Uma espécie vegetal como *Arabidopsis* possui cerca de 25.000 genes. Já o arroz tem o número de genes estimado em acima de 40.000. É, portanto, possível catalogar os genes e a variabilidade gênica existente na espécie. Esta tarefa, de acordo com algumas iniciativas em fase embrionária, deverá ser em grande parte concluída através de um esforço conjunto nos próximos 10 anos. Afinal, uma espécie autógama mantém somente um limitado grau de diversidade gênica. Somente um pequeno número de variantes para cada gene existe no genoma. Há, portanto, um potencial de que será possível catalogar todos os variantes comuns (alelos) de todos os genes de uma espécie vegetal. Haverá, portanto, forte ênfase no desenvolvimento de coleções de mutantes de espécies e de técnica que permitam a atribuição de função gênica em espécies que tiveram o genoma seqüenciado. Estes variantes atraem enorme interesse por potencialmente estarem relacionados com características de interesse econômico. O desafio será identificar a coleção completa de alelos e testar a correlação com características de interesse.

Da mesma forma que as comparações dentro da espécie, as realizadas entre espécies são também reveladoras. As comparações evolucionárias entre organismos permitem identificar as seqüências que possuem importantes

papéis funcionais na estrutura protéica e regulação gênica e, portanto, têm sido mantidas inalteradas ao longo de períodos de tempo evolucionário. Assim, a análise comparativa de genomas tem despertado muito interesse nos últimos anos. Isto significa que a informação de genes que controlam determinada característica em arroz é relevante para o conhecimento de característica similar em trigo ou milho. Comparação de seqüências deve permitir a identificação de genes que são cruciais para a distinção e criação de novas espécies. Estes genes devem ter sido submetidos a forte seleção, culminando em uma evolução de seqüência mais rápida, separando as espécies.

A integração de métodos clássicos de melhoramento genético com as estratégias e tecnologias da genômica levará ao estabelecimento de novos paradigmas para o desenvolvimento de cultivares superiores de plantas. Ela se sustenta no chamado melhoramento molecular ("molecular breeding") que utiliza informações de mapas genéticos saturados com marcadores moleculares, mapas físicos construídos com base em bibliotecas de grandes insertos de DNA e análise genético-quantitativa, para identificar as regiões do genoma que contêm genes de interesse econômico. A clonagem de genes através de estratégias de avaliação de genes candidatos nas regiões identificadas, ou de seleção assistida por marcadores moleculares integrada a métodos de retrocruzamento ou seleção recorrente são métodos já empregados com sucesso em programas públicos e privados.

No momento, faz-se necessário o aprimoramento da tecnologia de identificação dos genes de uma célula que estão sendo expressos e dos que são silenciosos. Várias estratégias estão sendo desenvolvidas para estudar misturas complexas de proteínas expressas, a nova ciência da proteômica. Esta área deve receber grande atenção nos próximos anos e por certo culminará em uma expansão do conhecimento biológico numa escala ainda mais acelerada. Uma vez que os espectros de proteínas expressas dentro da célula determinam sua biologia, descrições amplas da constituição protéica e da escala de expressão temporal dos genes serão a base para a compreensão precisa da diferenciação celular.

A disponibilidade de catálogos completos de genes e, portanto, proteínas, dos organismos, vem redirecionando os biólogos para uma perspectiva global nos processos da vida: o estudo do papel de todos os genes e todas as proteínas ao mesmo tempo. A nova estratégia promete uma grande perspectiva. Ao mesmo tempo, ameaça inundar os cientistas com uma enorme quantidade de dados que coibirá a capacidade de interpretação. Será necessário grande investimento e desenvolvimento

em bioinformática nos próximos anos para assimilar e interpretar os dados provenientes de vários tipos de pesquisa genômica. O objetivo, a longo prazo, é utilizar esta informação para reconstruir o complexo circuito molecular que opera dentro da célula, ou seja, mapear a rede de interações protéicas que determina a lógica das várias funções celulares, incluindo resposta a estresses fisiológicos, etc. Serão necessárias estratégias gene-específicas para interromper a função de cada componente celular e estudar os efeitos de tais dissociações nos outros genes e proteínas celulares.

Dentro do contexto acima exposto, a Plataforma de Sequenciamento de DNA do Laboratório de Genoma Funcional da Embrapa Recursos Genéticos e Biotecnologia é uma unidade central de prestação de serviços de sequenciamento de DNA e procedimentos correlatos para todos os pesquisadores da Embrapa Recursos Genéticos e Biotecnologia, bem como de outras unidades descentralizadas da Embrapa. Além disso, os serviços de sequenciamento são oferecidos, também, a toda a comunidade científica de acordo com contratos específicos ou disponibilidade de tempo e recursos.

A Plataforma de Sequenciamento foi criada sobre a premissa de que, para a geração de dados de sequenciamento em escala utilizando métodos de alto desempenho, é economicamente mais viável e tecnicamente mais eficiente a existência de uma infraestrutura centralizada para o atendimento das necessidades de sequenciamento de DNA de vários projetos (Café, Eucalipto, *Crinipellis pernicioso*, Banana, Bovinos entre outros), que são fruto de parcerias entre vários centros de pesquisa da empresa e de instituições nacionais e internacionais. Nestes primeiros dezessete meses de funcionamento foram produzidas, aproximadamente, 110.000 seqüências válidas (Tabela 1) nos diversos projetos citados, mostrando o importante papel do laboratório no apoio à pesquisa genômica.

Tabela 1: Totais de seqüências válidas produzidas pela Plataforma de Sequenciamento de DNA e placas formato 96 seqüenciadas, em diversos projetos, no período de agosto de 2002 a dezembro de 2003.

Projeto	Total de Seqüências Válidas	Total de Placas
Arachis	423	8
Algodão	2.338	29
Banana bac	7.596	122
Banana est	2.052	30
Banana rga	498	3
Bovinos	13.488	203
Café	35.347	507
Capsicum	128	2
Cenoura	193	7
Crinipellis	17.628	270
Eucalyptus bac	801	11
Eucalyptus est	24.067	301
Guaraná	1.487	24
Laranja	1.343	20
Maçã	63	1
Mamão	127	3
Nematóides	2.385	30
Geral	109.964	1.571

**Comunicado
Técnico, 91**

Ministério da Agricultura,
Pecuária e Abastecimento

Exemplares desta edição podem ser adquiridos na:
Embrapa Recursos Genéticos e Biotecnologia
Serviço de Atendimento ao Cidadão
Parque Estação Biológica, Av. W/5 Norte (Final) -
Brasília, DF. CEP 70.770-900 - Caixa Postal 02372
PABX: (61) 448-4600 Fax: (61) 340-3624
<http://www.cenargen.embrapa.br>
e.mail:sac@cenargen.embrapa.br

1ª edição

1ª impressão (2003): 150 unidades

**Comitê de
publicações**

Presidente: *José Manuel Cabral de Sousa Dias*
Secretário-Executivo: *Maria José de Oliveira Duarte*
Membros: *Maurício Machaim Franco*

Regina Maria Dechechi G. Carneiro

Luciano Lourenço Nass

Sueli Correa Marques de Mello

Vera Tavares Campos Carneiro

Supervisor editorial: *Maria José de Oliveira Duarte*

Normalização Bibliográfica: *Maria Alice Bianchi*

Editoração eletrônica: *Giscard Matos de Queiroz*

Expediente