Genoma Cana: o Brasil na pesquisa de genoma de plantas

COPERSUCAR

Um *workshop* internacional nos dias 12 e 13 de abril dará partida ao Genoma-Cana, projeto que integra o programa Genoma da FAPESP e que poderá colocar o Brasil lado a lado com os países que hoje lideram a pesquisa em genoma de plantas. A área é considerada altamente estratégica tanto pelas mudanças radicais que deverá provocar na pesquisa básica de plantas quanto pelos esperados impactos dos resultados dessa pesquisa sobre a agricultura, meio ambiente, energia e saúde, entre outros setores. E sintoma muito claro do peso que lhe vem sendo atribuído foram os US$ 85 milhões investidos, em 1998, só pela National Science Foundation, dos Estados Unidos, em seu programa de pesquisa em genoma de plantas.

Na reunião de trabalho organizada pela FAPESP, os pesquisadores brasileiros envolvidos no Programa Genoma vão debater com colegas europeus, norte-americanos, australianos e da África do Sul a situação presente — ou, como preferem os cientistas, o atual estado da arte — da pesquisa em genoma de plantas e, particularmente, em genoma de cana-de-açúcar. E, é claro, nessa oportunidade o projeto Genoma-Cana será apresentado à comunidade científica internacional.

A meta do projeto, é seqüenciar cerca de 50 mil genes e identificar um certo número deles, de grande significado para a agroindústria da cana-de-açúcar. Muito mais importante que o seqüenciamento em si, o esforço paralelo de identificação vai se centrar prioritariamente nos genes responsáveis pelo metabolismo da sacarose, nos genes relacionados à resistência da planta a patógenos, e em outros envolvidos com a tolerância da cana a condições adversas de clima e solo.

Expectativa justificada

O projeto paulista surge já provocando uma certa expectativa entre os especialistas estrangeiros em pesquisa de cana porque, nesse segmento, será um dos primeiros a se valer das ESTs ¬ sigla referente a *Expressed Sequence Tags* ou, em português, etiquetas de seqüências expressas, tecnologia de seqüenciamento rápido baseada na identificação apenas das porções dos genes que codificam proteínas. Essa tecnologia facilita enormemente a descoberta de novos genes e, em cana-de-açúcar, entrar nesse terreno, inclusive determinando a estrutura e a função dos genes, equivale quase a explorar solo virgem. Porque, embora a pesquisa da planta seja bastante disseminada pelo mundo, ela está mais concentrada em mapeamento genético e, até aqui, apenas a África do Sul deu início a um projeto com o uso de ESTs.

O interesse pelo Genoma-Cana alcança também a agro-indústria do setor. Registre-se, a propósito, que foi a Copersucar, associação dos produtores paulistas de cana-de-açúcar, quem apresentou à FAPESP, em meados de 1998, a proposta de seqüenciamento da cana. A razão de tal interesse é muito clara: A descoberta de genes pode significar o caminho para uma evolução extraordinária nos programas de melhoramento genético da cana-de-açúcar que, até o momento, se valem de um número limitado de genes, e todos estranhos à planta. Aposta-se, com razão, que a utilização de genes próprios deve resultar, não só em maior resistência das variedades convencionais a doenças e pragas, mas especialmente em novas e melhores características agronômicas da planta, o que inclui teor de sacarose , produtividade, despalha natural, resistência a seca e geadas, e precocidade.

A importância do Genoma-Cana, no entanto, ultrapassa largamente os interesses setoriais. Entre outras razões porque a utilização da técnica de EST no projeto permite a comparação das seqüências obtidas da cana-de-açúcar a seqüências de outras plantas, já depositadas em bancos de dados, ge-

Parceria necessária com a agroindústria

A cana-de-açúcar tem especial significado econômico para o Brasil, que lidera a lista dos 80 países produtores, garantindo 25% da produção mundial. No país são cultivados aproximadamente 5 milhões de hectares, que produzem mais de 300 milhões de toneladas de cana, utilizadas como matéria prima na produção de 14,5 milhões de toneladas de açúcar e 15,3 bilhões de litros de álcool. O setor é responsável por cerca de 1 milhão de empregos diretos na área rural, e outros 300 mil diretos e indiretos na área industrial. Há 350 unidades industriais, todas privadas, e cerca de 50 mil produtores independentes de cana-de-açúcar.

Em São Paulo, responsável por 60% da produção nacional, o agribusiness da cana movimenta R$8 bilhões por ano e proporciona 600 mil empregos diretos. Este estado é, em nível mundial, líder em competitividade (menor custo de produção) e em exportações de açúcar. Não bastasse isso, a importância econômica da cana para São Paulo deve crescer muito em função de seu potencial para a produção de energia renovável. O volume de exportação de açúcar cresceu sensivelmente na década de 90, até os níveis atuais de cerca de 6 milhões de toneladas (previsão para a safra 1997/98). As receitas em divisas estão variando entre US$ 1,5 bilhão a US$ 1,8 bilhão por ano, o que representa cerca de 3,5% do total das exportações brasileiras.

rando uma série de informações sobre genes homólogos em espécies diferentes e facilitando a descoberta de novos genes, segundo o pesquisador Paulo Arruda, que atuará como coordenador de DNA do Projeto Genoma-Cana.

Por conta de tais analogias, o trabalho com a cana-de-açúcar pode levar a um volume extraordinário de novas informações sobre genética vegetal, úteis para todos os outros projetos genona de planta em desenvolvimento pelo mundo. "O volume de informações que o Genoma-Cana tem potencial para gerar poderá colocar o Brasil como referência mundial em EST de plantas", diz, com entusiasmo, Arruda, que é professor e coordenador do Centro de Biologia Molecular e Engenharia Genética da Unicamp.

É por isso mesmo que, para o diretor científico da FAPESP, professor José Fernando Perez, "a importância do projeto Genoma-Cana, dentro da pesquisa em genoma de plantas, não pode ser subestimada". A propósito, ele observa que o Conselho Nacional de Ciência e Tecnologia dos Estados Unidos, criado em fins de 1993 pelo presidente Bill Clinton, lançou, em janeiro deste ano, um documento alentado sobre a estratégia que o país deve seguir para dar largos passos na pesquisa de genoma de plantas, batizada de *National Plant Genome Initiative*. Ao justificá-la, o documento observa que os maiores desafios a serem enfrentados pela humanidade, no século 21, são a ampliação da produção de alimentos e fibras, a obtenção de um meio ambiente mais limpo e saudável e novas fontes renováveis de energia e produtos químicos. E ¬ ressalta ¬ são justamente as tecnologias ligadas à pesquisa vegetal que podem desempenhar o mais importante papel para vencer cada um desses desafios. "É essa compreensão do valor estratégico da pesquisa de genoma de plantas que temos que ter. E no caso do projeto da cana, precisamos perceber que ele nos insere de uma maneira muito feliz no rumo internacional dessa área de pesquisa", observa o professor Perez.

A publicação do edital para seleção dos laboratórios que vão realizar o Genoma-Cana estava prevista para o começo do mês de março. Devem participar do projeto entre 20 e 25 laboratórios, que terão a incumbência, desde o início dos trabalhos, não só de fazer o seqüenciamento, mas também sua anotação, e identificação de genes.

Para se chegar ao seqüenciamento previsto de 50 mil genes, cerca de 400 mil seqüências devem ser trabalhadas num período estimado de dois anos. Isso não chega a assustar Paulo Arruda. Até porque, antes mesmo do Genoma-Cana ser oficialmente iniciado, já tinham sido feitas as primeiras bibliotecas de cDNA e estavam clonados 500 genes de tecidos variados da cana-de-açúcar. "A depender da velocidade com que se consiga suprir as bibliotecas, podemos gerar cerca de 2 mil seqüências por dia", diz ele, otimista.

Nos mesmos moldes do primeiro projeto genoma da FAPESP, o da cana terá em sua coordenação central, além do coordenador de DNA, um laboratório de Bioinformática. Da mesma forma, um comitê internacional vai supervisionar o projeto e, na parte de tecnologia de ESTs, haverá estreita cooperação com o grupo do Genoma-Câncer.

Identificando proteínas

Na última semana de janeiro, 24 pesquisadores de instituições paulistas e de outros países da América Latina puderam de atualizar seus conhecimentos sobre identificação de proteínas e seqüências de DNA através da busca de homologias em bancos de dados. Isso aconteceu graças ao Computational Genomics Course, organizado pela FAPESP, Academia Nacional de Ciências dos EUA e Academia de Ciências do México e realizado na Faculdade de Medicina da USP (FM/USP).

Os participantes do curso se capacitaram a utilizar os *softwares* — inclusive com a orientação de um dos pioneiros na área, o bioquímico William Pearson, co-autor do pacote de programas Fasta – e receberam ensinamentos conceituais para analisar os resultados obtidos. Isso permitiu que na metade do **workshop** já pudessem realizar pesquisas em bancos de dados de proteínas existentes em vários países, sobre questões práticas enfrentadas nos laboratórios.

Essa capacitação é fundamental para várias áreas ligadas à pesquisa com proteínas e genomas e é importante nesse momento que o Projeto Genoma *Xylella* está em fase final e que surge o Projeto Genoma Funcional. "Agora temos os genes da *Xylella* e codificaremos as proteínas. Para se projetar um remédio contra o "amarelinho", porém, temos que saber qual é a função das proteínas que a *Xylella* produz e encontrar uma maneira para que ela não sobreviva na laranjeira ou não se reproduza", esclarece Luís Fernandez Lopez, organizador do curso junto com Eduardo Massad, ambos da FM/USP. "Teremos que conhecer a estrutura e a função de cada proteína, saber quais atacar; como atacá-las." Esse trabalho segue-se à identificação das proteínas, com a utilização dos programas de busca de homologias com proteínas já conhecidas.

Essa busca de homologias é possibilitada pelo estudo evolutivo das proteínas. Segundo Lopez, utiliza-se uma matriz da evolução em um milhão de anos e multiplica-se a matriz por ela mesma o número de vezes necessário para se recuar no tempo os milhões de anos presumidos para a ocorrência da mutação.

Além de Pearson, o curso teve como instrutores Randy Smith (SmithKline Beecham), Lincon Stein (Cold Spring Harbor Laboratory), Gary Stormo (Universidade do Colorado), Jacques Retief (Universidade da Virgínia), Peter Cooper (National Institutes of Health) e Warren Gish (Universidade de St. Louis).

Trabalho pioneiro

Um dos autores do Fasta — o mais conhecido conjunto de programas de busca de homologias entre proteínas e seqüências genéticas em bancos de dados—, o bioquímico William Pearson, da Universidade da Virgínia, foi um dos instrutores do Computational Genomics Course. Com quase 60 artigos publicados em revistas de prestígio internacional, é uma dos pesquisadores mais citados na área (cerca de 9.500 menções no *Citation Index*).

Pearson está trabalhando no momento com um recuo evolutivo de 4 bilhões de anos. Isso lhe permite identificar organismos simples cujas proteínas são relacionadas com aquelas do ser humano. Um exemplo: "A base genética de uma forma de câncer foi descoberta quando determinou-se que a região cromossômica humana envolvida com a doença continha um gene relacionado com um gene reparador da *E. coli* e do levedo. O gene da *E. coli* divergiu do gene humano há mais de 2,5 bilhões de anos. Como as proteínas mudaram vagarosamente, foi possível determinar que são relacionadas."

Pearson acredita que mais de 75% das proteínas humanas serão vinculadas a uma família protéica quando o Projeto Genoma Humano terminar. "Mas isso não quer dizer que saberemos exatamente o que elas fazem, ou, mais importante, provavelmente não saberemos como elas interagem com outras proteínas. Se podemos entender uma das funções de muitas proteínas, provavelmente não saberemos muito das funções da maioria delas."