

## Genoma

# Projeto EST-Café permitirá acelerar lançamentos de novas variedades

Paulo Mazzafera\*

ANNANETTO



*Caféiros cultivar Catuaí Vermelho: Sul de Minas Gerais, 2012*

O genoma é o conjunto das informações genéticas, na forma de genes, necessárias à formação de um organismo. Projetos genomas, portanto, objetivam conhecer todas as informações genéticas de uma determinada espécie, ou seja, “ler” os genes que ela possui. O projeto conhecido como Genoma Café (Vieira et al., 2006) não é, na verdade, o sequenciamento dos genes do café, mas sim de uma parte das informações contidas nesses genes. Genes são formados por moléculas de DNA (ácido desoxirribonucleico), que possuem em suas estruturas quatro unidades de bases (adenina e guanina = bases purínicas; citosina e timina = bases pirimidínicas), além de fosfato e um açúcar, conhecido como ribose.

Algumas regiões dos genes podem ser “lidas” por enzimas específicas (RNA polimerases), as quais produzem o RNA mensageiro (mRNA), etapa que ocorre no núcleo das células e é conhecida como transcrição. Nas moléculas de RNA, a base timina é substituída pela base uracila. No citoplasma o mRNA se liga a estruturas conhecidas como ribossomos, que mediam a formação das proteínas. O conjunto de três bases identifica um aminoácido; ao ser lido nos ribossomos, o mRNA forma uma cadeia de aminoácidos, etapa conhecida como tradução.

As proteínas, por sua vez, permitem a formação das muitas substâncias presentes nos organismos vivos, além de fazer parte de estruturas. Por exemplo,

nas plantas, uma enzima apelidada Rubisco (abreviação de RibuloseBisfosfato-Carboxilase/Oxigenase) é responsável na fotossíntese, por catalisar a incorporação do carbono do CO<sub>2</sub> em uma molécula de açúcar de 5 carbonos – a ribulose bisfosfato. Outras proteínas que não têm atividades enzimáticas como a Rubisco (ou seja, não participam de reações químicas), podem fazer parte de estruturas (proteínas na membrana celular) ou auxiliar os processos celulares (por exemplo, aquaporinas, que facilitam a entrada de água nas células). Ou seja, as informações contidas no DNA (gene) têm que ser copiadas (transcritas) em uma molécula de mRNA, que por sua vez será traduzida em uma proteína.

Muitas proteínas são produzidas constantemente, outras não... De outra forma, diversos genes são constantemente expressos e outros passam a ser expressos ou, ainda, mais ou menos expressos, em resposta a algum tipo de estímulo. Por exemplo, uma planta pode passar a expressar um gene e produzir a proteína a ele relacionada mediante o estímulo causado pelo ataque de um inseto. Na mesma situação, outro gene pode deixar de ser expresso. Portanto, se aplicarmos determinado estímulo a uma planta e conseguirmos separar as moléculas de mRNA, num dado momento, o que faremos é coletar informações extraídas dos genes em resposta àquele estímulo, além daqueles genes que são constantemente expressos. É aqui que entra o sequenciamento por EST, usado no chamado projeto Genoma Café.

EST vem da expressão em inglês *Expressed Sequence Tag* (Etiqueta de Sequência Expressa), referindo-se a genes que estão funcionando, em determinado momento, na forma de mRNA. O que se faz, inicialmente, é extrair e purificar as moléculas de mRNA e, com a ajuda de uma enzima chamada DNA polimerase, realiza-se uma cópia do mRNA na forma de um cDNA – cuja letra “c” significa complementar, indicando tratar-se de uma cópia do RNA. Em seguida, cada cDNA tem a sequência de suas bases determinadas por aparelhos chamados sequenciadores. Por limitações técnicas, a sequência de cada cDNA é, na maior parte das vezes, parcial e, devido à maneira como é sequenciada, muitas das sequências podem ser obtidas para um mesmo cDNA.

Assim, para se obter a sequência inteira ou a mais completa possível, lança-se mão da bioinformática: programas de computadores (*softwares*) específicos fazem comparações das sequências e verificam suas similaridades (a sequência de bases); ou seja, organizando as sequências de tal forma que se sobreponham nas partes semelhantes. Com isso, essas sequências parciais resultarão em uma sequência completa e precisa cha-

mada de “consenso deduzida” ou, ainda, “contig”. Falta, então, identificar o que é cada contig.

Novamente, com o auxílio da bioinformática, os contigs são comparados a milhões de sequências de genes já encontradas e identificadas em outras plantas, depositadas em bancos públicos disponíveis pela Internet e, a depender do grau de semelhança, pode-se dizer que tal sequência identifica o gene de nome X, que tem a função Y. Esses bancos públicos são muito completos e trazem muitas informações mais, como, por exemplo, qual a função metabólica, quem depositou a sequência no banco, se existe algum trabalho publicado a ela relacionado, como é a sequência de aminoácidos traduzida, se existe alguma característica na sequência que permita relacioná-la a um grupo de proteínas, entre muitas coisas mais.

Resumindo, DNA é “transcrito” em mRNA, que é extraído, purificado e copiado em cDNA. Este último, por sua vez, é sequenciado e suas sequências são usadas na montagem de contigs, então comparadas com sequências de bancos de dados para se determinar a identidade e a função do gene. Mas qual gene? Aquele encontrado no DNA, no início do processo. A palavra transcrito, entre aspas porque, ao determinar as sequências dos mRNAs, o que fazemos é chamado de “transcritômica”, ou seja, determinamos quais genes são expressos; em outras palavras, transcritos. Aqui surge a grande utilidade de um programa de sequenciamento de ESTs, que é retratar determinado momento e determinado estímulo ou situação, em termos da expressão dos genes.

Exemplificando, imagine um cafeeiro sozinho crescendo no meio de um campo, onde não falta água. Ali uma série de genes são expressos, produzem suas respectivas proteínas e atuarão em muitos processos, tais como o de biossíntese de centenas de substâncias, a fotossíntese, a respiração, a expansão celular, a absorção de nutrientes etc. Além disso,

durante um ciclo, esta planta passará por diferentes fases, como vegetação, florescimento e frutificação. Nessas condições, já se pode imaginar que para cada processo metabólico e para cada fase do desenvolvimento, vários e diferentes genes se expressam ou, ainda, deixam de ser expressos. Sem esquecer que, ao longo do ano, existe variação de temperatura, insolação e umidade.

Em seguida, vamos colocar um cafeeiro de cada lado daquele inicial. Com isso, haverá um certo sombreamento entre os cafeeiros, que poderão começar a competir por água e nutrientes no solo. Imagine agora o plantio de milhares de cafeeiros, criando a possibilidade de ocorrer, por exemplo, o desenvolvimento da ferrugem por conta do aumento da umidade nas folhas. E, numa época mais seca, haverá não só competição por água, como o desenvolvimento de bicho-mineiro. Assim, ao crescer em uma cultura, nosso solitário café não só terá que desenvolver seu próprio ciclo, como estará sujeito a fatores ambientais naturais, ao longo do ano, e a uma série de fatores decorrentes do manejo da cultura. Haverá maior complexidade nas interações e, conseqüentemente, alterações na expressão de muitos genes.

Portanto, ao analisar as sequências de mRNAs produzidos em determinado momento (o transcritôma), o que se faz é, na verdade, analisar as respostas a estímulos impostos pelo ambiente (estímulos abióticos) e outros organismos (estímulos bióticos) nas atividades metabólicas dos diferentes tecidos do cafeeiro, em determinado momento. E que conseqüências práticas um projeto de sequenciamento EST pode ter? Um exemplo simples é a reação do cafeeiro à ferrugem. Conhecendo-se bem o ciclo da doença, após a inoculação com os esporos, pode ser feita a coleta de folhas, ao longo do desenvolvimento da doença, e, a cada etapa, serão obtidas sequências contigs, como descrito acima. Em outras palavras, será obtido o transcritôma das folhas em resposta à ferrugem,

permitindo conhecer quais genes são expressos a cada etapa amostrada.

Como se conhece bem o ciclo da doença pode-se inferir que, na etapa em que a hifa do fungo penetra a folha, tais e tais genes passam a ser mais expressos; e, também, que tais e tais genes são menos expressos. Uma vez obtidas as informações sobre os genes e, eventualmente, sobre suas funções em outras plantas, isso nos permitirá compreender como o cafeeiro responde à ferrugem, naquela etapa específica do ciclo da doença. Se, porém, forem cruzadas as informações de um transcrito de uma cafeeiro tolerante com as de um suscetível, podemos encontrar genes diferencialmente expressos nas duas plantas e que podem estar, eventualmente, envolvidos na tolerância. Neste caso, o número de genes diferencialmente expressos pode ser razoavelmente grande e, novamente, a bioinformática vem em socorro. Com o uso deste programa de computador, podemos construir redes de correlações entre os genes e, uma vez conhecidas suas funções, eles podem ser posicionados em um mapa metabólico, do qual podemos extrair ainda mais informações.

Outros exemplos de emprego do projeto de sequenciamento EST seriam entender como cafeeiros se comportam em situações de estresse hídrico ou quando cultivados em locais mais altos e propícios ao desenvolvimento de bebida de melhor qualidade, e assim por diante... Apesar da simplicidade como o processo está descrito aqui, é preciso lembrar que ele não é tão fácil assim. Poucos são os processos metabólicos em plantas que dependem de poucos genes, e existem sempre relações entre diferentes vias metabólicas, de tal forma que não se “pinça um gene mágico” que resolve determinada doença. Outro agravante é que nem todos os contigs obtidos no sequenciamento de ESTs podem ser identificados, quando se faz a comparação com os bancos de dados. Muitos terão identidade e função desconhecida, uma

vez que – embora os bancos de genes sejam grandes e ricos – há ainda muita informação sobre as funções dos genes a serem estabelecidas.

Porém, caso sua expressão esteja fortemente associada a determinado evento (como no exemplo da infecção de ferrugem), pode-se tentar descobrir sua função, ao se isolar o gene e inseri-lo em plantas-modelo, como *Arabidopsis thaliana*; ou seja, faz-se uma planta transgênica capaz de expressar o gene desejado. Isto é possível por meio de técnicas de biologia molecular. A planta transformada que contém o gene de café poderá, então, ser estudada ao se inocular nela outra doença e verificar a sua reação. *A. thaliana* é uma planta pequena, com seu genoma inteiramente conhecido (agora, sim, todo o seu DNA sequenciado), de ciclo curto e que facilmente aceita ser transformada com um gene de outro organismo.

Outra grande utilidade do sequenciamento EST feito no café é a possibilidade de se encontrar marcadores moleculares. Algumas alterações (mutações) no genoma de uma planta passam de geração a geração e podem servir para identificar populações que carregam determinados tipos de características; como no nosso exemplo, resistência à ferrugem. Não necessariamente, um determinado marcador está em um gene relacionado à resistência, mas estaria próximo o suficiente para permitir que se estabeleça relação entre sua presença e a característica procurada/estudada. Vários marcadores podem ser associados a uma única característica e isso é bastante desejável. Marcadores moleculares são muito práticos em estudos de melhoramento genético, usados para selecionar plantas em estágios iniciais de crescimento, o que é muito importante no cafeeiro que tem um ciclo de vida muito longo; às vezes, são necessários vários anos de espera até que uma determinada característica se manifeste, como, por exemplo, a resistência à ferrugem.

O melhoramento genético convencional utiliza apenas a observação das características da planta para se obter uma nova variedade, o que demanda muito tempo. Por exemplo, a variedade Icatu demorou mais de 30 anos para ser lançada. Assim, analisar plantas em estágios precoces do desenvolvimento é uma vantagem grande que acelera o lançamento de novas variedades. Hoje, o sequenciamento por EST deixou, contudo, de ser uma técnica aplicada. Máquinas e métodos bem mais rápidos, eficientes e baratos permitem sequenciar organismos mais facilmente, como é o caso da técnica conhecida como RNAseq. De qualquer forma, o RNAseq também é um transcrito e permite obter as mesmas informações de um sequenciamento EST. O importante é que, com tais métodos, o avanço na pesquisa básica é enorme. Tanto que existe um considerável descompasso entre a quantidade de informações geradas em sequenciamentos e seu processamento por cientistas. Nisso a bioinformática é essencial, pois permite a comparação entre milhões de sequências de organismos diferentes ou não, e possibilita, ainda, integrar os dados com trabalhos de proteômica, metabolômica e outras ômicas (sim, existem várias ômicas além da genômica). Isto poderá, sem dúvida, apontar caminhos e economizar tempo; mas sempre, após descobrir um gene ou sua função em uma determinada situação, ele deverá ser testado em campo. Ou seja, a ciência sempre funde passado e presente. ☺

---

\*Paulo Mazzafera é professor no Departamento de Biologia Vegetal do Instituto de Biologia da Unicamp (pmazza@unicamp.br).

## REFERÊNCIAS BIBLIOGRÁFICAS

- VIEIRA, Luiz Gonzaga Esteves et al. Brazilian coffee genome project: an EST-based genomic resource. In: *Brazilian Journal of Plant Physiology*, 18, p. 95-108, 2006.
- MONDEGO, J. M. et al. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. In: *BMC Plant Biology* 11, 30, 2011.