

BIOINFORMÁTICA

A dança dos genes

Brasileiros criam modelo matemático que define limite de alterações e evolução de genomas

A bioinformática, ramo das ciências da computação dedicado a criar *softwares* e ferramentas matemáticas para a área biológica, acaba de demonstrar um novo teorema que pode ser útil para o estudo da evolução de genomas. Os pesquisadores João Meidanis e Zaroni Dias, da Universidade Estadual de Campinas (Unicamp), e Maria Walter, da Universidade de Brasília (UnB), calcularam o número máximo de vezes que podem ocorrer dois tipos básicos de rearranjos no interior de um genoma: movimentos de blocos de seus genes e inversões na seqüência dos pares de bases (unidades químicas) que compõem esses genes.

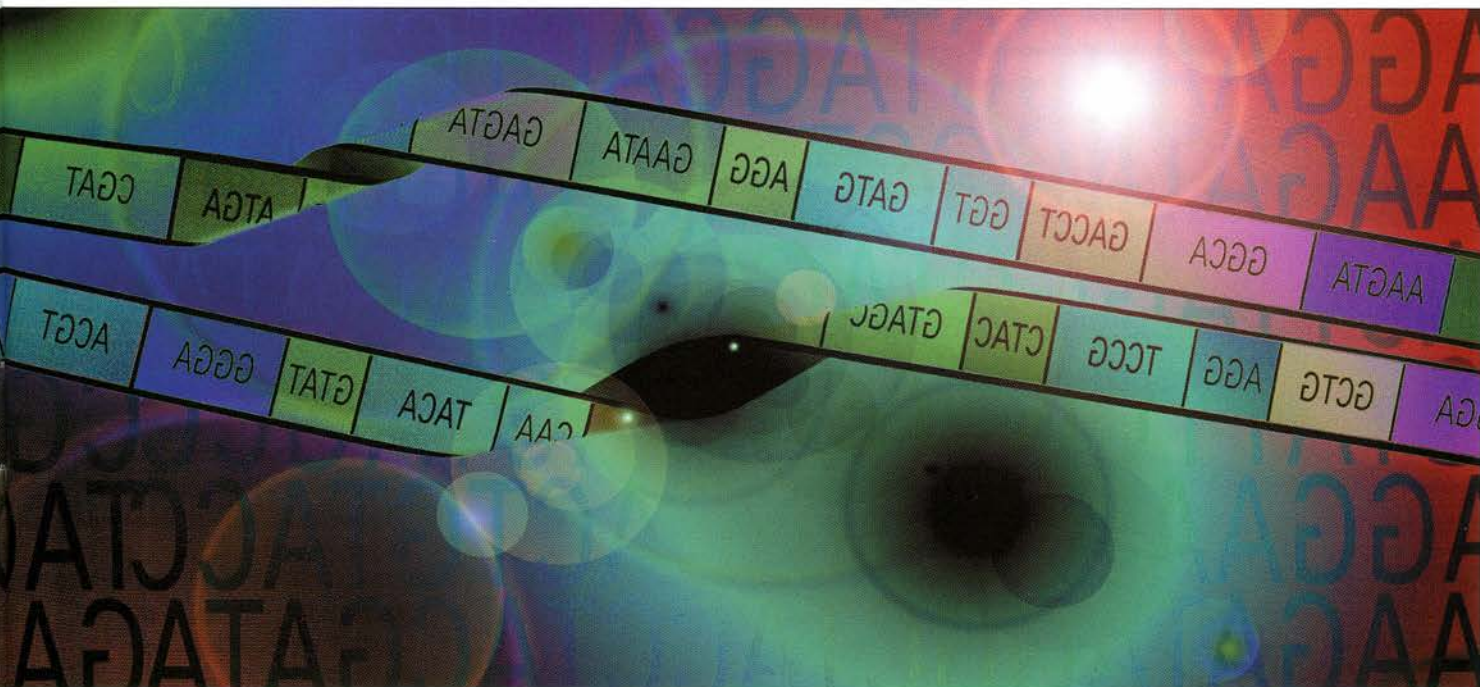
Segundo os autores do estudo, publicado no final do ano passado no *Journal of Computational Biology*, a respos-

ta a essa indagação é igual à metade do número de genes do genoma em questão mais o algarismo 2. Ou seja, num genoma com 100 genes, podem acontecer, no máximo, 52 rearranjos dos tipos descritos acima. Nesse caso, a conta é $100 \div 2 + 2 = 52$. Se o número de genes for ímpar, o resultado da equação, uma fração, deve ser arredondado para baixo. “Essa equação, a principal de nosso trabalho, é válida para genomas com três ou mais genes”, diz Meidanis, líder do grupo de bioinformatas à frente do estudo. “Ela é um aperfeiçoamento de teoremas propostos por outros autores.”

O teorema pode ser útil no estudo da evolução dos genomas por permitir compará-los e, desse modo, ver o que há de comum e de diferente entre eles. Ao sofrer rearranjos como os mencionados, um genoma se transforma e gera um

outro, diferente do original. Do ponto de vista evolutivo, a distância entre dois genomas pode ser encarada como diretamente proporcional ao número de arranjos processados: quanto maior a quantidade de operações de reordenamento, maior a distância evolutiva entre eles. Há outras formas de medir a proximidade evolutiva entre genomas, mas essa foi a usada como parâmetro pelos pesquisadores nesse trabalho.

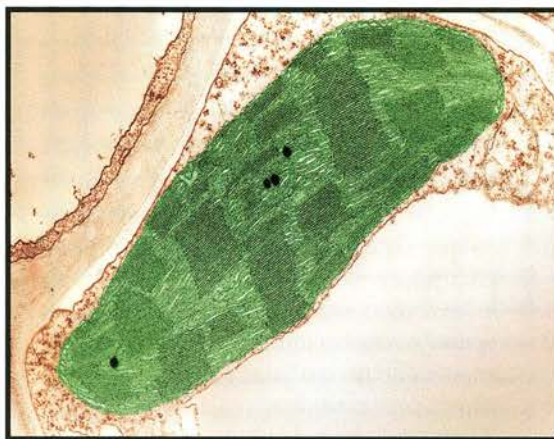
Portanto, o que a equipe de Meidanis fez foi calcular o número máximo de rearranjos – o equivalente à maior distância possível em termos evolutivos – que pode separar dois genomas que guardam certa semelhança. Um genoma com dez rearranjos em relação ao seu genoma de base, do qual derivou, está mais próximo de sua seqüência-mãe do que um terceiro genoma, que



apresenta 15 reordenamentos. “Nossa contribuição foi mostrar que o número máximo de rearranjos possíveis entre dois genomas similares é menor do que se pensava”, explica Meidanis. As equações defendidas por outros autores para tratar dessas questões chegam, invariavelmente, a resultados numéricos superiores aos obtidos pelo teorema demonstrado pelos cientistas da computação da Unicamp e UnB.

Em termos práticos, a equação proposta pela equipe de Meidanis pode ser aplicada apenas na comparação de genomas com características bastante específicas. Primeiro condicionante: ela se presta somente ao confronto de pares de genomas, um conjunto de genes contra outro conjunto de genes. Se houver dez genomas para serem comparados, a análise terá de ser feita dois a dois. Outra restrição: só faz sentido empregar a equação para comparar dois genomas muito parecidos – ambos devem ter os mesmos genes (pelo menos três para o teorema valer) e na mesma quantidade. Normalmente, os genomas comparados em laboratório não exibem essas condições ideais para o emprego da equação, mas isso não é motivo de preocupação. “O teorema é parte de um modelo teórico mais amplo que ainda precisa ser refi-

nado”, afirma Meidanis. Então, não dá para testar efetivamente se a equação é válida? “Dá, sim. Vamos usá-la em breve para comparar genomas de vírus, que são bastante semelhantes entre si.” Outra possibilidade é empregar a equa-



COLORIZAÇÃO SOBRE FOTO DE E. KITAJIMA/USP

Cloroplasto: evolução genética mensurável

ção para a análise da evolução de genomas de cloroplastos (estrutura celular responsável pela fotossíntese nas plantas), mitocôndrias (organela responsável pela produção de energia) e talvez de algumas bactérias.

Os rearranjos dos quais se ocupa o teorema recebem o nome técnico de transposição e reversão. Mais um exemplo ajuda a entender em que consistem essas duas operações. Imagine dois ge-

nomas, chamados de X e Y, ambos com a mesma quantidade de genes, cinco. Cada um desses genes – distintos entre si e denominados pelos números 1, 2, 3, 4 e 5 – aparece apenas uma única vez nos genomas. Em X, o genoma de referência, a seqüência dos genes é 1, 2, 3, 4 e 5. Em Y, por causa de um rearranjo interno, a seqüência padrão se alterou para 1, 4, 5, 2 e 3. No segundo genoma, o bloco 4,5 se posicionou entre o gene 1 e o gene 2. Tecnicamente, essa operação de rearranjo na ordem dos genes de um genoma é a transposição.

Já a reversão é uma alteração na ordem dos pares de base que compõem um gene (adenina, timina, citosina e guanina ou, simplesmente, A, T, C e G). Recorrendo ao exemplo acima, considere que a seqüência de pares de base do gene 1 em X, o genoma de referência, é ATCG. Após sofrer uma reversão, um processo complicado para não-iniciados em genômica, a seqüência resultante em Y será CGAT. “Existem outros tipos de rearranjos entre genomas, mas não os levamos em conta na criação do teorema, que se aplica a situações mais simples”, comenta Meidanis, fundador da Scylla, empresa de bioinformática de Campinas, que desenvolve *softwares* para a área de genômica. •