



SIMPLICIDADE

NATURAL

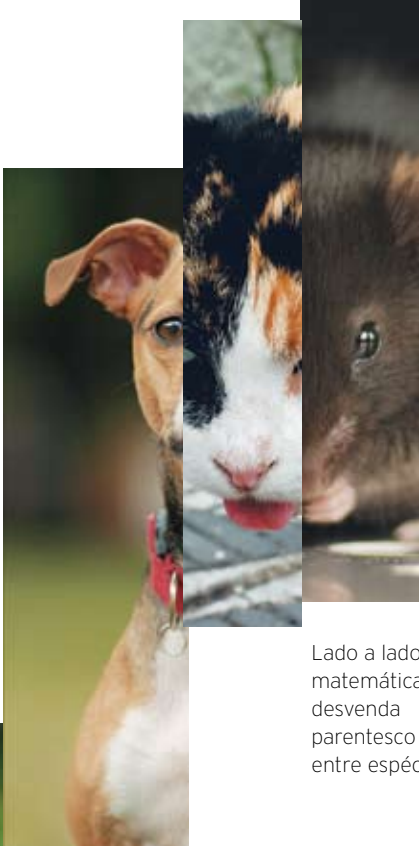
Novo método para comparar genomas pode ajudar a construir a árvore da vida

Cada vez é mais rápido e fácil ler o DNA completo de qualquer ser vivo, com os equipamentos modernos. Mas mesmo essa evolução da tecnologia ainda não tornou banal comparar genomas para avaliar semelhanças entre espécies e montar uma árvore da vida. “Com os métodos atuais, os computadores demoram muito para comparar o material genético integral de um conjunto com mais de vinte espécies”, avalia o matemático João Meidanis, do Instituto de Informática da Universidade Estadual de Campinas (Unicamp). Não satisfeito com a solução habitual de buscar aproximações para dar sentido aos dados, ele e seu aluno Pedro Feijão desenvolveram um método novo para comparar genomas, que em setembro passará por uma prova de fogo: será apresentado a colegas do mundo todo no *workshop* Algoritmos em Bioinformática, nos Estados Unidos.

O que torna tão lenta a análise dos dados não é mais obter as sequências, mas compará-las. Isso porque cada genoma é representado por bilhões de letras enfileiradas (cerca de 3 bilhões, no caso humano). Os métodos de comparação entre espécies usam representações matemáticas dos modelos de como as mutações naturais aos poucos substituem letras ou quebram essa longa cadeia que volta a emendar-se em outro ponto – uma contabilidade extenuante até para os computadores mais possantes.

A fórmula proposta pela dupla de matemáticos simula uma situação em que o genoma seria quebrado num só ponto e depois emendado outra vez de maneira aleatória. Se isso acontece sucessivas vezes, a sequência genética é aos poucos embaralhada. Daí vem o nome *single-cut-or-join* (único corte ou ligação) que batiza o método. O processo simula o tipo mais comum de rearranjo genético, em que um trecho do DNA fica invertido. Se a vírgula da frase anterior fosse o ponto de ruptura, a frase poderia virar “ociténeg ojnarraer ed mumoc siam opit o alumis ossecorp O, em que um trecho do DNA fica invertido” ou “em que um trecho do DNA fica invertido, O processo simula o tipo mais comum de rearranjo genético”, entre outras possibilidades. “Essa é uma das formas de alteração mais comuns no genoma”, explica Meidanis, “pois ela preserva trechos intactos e assim mantém propriedades genéticas”. O programa que ele desenvolveu





Lado a lado:
matemática
desvenda
parentesco
entre espécies



faz uma série de cortes aleatórios no genoma selecionado e determina a semelhança com outro genoma pelo número de cortes necessários para que o primeiro fique igual ao segundo. Ao comparar o material genético de várias espécies – com esse método é possível comparar até 100 genomas em poucos dias – o programa de Meidanis e Feijão produz uma árvore filogenética que mostra o parentesco entre os seres vivos comparados.

Debate - O trabalho teve uma recepção longe de unânime na comissão científica que analisou os trabalhos submetidos à conferência. “Dois revisores acharam que não estávamos apresentando nada útil e três ficaram em cima do muro”, conta Meidanis. Em vez de motivo para desânimo, a resposta foi um estímulo. A começar pelo artigo ter sido analisado por cinco revisores em vez dos habituais três. “Pelo visto eles tiveram dificuldades em decidir, mas mesmo assim possivelmente aceitaram o trabalho porque é algo novo que pode dar origem a um debate importante”, postula.

Eles estão preparados para a discussão. Já refizeram todos os cálculos para demonstrar que sua proposta é sim matematicamente distinta dos métodos em uso: o ponto de quebra

(*breakpoint*), usado desde o início do século XX quando surgiu a genética de populações, e o *double-cut-or-join* (duplo corte ou ligação), mais usado nos últimos tempos. O método mais antigo é conceitualmente muito parecido com o agora proposto, mas difere na formalização matemática; o mais recente considera que o genoma é quebrado em três partes que voltam a se juntar aleatoriamente – uma complexidade desnecessária, aos olhos de Meidanis. Para ele, a simplicidade de seu modelo torna mais fácil a resolução dos problemas. E talvez torne sua solução mais próxima da realidade, completa, citando o físico Albert Einstein: “Ele disse que tudo deveria ser considerado da maneira mais simples possível, mas não mais simples do que isso”.

Num próximo passo, discussões e colaborações com geneticistas serão essenciais para avaliar se a simplificação matemática excede a da natureza. Por enquanto, Meidanis e Feijão têm testado o modelo com conjuntos de dados que a comunidade de bioinformatas usa justamente para testar métodos novos. Ao comparar o formato das árvores obtidas pelo programa, eles constatam que seu método chega a resultados semelhantes ao que outros encontraram – mas com um tempo de processamento muito mais curto.

Mesmo antes da discussão deste mês e da publicação formal do trabalho, um grupo de pesquisa alemão já mostrou interesse em receber a versão final. Mais um sinal, para o professor da Unicamp, de que sua proposta é inovadora. ■

MARIA GUIMARÃES

► Artigo científico

FEIJÃO, P. e MEIDANIS, J. SCJ: a variant of breakpoint distance for which sorting, genome median and genome halving problems are easy. *9th Workshop on Algorithms in Bioinformatics*. 2009.