

NOÇÕES DE FILOGENÉTICA MOLECULAR

**Flora Maria de Campos Fernandes-Matioli**

Depto. de Biologia, Instituto de Biociências, USP

São Paulo - SP

E-mail: fmcampos@ib.usp.br

Filogenética é o campo da Biologia que busca identificar e compreender as relações evolutivas entre as diferentes formas de vida na Terra. Os primeiros critérios objetivos para a reconstrução filogenética baseavam-se em dados morfológicos. Com o acesso recente à estrutura de macromoléculas (DNA, RNA e proteínas) as análises filogenéticas passaram a ter um avanço vertiginoso.

Para se trabalhar com reconstrução filogenética baseada em dados moleculares é preciso conhecer algumas propriedades das seqüências a serem comparadas. Primeiramente, elas têm de ser homólogas, ou seja, apresentar ancestralidade em comum. Há três tipos de homologia: ortologia, quando as seqüências têm um único e mesmo ancestral comum; paralogia, quando se originam de uma duplicação gênica; e xenologia, quando se originam por incursão lateral (ou horizontal). Apenas seqüências ortólogas poderão fornecer informações filogenéticas na história de organismos.

Uma vez obtidas as seqüências, estas devem ser alinhadas corretamente. Para tal, apesar de haver muitos programas computacionais que realizam essa tarefa, as seqüências devem ser alinhadas manualmente. Entretanto, alinhamentos automáticos podem ser um primeiro passo. Essa etapa é de extrema importância, pois um alinhamento errado comprometerá toda a análise.

Para um alinhamento coerente, deve-se conhecer os diferentes tipos de mutações e substituições de nucleotídeos. Mutações ocorrem ao acaso em qualquer parte do genoma e podem ser transições, tranversões, deleções/inserções e inversões. Substituições referem-se a regiões codificantes e podem ser sinônimas (quando não há alteração do aminoácido codificado), não sinônimas (quando há alteração) e sem sentido (quando um códon de parada é gerado).

Em todas as metodologias de filogenética molecular, cada posição ocupada na seqüência (nucleotídeo ou aminoácido) é considerada como um caráter do tipo multi-estado (podendo ser um dos quatro nucleotídeos ou um dos vinte aminoácidos) e cada caráter é considerado independente dos demais. A variação dos estados de caracteres fornecerá informações filogenéticas. Metodologias filogenéticas são essencialmente estatísticas e podem ser classificadas em dois

grupos principais, de acordo com seus critérios: 1) quantitativos (métodos de distância); 2) qualitativos (métodos de parcimônia e verossimilhança).

Nos métodos de distância, as diferenças entre duas seqüências são reduzidas a uma só variável (número de diferenças) e suas relações evolutivas não são consideradas. Nesse método, primeiramente calcula-se a distância, para, a seguir, reconstruir a árvore filogenética, utilizando um algoritmo específico. Existem diferentes modelos que podem ser utilizados para a construção de uma matriz de distância (NEI *et al.*, 2000; MATIOLI, 2001). A partir dessa matriz será reconstruída a árvore com um dado algoritmo, sendo os mais utilizados o UPGMA e Neighbor-Joining.

Nos critérios qualitativos, as diferenças entre as moléculas são consideradas como uma série de variáveis descontínuas. Os métodos mais empregados, embora não sejam os únicos, são a Máxima Parcimônia (MP) e a Máxima Verossimilhança (MV).

O princípio da MP é que a hipótese mais simples, ou seja, aquela em que a árvore que apresente o menor número de passos (=menor número de mudanças de estado de caráter) será a escolhida para explicar um determinado conjunto de dados. Desse modo, não há um modelo evolutivo explícito embasando essa metodologia. Além disso, podem ocorrer alguns problemas operacionais, como a obtenção de um número muito grande de árvores igualmente mais parcimoniosas ou a influência demasiada de homoplasias. Porém, a questão mais crítica é de cunho filosófico: A Natureza é parcimoniosa?

A MV baseia-se em modelos evolutivos explícitos de substituição de nucleotídeos. Esses modelos são avaliados quanto a sua probabilidade de explicar um conjunto de dados de forma que reflita a história evolutiva mais verossímil. O modelo que apresentar o melhor valor de verossimilhança, que, por questões operacionais, é dado em forma logarítmica, será o escolhido como base para a reconstrução da árvore. Uma descrição detalhada desses modelos pode ser encontrada em NEI *et al.* (2000).

Tanto para a MP, quanto para MV, existem dois tipos de algoritmo de busca da melhor árvore: 1) Busca Exaustiva, em que todas as possibilidades de topologias são verificadas; 2) Busca Heurística, em

que apenas um subconjunto de árvores probabilisticamente mais prováveis é examinado. Esta último, em termos práticos, reduz sensivelmente o tempo de processamento computacional que, em alguns casos, dependendo do número de seqüências comparadas, pode durar dias e até meses!

## REFERÊNCIAS BIBLIOGRÁFICAS

- MATIOLI, S.R. (Ed.). *Biologia molecular e evolução*. Ribeirão Preto: Holos, 2001.
- NEI, M. & KUMAR, S. *Molecular evolution and phylogenetics*. Oxford Univ. Press, 2000.