

Capítulo 8

Versão 0.4

Filogenômica

O termo "Filogenômica" é definido de várias maneiras, mas, em geral, a definição é relacionada com a intersecção dos campos da Genômica e da evolução biológica. Aqui, nesse capítulo, ele será restrito à questão da reconstrução de filogenias com dados de genomas completos.

Existem dois extremos de abordagem na Filogenômica. Uma primeira abordagem, que poderia ser chamada de aglomerativa, advoga que os dados de todo o genoma poderiam ser considerados em conjunto, como se fosse uma única molécula, e as análises filogenéticas seriam feitas a partir de uma única matriz, como visto no capítulo de filogenética. Essa abordagem é conhecida como o método da "super-matriz", ou da "evidência total".

Em algumas situações, a abordagem de molécula única faz todo o sentido do ponto de vista biológico. Os genomas mitocondriais, na maioria dos eucariontes, têm herança uniparental e baixa taxa de recombinação. Assim, não haveria qualquer motivo para que não se considerasse todo o genoma como uma única molécula (caso dos genomas circulares, ou mesmo lineares de muitas espécies), utilizando-se os procedimentos usuais.

Essa regra pode ser estendida para outros genomas de organelas, como os de cloroplastos. Isso não seria automático, no entanto, para qualquer genoma pequeno, como aqueles dos vírus. Durante a evolução de diversos vírus, os eventos de recombinação são frequentes e ocorrem nos casos de coinfeção de uma célula por duas ou mais linhagens diferentes.

Para organismos mais complexos, várias outras abordagens têm sido empregadas nos estudos filogenômicos. Essas abordagens podem ser consideradas como cada uma das gradações de um leque de opções. Em um extremo desse leque podemos situar a abordagem da "super-matriz".

Evidência total

A abordagem da evidência total é bastante simples e também tem o apelo da simplicidade epistemológica atribuída ao método de máxima parcimônia, como visto no capítulo sobre filogenética. Todo e qualquer caráter disponível é utilizado sem que se empregue qualquer tipo de pesagem diferente, seja o caráter um nucleotídeo de um "espaçador" intergênico ou a presença de um determinado folheto germinativo. Utiliza-se então o método de máxima parcimônia para a reconstrução da filogenia.

Do ponto de vista algorítmico, o método consiste na procura, dentre o universo de árvores filogenéticas possíveis, daquela que contenha o menor número de eventos de mudança de estado de caráter.

Informações além da estrutura primária do genoma

A organização dos genomas, além da sequência de nucleotídeos, possui certas características que podem ter sido modificadas durante a evolução dos organismos e que podem servir para a inferência das relações de parentesco das espécies que foram consideradas na análise.

A ordem dos genes nos genomas é determinada pelos eventos que tiveram sucesso com algum dos mecanismos de duplicação, deleção, translocação de trechos cromossômicos, assim como os mecanismos de fusão e fissão que envolvem cromossomos inteiros. Tradicionalmente esses eventos têm sido considerados na reconstrução filogenética desde a década de 1930 (quase 15 anos antes de se estabelecer o DNA como material genético!), com a observação microscópica de morfologia cromossômica de espécies próximas, como mostrado na figura 1.

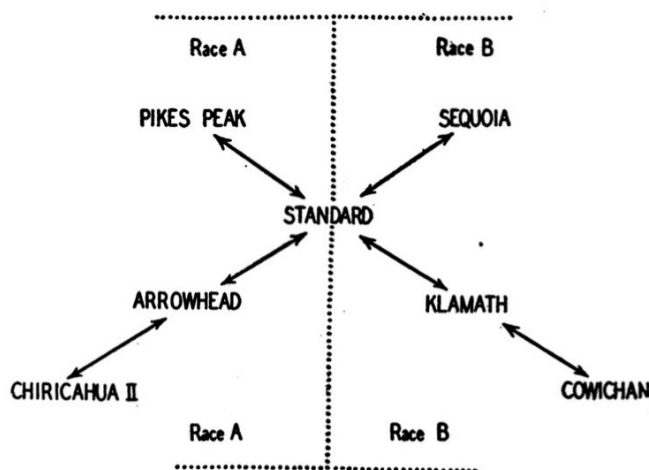


Figura 1. Filogenia de linhagens cromossômicas de *Drosophila pseudoobscura* baseada na reconstrução de seus rearranjos.

Os algoritmos que fazem a análise de rearranjos de genomas utilizam uma abordagem que se assemelha àquelas de *dot plots* e dos algoritmos de alinhamento par a par locais. O que é marcado, entretanto, é não só a coincidência de nucleotídeos mas também os possíveis emparelhamentos entre os nucleotídeos compatíveis.

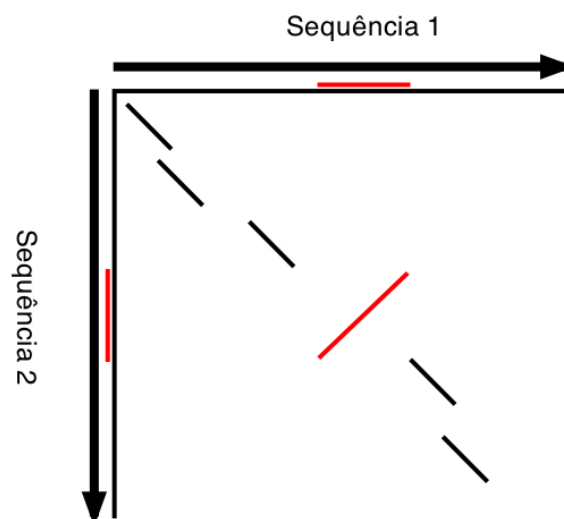


Figura 2. Estratégia de procura de rearranjos. Sequências de dois genomas são dispostas ortogonalmente e os trechos similares são detectados como nos algoritmos de alinhamento local. Os trechos invertidos de um genoma com

relação ao outro aparecem como regiões de possível emparelhamento (As com Ts e Cs com Gs), em vermelho na figura.

Uso de consenso ou "super-árvores"

Outra alternativa bastante empregada na análise de dados de genomas inteiros é a análise separada de trechos desses genomas. Em geral os trechos correspondem a unidades estruturais funcionais, como genes ou exons. Cada um desses trechos é utilizado para se fazer uma análise filogenética de acordo com algum dos procedimentos vistos no capítulo sobre filogenética. A seguir essas árvores são combinadas para possibilitar um panorama da evolução do organismo no nível molecular.

Consenso rigoroso

Normalmente traduzido como consenso estrito (do inglês *strict consensus*), é um procedimento que adota a politomia como uma forma de acomodar toda e qualquer relação filogenética que surgiu nas análises individuais.

Consenso de maioria

No consenso de maioria, são mostradas todos os agrupamentos que ocorreram em 50% ou mais das árvores. Na figura 3 estão mostrados os resultados dos dois tipos de consenso em um mesmo conjunto de árvores.

As árvores de consenso obtidas do conjunto de árvores de trechos separados de genomas inteiros tem sido descritas como sendo "super-árvores", em contraposição às árvores que são reconstruídas a partir de todos os dados disponíveis, que são conhecidas como árvores feitas a partir de "super-matriz".

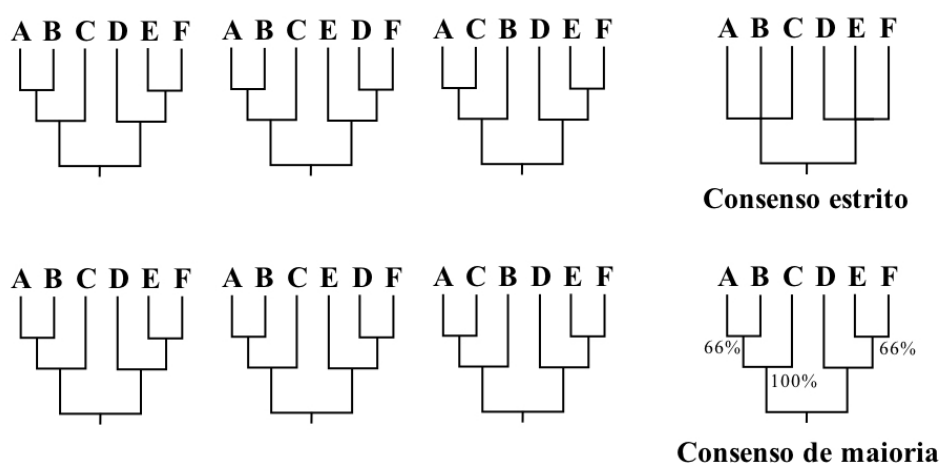


Figura 3. Consenso estrito e consenso de maioria em um mesmo conjunto de árvores.

Uma outra abordagem, que pode ser bastante útil por expressar a variedade existente, é a utilização de métodos gráficos que preservam cada uma das árvores independentes. Uma amostra do uso desse procedimento está mostrado na figura 4.

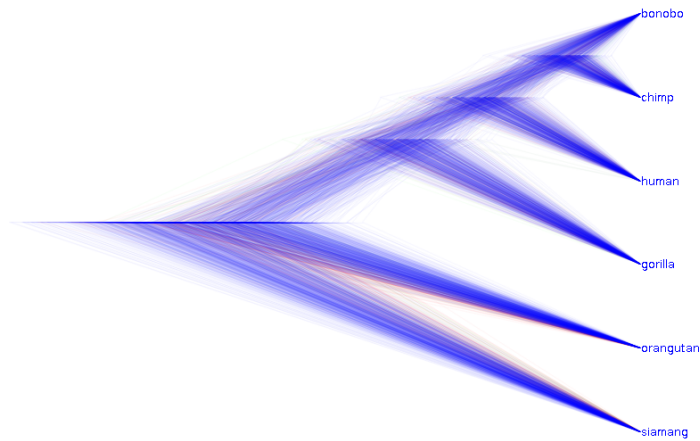


Figura 4. Representação onde árvores individuais são sobrepostas para mostrar a variabilidade delas. (produzida pelo programa Densitree).

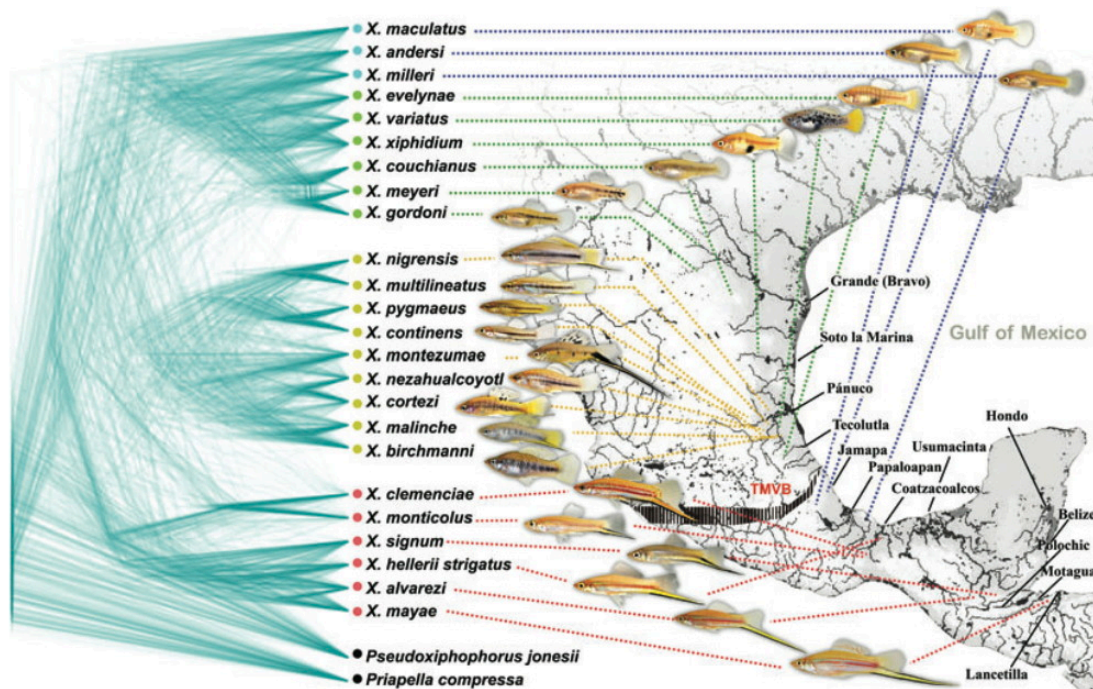


Figura 5. Filogenômica mostrando evolução parcialmente reticulada de peixes de água doce.

Uso de proteomas e exomas.

Sem se considerar os métodos de reconstrução filogenômica propriamente ditos, um aspecto importante é a escolha das parcelas dos genomas que serão utilizadas quando a análise dos genomas completos são inviáveis. Isso pode acontecer quando o estabelecimento de homologies minimamente aceitáveis é impraticável. Isso acontece, por exemplo, no caso de haver trechos abundantes e/ou longos dos genomas que divergiram muito durante a evolução pregressa dos organismos considerados.

Uma das maneiras de se selecionar trechos dos genomas mais utilizadas baseia-se na escolha dos trechos que são transcritos e traduzidos para

polipeptídeos. Um fato bastante bem estabelecido a partir dos estudos de evolução de polipeptídeos que remontam à década de 1970 é de que as sequências polipeptídicas evoluem muito mais lentamente quando se considera as sequências de polipeptídeos do que as sequências nucleotídicas codificantes correspondentes. A razão dessa diferença reside na natureza degenerada do código genético (correspondência entre trinças de nucleotídeos e resíduos de aminoácidos) e na maior intensidade da seleção que envolvem diferenças entre resíduos de aminoácidos nos polipeptídeos.

Grupos ortólogos

Mesmo depois de separados os trechos correspondentes aos proteomas das espécies que se pretende analisar, surge uma dificuldade. Muitos dos genes codificantes para polipeptídeos existem em famílias cujo número de exemplares pode diferir entre cada um dos grupos de genes parálogos nas espécies consideradas. Quando isso acontece em uma quantidade apreciável de grupos de genes parálogos, uma opção é a utilização de grupos ortólogos, onde um único exemplar das famílias de parálogos é escolhido. Nesse caso a análise é feita segundo o critério dos grupos ortólogos. No caso de organismos eucarióticos, o estabelecimento de homologias entre os trechos correspondentes aos introns pode ser problemática e a solução para isso é a escolha do conjunto de exons dos genomas, que é conhecido como "exoma".