Agricultura
Ministério da Agricultura, Pecuária e Abastecimento

Documentos

ISSN 0102-0110
Agosto, 2005 137

MARCADORES MOLECULARES E ANÁLISE FILOGENÉTICA

Embrapa

República Federativa do Brasil

Luiz Inácio Lula da Silva
Presidente

Ministério da Agricultura, Pecuária e Abastecimento

Roberto Rodrigues
Ministro

Empresa Brasileira de Pesquisa Agropecuária

Conselho de Administração

Luis Carlos Guedes Pinto
Presidente

Silvio Crestana
Vice-Presidente

Alexandre Kalil Pires
Ernesto Paterniani
Helio Tollini
Marcelo Barbosa Saintive
Membros

Diretoria-Executiva da Embrapa

Silvio Crestana
Diretor Presidente

José Geraldo Eugênio de França
Kepler Euclides Filho
Tatiana Deane de Abreu Sá
Diretores Executivos

Embrapa Recursos Genéticos e Bioteconologia

José Manuel Cabral de Sousa Dias
Chefe-Geral

Maurício Antônio Lopes
Chefe-Adjunto de Pesquisa e Desenvolvimento

Maria Isabel de Oliveira Penteado
Chefe-Adjunto de Comunicação e Negócios

Maria do Rosário de Moraes
Chefe-Adjunto de Administração

ISSN 0102 0110
Agosto, 2005

Embrapa
Recursos Genéticos e Biotecnologia

Documentos 137

Marcadores Moleculares e Análise Filogenética

Glaucia Salles Cortopassi Buso

Brasília, DF
2005

Exemplares desta edição podem ser adquiridos na

Embrapa Recursos Genéticos e Biotecnologia
Serviço de Atendimento ao Cidadão
Parque Estação Biológica, Av. W/5 Norte (Final) –
Brasília, DF CEP 70770-900 – Caixa Postal 02372 PABX: (61) 3348-4739 Fax: (61) 3340-3666 http://www.cenargen.embrapa.br
e.mail:sac@cenargen.embrapa.br

1ª edição
1ª impressão (2005):

B 976 Buso, Gláucia Salles Cortopassi.
Marcadores Moleculares e Análise Filogenética / Gláucia Salles Cortopassi Buso. – Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia, 2005.
22 p.: il. – (Documentos / Embrapa Recursos Genéticos e Biotecnologia, ISSN 0102-0110; 136)

1. Marcadores moleculares 2. Análise 3. Filogenética. II. Série.

574.87328 – CDD 20

AUTORA

Glaucia Salles Cortopassi Buso
Ph.D., Embrapa Recursos Genéticos e Biotecnologia, e-mail: buso@cenargen.embrapa.br.

SUMÁRIO

Marcadores Moleculares e Análise Filogenética

Glaucia Salles Cortopassi Buso

O conhecimento da evolução das espécies e a relação entre as espécies cultivadas e seus parentes silvestres tem sido de considerável valor na conservação e utilização dos recursos genéticos. Os parentes silvestres das plantas cultivadas potencialmente contêm muitos caracteres importantes que podem ser introduzidos nas espécies cultivadas. Estas introduções tendem a ser mais efetivas quando a espécie silvestre é parente próximo da espécie cultivada alvo ou mesmo ancestral direto da mesma. Portanto, o conhecimento da relação entre as espécies é de extrema importância para a utilização eficiente dos recursos genéticos. Além disso, o conhecimento da relação entre as espécies e sua evolução é fundamental para a sistemática que visa o estudo dos padrões evolucionários da diversidade biológica. Ela inclui a classificação e a identificação das espécies, da identificação e entendimento das suas relações, e o estudo de desenvolvimento das espécies e sua relação num determinado período. A sistemática abrange duas grandes áreas da biologia: Taxonomia: classificação dos organismos com base em similaridades de estruturas ou características; Filogenia: eventos evolucionários que levam à relação entre os organismos.

A filogenia lida com a identificação e entendimento das relações entre as espécies que resultam da evolução. Sob a premissa que as espécies evoluem de um ancestral comum, as espécies mais próximas têm mais características em comum do que as mais distantes. A sistemática filogenética é utilizada para identificar características similares e definir a relação histórica ou evolucionária com base nessas características – incluindo genes ou fragmentos de DNA. Com base nessas relações, árvores filogenéticas são desenvolvidas mostrando as relações evolucionárias ou agrupamentos de organismos.

Utilização de DNA na análise filogenética

Com o aprimoramento e facilidade das técnicas moleculares, estas têm sido muito utilizadas nas análises filogenéticas. A maior vantagem dos métodos moleculares é a investigação direta da situação genotípica, o que permite a detecção de variação ao nível de DNA, excluindo, portanto, influencias ambientais. Além disso, a análise pode ser feita em estágios de desenvolvimento primordiais. Dependendo da metodologia escolhida os métodos moleculares podem ser mais sensíveis a qualquer diferença genética e detectar muito mais diversidade genética do que os métodos clássicos de caracterização morfológica. Os marcadores moleculares têm se tornado ferramentas poderosas na análise filogenética.

Algumas áreas do DNA são neutras com respeito à seleção, elas não evoluem por meio da seleção natural porque as mutações nestas áreas não

afetam o fenótipo do organismo. A seleção natural depende de diferenças no fenótipo que representem diferenças de adaptação; se indivíduos com mutações diferentes no DNA tiverem o mesmo fenótipo, eles terão provavelmente a mesma capacidade de adaptação. Algumas destas áreas são:

1) DNA não codante ("junk DNA"): áreas que aparentemente não codificam nenhum RNA. Estas áreas são de tamanho variado e mutações nelas não afetam o fenótipo;

2) posições na região codante do DNA onde as mutações não afetam o amino-ácido codificado naquela posição – Substituição silenciosa – Frequentemente estas substituições silenciosas são mutações na terceira base da seqüência que codifica a proteína; geralmente seqüências de três bases com as mesmas primeiras bases, só diferindo na terceira, codificam o mesmo aminoácido;

3) DNA que codifica introns: intron é uma seqüência que é retirada do mRNA antes que ele seja utilizado como "template" para proteínas. O DNA que codifica introns, portanto não é base para codificação das proteínas e mutações nos introns são aparentemente neutras com respeito à seleção.

Ao contrário, outras áreas do DNA podem afetar o fenótipo e então não são neutras com respeito à seleção. São áreas do DNA que incluem DNA codante de proteína e DNA que codifica tRNA ou mRNA. Se o DNA codifica para alguma característica que quando sofre mutação afetará o fenótipo, muitas mutações provavelmente resultarão num fenótipo que não funciona apropriadamente. Estas mutações serão letais, portanto serão perdidas ou permanecerão a uma freqüência muito baixa nas populações. Como somente poucas mutações nestas áreas seguem adiante, a evolução nestas áreas tende a ser muito lenta. Ao contrário, nas áreas de DNA neutro com respeito à seleção, qualquer mutação pode ir adiante e vir a ser comum e a evolução provavelmente será mais rápida. Por exemplo, o DNA que codifica rRNA evolui muito lentamente, provavelmente porque mutações nesta área do DNA resultariam em ribossomos não funcionais.

Em resumo, quando se utiliza DNA para análise filogenética deve-se escolher DNA que: 1) evolua rápido o bastante para que as espécies em estudo mostrem diferenças umas das outras; 2) evolua devagar o bastante para que não haja muita divergência. Se quisermos conhecer a relação de parentes próximos como espécies dentro de um gênero, escolheremos áreas do DNA que evoluem rapidamente, ou seja, áreas neutras do DNA, como introns. Se quisermos saber a relação de parentes distantes, como famílias ou ordens dentro de classes diferentes, deveríamos usar DNA que evolua a uma velocidade moderada, como codante de proteínas. Para grupos muito distantes, como classes ou filos, provavelmente usaríamos DNA que codifica ribossomos.

Além do DNA ribossomal, os polimorfismos de seqüências de DNA no genoma nuclear fornecem um número praticamente inexaurível de marcadores, quando acessados eficientemente. Para análises de variação genética, tem-se dado atenção a seqüências de cópias simples e seqüências repetitivas,

particularmente as hipervariáveis como VNTRs e microssatélites, e rDNA (FERREIRA e GRATTAPAGLIA, 1995; DOWLING et al., 1996).

Em plantas, nos cloroplastos, a organela responsável pela fotossíntese, existem genes para proteínas especiais sintetizadas pelo cloroplasto, rRNAs e tRNAs específicos para produção de proteínas ribossomais do cloroplasto. Vários componentes do aparato fotossintético são codificados no cloroplasto, incluindo a maior das subunidades da enzima responsável pela fixação do dióxido de carbono. Estudos recentes sobre a evolução do genoma do cloroplasto têm revelado um alto grau de conservação em tamanho, estrutura, conteúdo gênico e ordem linear dos genes entre espécies aparentadas. Este modo conservativo de evolução do cpDNA sugere que qualquer mudança em estrutura, arranjo ou conteúdo do genoma pode ter implicações filogenéticas significativas. Na maioria dos aspectos, a evolução molecular dos genes do cloroplasto é similar à dos genes nucleares. No entanto, os genes que codificam proteínas do cloroplasto evoluem a uma taxa cinco vezes mais lenta do que os genes nucleares (CLEGG e ZURAWSKI, 1992). A taxa média de substituições sinônimas dos genes do cpDNA variam de aproximadamente 0.2 a $1.0x10^{-9}$ substituições por sítio de DNA por ano. É provável que as reduzidas taxas de evolução gênica sejam conseqüência de uma baixa taxa de mutação neste genoma (CLEGG e ZURAWSKI, 1992). As mutações no cpDNA são de dois tipos: substituições de nucleotídeos (mutações de ponto) e rearranjos. A detecção de substituições através de restrição ou comparação direta de seqüências tem sido muito utilizada na reconstrução filogenética. Os rearranjos maiores incluem inversões, inserções ou deleções de genes e introns, e perda de uma cópia do IR ("inverted repeat"). Estes eventos são usualmente detectados com sondas, na reconstrução filogenética de grupos taxonômicos superiores (comparação de famílias, gêneros etc.) ou seqüenciamento. Os rearranjos menores incluem pequenas inserções e deleções (1-1000 bp) e ocorrem principalmente em regiões não codificantes. Portanto, a variação do cpDNA tem provado ser imensamente válida na reconstrução filogenética ao nível de espécies (DOWNIE e PALMER, 1992).

Um fator que distingue a evolução de genes do cpDNA daquela de genes nucleares é a falta de atividade de transposons, não evidenciada no genoma do cloroplasto. Devido à rara transmissão biparental e à baixa diversidade intraespecífica das seqüências de cpDNA, os processos recombinatórios não têm papel importante na evolução de seqüências de DNA cloroplástico. Outras razões da adequação do cpDNA ao estudo filogenético incluem: facilidade de extração e análise por ser um componente abundante do DNA total; quantidade de informação genética-molecular (em algumas plantas, incluindo o arroz, o cpDNA já foi totalmente sequenciado); taxas conservativas de substituição de nucleotídeos. A taxa conservativa de evolução do cpDNA permite que clones ou primers específicos de cpDNA possam ser utilizados virtualmente para todo o reino vegetal, facilitando a comparação de diferentes taxons. Ao mesmo tempo, a taxa conservativa de substituição de nucleotídeos, permite que a detecção de polimorfismos geralmente assegure soluções para estimativas de relações filogenéticas. É neste nível que os métodos convencionais de inferência filogenética são necessários para evitar problemas

de mudanças de caráter paralelas e convergentes (CLEGG e ZURAWSKI, 1992).

A variação do cpDNA tem sido usada desde 1980 para melhor entendimento de problemas filogenéticos, incluindo inferências sobre a origem de plantas cultivadas, identificação do ancestral materno e paterno de híbridos e de espécies poliplóides, detecção de introgressão, e identificação do gênero progenitor de um outro isolado morfologicamente (DOWLING et al., 1990). O modelo de variação de seqüência de cpDNA ou outras seqüências baseia-se na proporcionalidade de acúmulo de mudança de seqüência em relação ao tempo, originalmente conhecida como relógio biológico (ZUCKERKANDL e PAULING, 1962). A análise de restrição do cpDNA em vários gêneros (RAINA e OGIHARA, 1994; LLACA et al., 1994; BADENES e PARFITT, 1995; TSUMURA et al., 1995; ZUNK et al., 1996; PROVAN et al., 1997) tem geralmente confirmado os agrupamentos derivados de classificações tradicionais, proporcionando um quadro mais claro e detalhado de informações.

O mtDNA de plantas é muito pouco estudado em comparação ao cpDNA. Aspectos da estrutura genômica, como tamanho do genoma, configuração e ordem gênica mudam extremamente rápido na mitocôndria de plantas, enquanto a seqüência primária do genoma é excepcionalmente lenta para mudar. A taxa de substituição de nucleotídeos é de três a quatro vezes menor que no cpDNA (PALMER, 1992). Portanto, a evolução estrutural do mtDNA de plantas parece muito maior do que no cloroplasto. Por outro lado, o mtDNA evolui mais lentamente do que os genomas do cloroplasto e nuclear quando as seqüências primárias são consideradas. Como no cpDNA, a herança maternal parece predominar no mtDNA das angiospermas.

Relógio molecular

Foi mostrado que devemos esperar que diferentes áreas do DNA evoluam a taxas diferentes. Agora consideremos outra possibilidade: a mesma área do DNA, por exemplo, um mesmo gene, em espécies diferentes, tenham a mesma taxa de evolução. Esta possibilidade é a base da hipótese do relógio molecular. Esta hipótese diz que assumindo que o mesmo gene evolui à mesma taxa em diferentes espécies, diferenças entre espécies diretamente refletem a época da especiação.

Métodos de inferência da filogenia

Os métodos usados para determinar as diferenças entre as unidades de estudo dependem do número e tipo de características a serem comparadas. Características podem incluir diferenças morfológicas, ou diferenças em seqüências de DNA, genes ou seus produtos (proteínas). Uma árvore geralmente deve ser feita a partir da comparação de múltiplos caracteres.

Existem dois tipos de análises utilizadas para o desenvolvimento de árvores filogenéticas. Os métodos **fenéticos** utilizam medidas de distância, que consolidam estatisticamente as diferenças entre os caracteres em um número. Uma matriz de distância entre todos os possíveis pares do grupo de estudo é

criada, e árvores são desenvolvidas agrupando aqueles com menor diferença num fenograma.

Os métodos **cladísticos** calculam árvores para cada caráter e então indicam a melhor árvore através da determinação daquela que requer menor número de mudanças (parcimônia) ou aquela mais provável estatisticamente (máxima verossimilhança). A idéia básica por trás da cladística é que membros do mesmo grupo compartilham uma história evolucionária comum, e são proximamente relacionados, mais com membros do mesmo grupo do que com outros organismos. Estes grupos são reconhecidos por compartilharem características únicas que não estão presentes num ancestral distante. O método de Máxima Verossimilhança avalia a hipótese da história evolucionária em termos da probabilidade de que um modelo proposto e a história hipotética originariam um conjunto de dados. A suposição é que a história com maior probabilidade de alcançar o estado observado é preferida à história com menor probabilidade. O método procura a árvore com maior probabilidade ou verossimilhança.

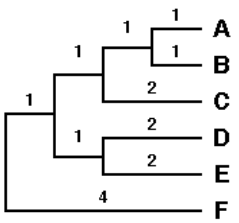
Métodos para construção das árvores filogenéticas

Construção de árvores filogenéticas a partir de dados de distância, utilizando agrupamento de pares não ponderados baseado na média aritmética (UPGMA).

UPGMA é o método mais simples de construção de árvores filogenéticas a partir de dados de distância. Ele foi desenvolvido originalmente para construção de fenogramas taxonômicos, ou seja, árvores que refletem similaridades fenotípicas entre táxons, mas pode também ser usado para construir árvores filogenéticas se as taxas de evolução forem aproximadamente constantes. UPGMA emprega um algorítmo seqüencial de agrupamento, no qual as relações são identificadas em ordem de similaridade, e a árvore é construída passo a passo. Primeiro, identifica-se entre todas as unidades estudadas as duas mais similares e aí trata como se fosse uma unidade. Subsequentemente, do resto do grupo identifica-se outra unidade com maior similaridade, e assim por diante.

Explicação do método

Suponha a árvore seguinte que consiste de 6 táxons:



As distâncias entre taxons são dadas pela matriz de distância seguinte:

	A	B	C	D	E

B	**2**				
C	4	4			
D	6	6	6		
E	6	6	6	4	
F	8	8	8	8	8

O método agrupa o par de táxons com a menor distância, sendo A e B, que estão separados a uma distância de 2 unidades. O ponto de formação dos galhos é posicionado a distância de 2 / 2 = 1. Portanto, uma sub-árvore é construída como se segue:

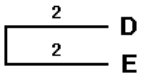
O primeiro agrupamento então é considerado um táxon(A,B) composto e sua distância com os outros táxons é calculada, como se segue:

dist(A,B),C = (distAC + distBC) / 2 = 4
dist(A,B),D = (distAD + distBD) / 2 = 6
dist(A,B),E = (distAE + distBE) / 2 = 6
dist(A,B),F = (distAF + distBF) / 2 = 8

Em palavras, a distância entre o táxon simples e o táxon composto é a média das distâncias entre o táxon simples e os constituintes do táxon composto. Então uma nova matriz de distância é calculada usando as novas distâncias e o ciclo é repetido:

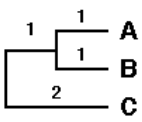
Segundo ciclo

	A,B	C	D	E
C	4			
D	6	6		
E	6	6	**4**	
F	8	8	8	8



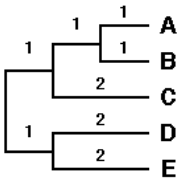
Terceiro ciclo

	A,B	C	D,E
C	**4**		
D,E	6	6	
F	8	8	8



Quarto ciclo

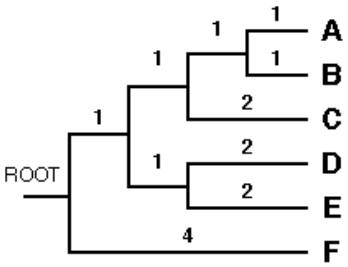
	AB,C	D,E
D,E	**6**	
F	8	8



Quinto ciclo

	ABC,DE
F	**8**

Embora o método leve a uma árvore não enraizada, UPGMA assume mesma taxa de evolução ao longo dos galhos da árvore, como modelo de evolução. O enraizamento teórico, portanto, fica equidistante de todos os táxons: dist (ABCDE),F / 2 = 4.



Análise de Máxima Parcimônia

Parcimônia implica que as hipóteses mais simples são preferíveis às mais complicadas. Como já foi dito, a parcimônia é um método baseado no estado do caráter que infere a árvore filogenética minimizando o número total de passos evolucionários para explicar um conjunto de dados. Os passos podem constituir na substituição de base do DNA numa seqüência ou perda de um sítio de restrição.

O método de parcimônia procura todas as topologias possíveis para achar uma árvore adequada. No entanto, o número de árvores não enraizadas que tem que ser analisadas aumenta rapidamente com o número de táxons. O número de árvores enraizadas é dado por:

- *Nr* = (2*n* -3)!/(2exp(*n* -2)) (*n* -2)!

O número de árvores não enraizadas para n táxons é dado por:

- *Nu* = (2*n* -5)!/(2exp(*n* -3)) (*n* -3)!

Alguns resultados estão mostrados na tabela:

Número de OTUs	Número de árvores não enraizadas	Número de árvores enraizadas
2	1	1
3	1	3
4	3	15
5	15	105
6	105	945
7	954	10,395
8	10,395	135,135
9	135,135	34,459,425
10	34,459,425	2.13E15
15	2.13E15	8.E21

Este rápido aumento no número de árvores a serem analisadas pode tornar impossível a aplicação do método em conjuntos grandes de dados (SWOFORD, 1991)

Explicação do método

Um exemplo do método de máxima parcimônia para uma matriz de 4 seqüências de DNA é dado abaixo:

Seqüência 1 2 3 4 5 6 7 8 9

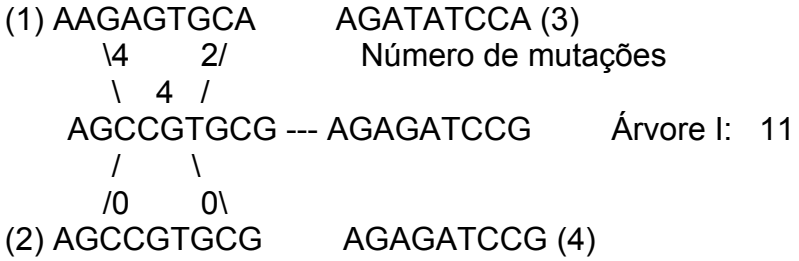
1 A A G A G T G C A

2 A G C C G T G C G

3 A G A T A T C C A

4 A G A G A T C C G

Para os 4 táxons existem 3 árvores não enraizadas possíveis. As 3 são então analisadas por meio da procura da seqüência ancestral e contando-se o número de mutações requeridas para explicar as respectivas árvores:



A árvore I tem a topologia com o menor número de mutações e, portanto é a mais parcimoniosa.

A máxima parcimônia procura pela árvore mínima. Neste processo mais do que uma árvore pode ser encontrada. Para garantir que a melhor árvore será encontrada uma avaliação exaustiva de todas as topologias tem que ser feita. Entretanto, isto é impossível quando se tem mais do que 12 táxons a serem analisados. Alternativas:

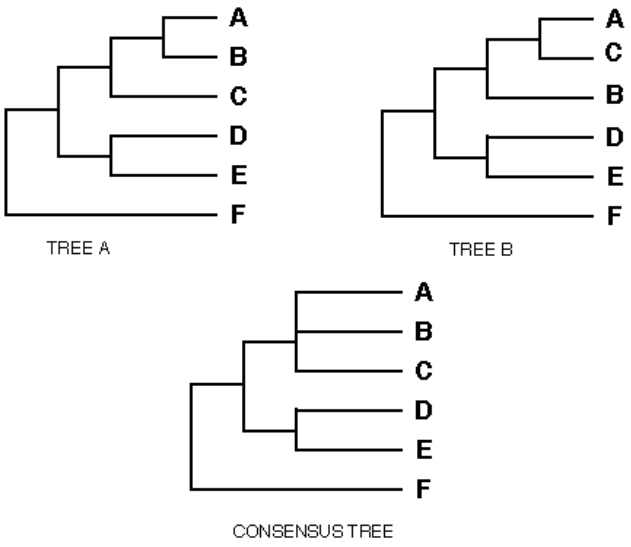
"Branch and Bound": é uma variação da parcimônia máxima, que garante achar a árvore mínima sem ter que avaliar todas as possíveis árvores. Desta forma um maior número de taxas pode ser avaliado. Ele realiza uma busca heurística inicial para obter uma árvore próxima do ótimo. O escore desta árvore será usado como limite superior para corte.

Procura Heurística (Heuristic searches): quando os dados contêm um número elevado de táxons, as árvores têm que ser avaliadas por meio de

métodos heurísticos. Pode-se iniciar a procura heurística de modos diferentes: 1) adição passo a passo: é um método com adição passo a passo (“step-wise”) e rearranjo dos táxons. O processo inicia-se com três táxons. Em seguida, um táxon é adicionado. Cada uma das três árvores resultantes é avaliada e a que tem melhor escore é armazenada para a próxima adição. Não há garantia de que a árvore ótima seja obtida. 2) rearranjo dos ramos (“branch swapping”): envolve a mudança de ramos para novas partes da árvore produzindo novas topologias.

Árvores de consenso

Já que os métodos de máxima parcimônia podem resultar em mais do que uma árvore parcimoniosa, uma árvore de consenso deve ser criada.



Método de máxima verossimilhança

É um método de inferência filogenética que avalia a hipótese sobre a história evolucionária em termos da probabilidade de que o modelo proposto e a história hipotética dariam origem ao conjunto de dados observados. A suposição é que a história com maior probabilidade de atingir o estado observado é preferida à história com mais baixa probabilidade. O método procura pela árvore com maior probabilidade ou maior verossimilhança.

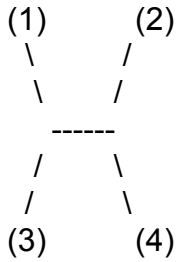
Explicação do método

O método de máxima verossimilhança avalia a probabilidade de que um modelo evolucionário escolhido tenha gerado as seqüências observadas. A filogenia é inferida achando-se as árvores com maior verossimilhança.

Suponha as seqüências de nucleotídeos para 4 táxons:

```
        1       j      ....N
(1)     A G G C U C C A A ....A
(2)     A G G U U C G A A ....A
(3)     A G C C C A G A A.... A
(4)     A U U U C G G A A.... C
```

e deseja-se avaliar a verossimilhança da árvore representada pelos nucleotídeos do sítio j na seqüência:

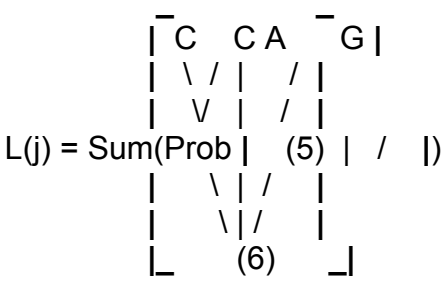


Qual a probabilidade de que esta árvore tenha sido gerada pelos dados presentes na seqüência, sob a visão do modelo escolhido?

Já que a maioria dos modelos utilizados é reversível no tempo, a verossimilhança da árvore é geralmente independente da posição da raiz. Portanto, seria conveniente enraizar a árvore num nó arbitrário, como no exemplo,

Sob a hipótese de que cada sítio do nucleotídeo evolui independentemente (modelo de evolução Markovian), pode-se calcular a verossimilhança para cada local separadamente e combinar as verossimilhanças num valor total. Para calcular a verossimilhança para o sítio j, tem que se considerar todos os cenários possíveis pelos quais os nucleotídeos presentes nas pontas da árvore possam ter evoluído. Então a verossimilhança para um local particular é a soma das probabilidades de todas as reconstruções possíveis de estados ancestrais, dado o modelo de substituição de bases. Neste caso, todos os

nucleotídeos possíveis A, G, C, and T podem ocupar os nós (5) e (6), ou 4 x 4 = 16 possibilidades:



L(j) = Sum(Prob

Após a estimativa das probabilidades para o sítio j e considerando-se todos os nucleotídeos possíveis nos nós 5 e 6, o processo é repetido para todos os sítios até que todas as probabilidades tenham sido estimadas. Aceitando-se que todos os sítios evoluem independentemente, a estimativa da verossimilhança para toda a seqüência será igual ao produto das probabilidades de cada sítio. Isso é repetido para todas as árvores possíveis e aquela com maior verossimilhança será escolhida. Para um grande número de táxons torna-se impossível a procura devido ao grande esforço computacional.

Como enraizar uma árvore filogenética?

- A maioria dos métodos produz árvores não enraizadas. Portanto, olhando-se a árvore somente, é impossível dizer em que táxon começou a ramificação.
- Para enraizar uma árvore deve-se adicionar um grupo externo (“outgroup”) ao conjunto de dados. Um grupo externo é um táxon para o qual se tem informação que indica que o mesmo está ramificado fora daquele grupo a ser analisado.

Algumas observações

- Não escolher um grupo externo muito distante do grupo a ser analisado. Isto pode resultar em erros topológicos porque os sítios podem ficar saturados de mutações;
- Não escolher um grupo externo muito próximo do grupo a ser analisado, pois ele pode não ser um verdadeiro grupo externo;
- A utilização de mais do que um grupo externo geralmente melhora a topologia da árvore estimada;
- Na ausência de um grupo externo, o enraizamento pode ser posicionado no meio do caminho mais longo entre dois táxons, assumindo-se a mesma taxa evolucionária em todos os ramos.

Como utilizar os programas de construção de árvores filogenéticas?

Para se ter uma idéia da consistência da topologia da árvore filogenética resultante, as observações abaixo devem ser seguidas:

- Aplicar mais do que um método de análise (distancia, máxima parcimônia, máxima verossimilhança) aos dados;
- Variar os parâmetros dos diferentes programas;
- Quando estiver em dúvida, aplicar vários modelos evolucionários para construção das matrizes;
- Adicionar ou remover um ou mais táxons e ver como isto influencia a topologia da árvore;
- Incluir no mínimo um grupo externo para enraizamento da árvore;
- Aplicar análises de "Bootstrap" ou "Jacknife" aos dados e preparar uma árvore consenso com 100 - 1000 replicas (dependendo do tamanho do conjunto de dados e da capacidade do computador). No caso da análise de bootstrap somente nós que ocorrem mais do que em 95% dos casos são confiáveis.

Somente quando métodos bem diferentes produzem topologias similares ou idênticas apoiados por valores de bootstrap acima de 95%, as árvores filogenéticas podem ser consideradas confiáveis.

Limitações dos métodos descritos

- Métodos baseados em distância não usam os dados originais, mas informação de distância derivada destes dados;
- Métodos baseados em estado do caráter utilizam os dados originais, mas utilizam somente uma fração dos dados (somente sítios informativos).

Exemplos de utilização de marcadores moleculares em estudos filogenéticos

Num estudo com espécies silvestres de arroz (BUSO et al, 2001) foram utilizados marcadores que amostram ao acaso regiões de diferentes taxas de mutação (RAPD) e regiões conservadas do genoma (análise de CAPS de cpDNA e mtDNA). Os resultados foram concomitantemente analisados fenética e cladisticamente. Duzentos e trinta acessos de *Oryza* foram utilizados numa seqüência de avaliações genéticas. Desses, cento e vinte e três distribuíam-se entre as espécies silvestres brasileiras de *Oryza*, que representam a maior amostra de *Oryza* americana já estudada. Inicialmente a citometria de fluxo, contagem cromossômica e marcadores RAPD genoma-específicos foram empregados para se verificar a classificação de ploidia e de espécies de todo o material. Cerca de 8% dos acessos foram reclassificados com base no uso conjunto destas metodologias. Em seguida esta amostra do gênero *Oryza* foi estudada para avaliar a diversidade genética e as relações filogenéticas de acessos silvestres de arroz coletados na América, África, Ásia e Oceania. Cladogramas e fenogramas foram em geral concordantes com as

classificações disponíveis, baseadas em dados morfológicos e citogenéticos. No entanto, os resultados indicaram que a espécie americana diplóide, *O. glumaepatula*, deveria ser considerada uma espécie distinta, e não como uma forma americana da espécie *O. perennis* ou *O. rufipogon*, como é sugerido na literatura. Confirmou-se que as três espécies tetraplóides de genoma CCDD aparentemente constituem uma única espécie e não espécies distintas, como também é sugerido na literatura. Os dados indicam ainda que as tetraplóides americanas surgiram provavelmente de um evento único de poliploidização. Ficou evidente que o citoplasma mais próximo dos tetraplóides americanos é do genoma CC e não houve indicação de outra espécie diplóide próxima ao genoma CCDD, persistindo assim várias questões sobre a origem do genoma DD: se encontra-se extinto, ou talvez não tenha sido reconhecido, ou ainda não tenha sido coletado.

Pela análise cladística, o grupo *O. sativa* constituiu o parente mais próximo de *O. glumaepatula*. No entanto, apesar da proximidade, os dados de polimorfismo de cpDNA permitiram estimar que a separação entre essas duas espécies ocorreu há 20 milhões de anos o que se aproxima da hipótese de Chang (1985), que considera que ancestrais das espécies modernas de *Oryza* se separaram com a divisão do continente de "Gondwanaland". Esta estimativa concorda ainda com a hipótese de que as espécies americanas são derivadas de ancestrais provenientes da África no período terciário (de 66,4 milhões à 1.6 milhões de anos), quando os continentes assumiram progressivamente sua configuração atual (CHATTERJEE, 1951). Pela análise de cpDNA, estimou-se que a separação entre ancestrais de genomas CC e CCDD ocorreu também, há aproximadamente 20 milhões de anos.

CAPS de DNA de cloroplasto foi também utilizada no estudo da relação filogenética em *Ananas* e gêneros relacionados (DUVAL et al., 2003). Neste estudo foram utilizados 115 acessos representando sete espécies de *Ananas* e sete outras *Bromelioideae* incluindo o gênero monoespecífico *Pseudananas*, dois *Pitcairnioideae* e um *Tillandsioideae*. As análises cladística e fenética deram resultados semelhantes confirmando alguns pontos polêmicos na classificação deste gênero. Confirmou-se a posição basal de *Bromelia* em *Bromelioideae*. *Ananas* e *Pseudananas* formaram um grupo monofilético, com três subgrupos. A maioria dos acessos de *A. parguazensis* constituem um grupo restrito ao Rio Negro e bacia do rio Orinoco. *A. ananassoides* mostrou-se muito próximo das espécies cultivadas, dando suporte à hipótese de que este é o ancestral silvestre do abacaxi domesticado. Os dados indicaram que o fluxo gênico é comum entre estas espécies. A comparação dos dados de cpDNA com dados publicados de DNA genômico apontam para a origem híbrida de *A. bracteatus* e apóia a autopoliploidia de *Pseudananas*. Os resultados e hipóteses surgidas deste estudo sugerem alguns pontos que devem ser reconsiderados na atual discussão sobre a taxonomia do gênero *Ananas*.

REFERÊNCIAS BIBLIOGRÁFICAS

BADENES, M. L.; PARFITT, D. E. Phylogenetic relationships of cultivated *Prunus* species from an analysis of chloroplast DNA variation. **Theoretical and Applied Genetics,** Berlin, v. 90, p. 1035-1041, 1995.

BUSO, G. S. C.; RANGEL, P. H.; FERREIRA, M. E. Analysis of random and specific sequences of nuclear and cytoplasmic DNA in diploid and tetraploid American wild rice species. **Genome**, Ottawa, v. 44, p. 476-494, 2001.

CHANG, T. T. Crop history and genetic conservation: Rice - a case study. **Iowa State Journal of Research,** v. 59, p. 425-455, 1985.

CHATTERJEE, D. A modified key and enumeration of species of *Oryza*. **Indian Journal of Agricultural Sciences,** v. 18, p. 185-192, 1951.

CLEGG, M. T.; ZURAWSKI, G. Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In: SOLTIS, P. S.; SOLTIS, D. E.; DOYLE, J. J. (Ed.) **Molecular systematics of plants.** New York: Chapman & Hall, 1992. p. 1-13.

DOWLING, T. E.; MORITZ, C.; PALMER, J. D.; RIESEBERG, L. H. Nucleic acids III: analysis of fragments and restriction sites. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. (Ed.). **Molecular Systematics.** Sunderland, Massachusetts: Sinauer Associates, 1996. p. 249-320.

DOWLING, T. E.; MORITZ, C.; PALMER, J. D. Nucleic acids II: restriction-site analysis. In: HILLIS, D. M.; MORITZ, C. (Ed.). **Molecular Systematics**. Sunderland, Massachusetts: Sinauer Associates, 1990. p. 250-315.

DOWNIE, S. R.; PALMER, J. D. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: SOLTIS, P. S.; SOLTIS, D. E.; DOYLE, J. J. (Ed.). **Molecular systematics of plants.** New York: Chapman & Hall, 1992. p. 14-35.

DUVAL, M. F.; BUSO, G. S. C.; FERREIRA, F. R.; NOYER, J. L.; D´EECKENBRUGGE, G. C.; HAMON, P.; FERREIRA, M. E. Relationships in Ananas and other related genera using chloroplast DNA restriction site variation. **Genome,** Ottawa, v. 46, n. 6, p. 1-15, 2003.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores RAPD e RFLP em análise genética.** 2. Ed. Brasília: EMBRAPA-CENARGEN, 1998.

LLACA, V.; SALINAS, A. D.; GEPTS, P. Chloroplast DNA as an evolutionary marker in the *Phaseolus vulgaris* complex. **Theoretical and Applied Genetics,** Berlin, v. 88, p. 646-652, 1994.

PALMER, J. D. Mitochondrial DNA in plant systematics: applications and limitations. In: SOLTIS, P. S.; SOLTIS, D. E.; DOYLE, J. J. (Ed.) **Molecular systematics of plants**. New York: Chapman & Hall, 1992. p. 36-49.

PROVAN, J.; CORBETT, G.; MCNICOL, J. W.; POWELL, W. Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. **Genome,** Ottawa, v. 40, n. 104-110, 1997.

RAINA, S. N.; OGIHARA, Y. Chloroplast DNA diversity in *Vicia faba* and its close relatives: implications for reassessment. **Theoretical and Applied Genetics,** n. 88, p. 261-266, 1994.

SWOFFORD, D. L. **Phylogenetic Analysis Using Parsimony (PAUP),** version 3.0s. Champaign, Illinois: Illinois Natural History Survey, 1991.

TSUMURA, Y.; YOSHIMURA, K.; TOMARU, N.; OHBA, K. Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. **Theoretical and Applied Genetics,** Ottawa, n. 91, p. 1222-1236, 1995.

WESTON, P. H.; CRISP, M. D. **Introduction to Phylogenetic Systematics.** Disponível em: < http://www.science.uts.edu.au/sasb/WestonCrisp.html >. Acesso em: 2005.

ZUCKERKANDL, E.; PAULING, L. Evolutionary divergence and convergence in proteins. In: BRYSON, V.; VOGEL, H. J. (Ed.) **Evolving Genes and Proteins.** New York: Academic Press, 1965. p. 97-166.

ZUNK, K.; MUMMENHOFF, K.; KOCH, M.; HURKA, H. Phylogenetic relationships of *Thlaspi* s.l. (subtribe Thlaspidinae, Lepidieae) and allied genera on chloroplast DNA restriction-site variation. **Theoretical and Applied Genetics,** Ottawa, n. 92, p. 375-381, 1996.