

NOÇÕES BÁSICAS DE SISTEMÁTICA FILOGENÉTICA **4** TÓPICO

Sônia Godoy Bueno Carvalho Lopes
Fanly Fungyi Chow Ho

4.1 Introdução

4.2 Escolas de classificação baseadas em princípios evolutivos

4.3 O método filogenético Interpretando uma árvore filogenética

4.4 Tipos de dados utilizados em inferência filogenética

4.5 Utilização da metodologia e principais limitações

4.1 Introdução

Como estudamos nos tópicos anteriores, todos os organismos vivos e fósseis descendem de um único ancestral comum. Todos evoluíram através de um processo de **descendência com modificação**, dando origem às milhares de formas hoje existentes.

No Tópico 1 foi apresentado como a humanidade historicamente tentou organizar e classificar esta incrível diversidade de animais, plantas, fungos e micro-organismos. Neste tópico, discutiremos como se deu o desenvolvimento das técnicas modernas de reconstrução filogenética, sua utilização para classificação e quais os principais fundamentos teóricos e metodológicos para sua prática.

Objetivos

Espera-se que o aluno compreenda:

- o que é sistemática filogenética: origens, principais metodologias e fundamentos teóricos, usos e limitações deste paradigma.

4.2 Escolas de classificação baseadas em princípios evolutivos

Há duas escolas principais que se pautam no princípio evolutivo central de **descendência com modificação**: a evolutiva e a filogenética ou cladística.

Ambas partem de um aspecto simples, porém fundamental: organismos com relação de parentesco próxima são mais semelhantes que organismos com relação de parentesco relativamente mais distante. Isto porque parentes próximos tendem a herdar características que estavam presentes em um ancestral em comum. É fácil notar como irmãos parecem mais entre si do que quando comparados com outros parentes mais distantes ou outras pessoas.

Essas escolas divergem, no entanto, no modo como interpretam as relações de parentesco.

Na escola evolutiva, foi partindo desta simples observação que foram desenvolvidas classificações buscando simplesmente agrupar organismos semelhantes que reflitam o maior ou menor grau de parentesco. Para tal, os pesquisadores procuraram observar **caracteres** que pudessem de alguma forma auxiliar na identificação dos seres mais aparentados, observando peculiaridades ou traços discretos nos organismos como, por exemplo, estruturas,

morfologia, citologia, embriologia, etc. É importante notar dois aspectos principais neste tipo de escola de classificação evolutiva. O primeiro é que as decisões sobre quais caracteres são importantes ou não é feita sem nenhum método objetivo e replicável (aspecto este fundamental em qualquer disciplina científica), e o segundo é que ela não permite a utilização de muitos caracteres simultaneamente.

O fato da escolha da importância dos caracteres ser relativamente arbitrária, dependendo do pesquisador ou dela ser mais evidente e fácil de analisar, tornou-se subseqüentemente uma questão central na biologia, pois eles não necessariamente poderiam refletir as proximidades de parentesco. Todos os caracteres têm a mesma importância? Se não, quais caracteres devem ser utilizados em sistemas de classificação?

Estas questões só começaram a ser resolvidas com o desenvolvimento, do conceito de **homologia**. O curador do Museu de História Natural de Londres, **Richard Owen**, foi um dos grandes responsáveis pelo desenvolvimento desse conceito. Ele argumentava que existiam tantas semelhanças entre os membros anteriores de diferentes animais vertebrados como a nadadeira de uma foca e a mão de um ser humano, que as estruturas deveriam ser derivadas de uma única estrutura presente no ancestral comum de todos aqueles animais. Este conceito entrelaça de maneira elegante e definitiva as observações provenientes da anatomia comparada com o pensamento evolutivo que começava a se desenvolver ao final do século 19. Este grande desenvolvimento intelectual resolve a primeira das questões: quais caracteres devem ser utilizados ao fazer reconstruções evolutivas? Os caracteres homólogos, por refletirem ancestralidade comum e, portanto, herdabilidade de caracteres, são os melhores candidatos para a reconstrução evolutiva.

No entanto, ainda resta um problema de ordem prática. Como determinar quais caracteres são homólogos? Que método deve ser utilizado para que possamos, de maneira objetiva e replicável, diferenciar os caracteres homólogos?

Por muito tempo, não existiam métodos analíticos para determinar a validade dos caracteres homólogos e eram consideradas racionalizações discursivas, ou seja, autoridades no assunto estudavam os diversos caracteres a fim de determinar a partir de sua opinião quais deveriam ser utilizados para reconstrução evolutiva e/ou classificação. No entanto, nas décadas de 1950 e 1960, o entomólogo alemão **Willi Hennig** propôs uma teoria capaz de lidar com todos os caracteres gerados pelos morfologistas de uma vez só, e ainda determinar através de um meio analítico quais seriam homólogos. Sua abordagem, hoje conhecida como **cladística**

ou **sistemática filogenética**, revolucionou o campo da sistemática. Atualmente, a técnica é utilizada não somente para reconstruções históricas de parentesco entre os organismos, mas também como ferramenta preditiva em estudos epidemiológicos (maiores detalhes sobre isso será tratado ao final do tópico).



Agora é com você:
Realizar **atividade online 1**

Na sistemática filogenética, entende-se que um caráter, eventualmente, poderá ser modificado na descendência, passando a se apresentar com variações, que serão subsequentemente, herdadas nas próximas gerações. Desta maneira, o caráter está presente no ancestral exclusivo de todos os herdeiros, e também em todos os herdeiros, mas nestes com uma modificação ou variação. Essa nova variação ou novo estado do caráter é considerado uma condição derivada, ou seja, surgiu a partir da modificação no estado do caráter previamente presente na linhagem ancestral. A condição derivada tem o potencial de servir como determinante para definir um novo grupo e é chamada **apomorfia** (do grego *apó* = longe de; *morphē* = forma) no paradigma moderno. Uma apomorfia pode ser exclusiva de apenas um grupo, sendo chamada nesse caso de **autapomorfia** (do grego *autos* = (eu) mesmo; *morphē* = forma), ou compartilhada por dois ou mais grupos, chamada **sinapomorfia** (do grego *synapsis* = ação de juntar; *morphē* = forma).

Há, no entanto, casos em que o caráter é herdado sem modificação, falando-se em estado **plesiomórfico** (do grego *plesios* = vizinho, significando próximo; *morphé* = forma) e não serve para definir um novo grupo. Quando esse estado plesiomórfico é compartilhado por mais de um agrupamento é chamado **simplesiomórfico**.

As **analogias**, ou seja, caracteres semelhantes que surgem em linhagens não aparentadas, devem ser discriminadas nas análises filogenéticas, pois elas podem gerar interpretações errôneas sobre relações de parentesco entre grupos de organismos. Esses caracteres análogos devem ser identificados para evitar que equivocadamente sejam usados para unir grupos não relacionados. Caracteres análogos são interpretados na filogenética como um tipo de **homoplasia** (do grego *homós* = semelhante, igual; *plásis* = ação de modelar, dar feição): semelhança estrutural decorrente de paralelismo ou convergência evolutiva, e não de ancestralidade comum.

Outro tipo de homoplasia que deve ser identificado nas análises filogenéticas é a perda, condição bastante comum na evolução. Vários caracteres são perdidos na evolução e precisamos saber se a ausência de um caráter na linhagem em estudo se deve a uma perda evolutiva (ou seja, o caráter estava presente no ancestral e foi perdido) ou se realmente aquele caráter nunca esteve presente na linhagem ancestral.

A diferenciação de todos esses tipos de condições é feita na sistemática filogenética com base em uma metodologia própria.

4.3 O método filogenético

Na sistemática filogenética, os caracteres de interesse são os apomórficos: derivados de um estado ancestral.

A questão restante, no entanto, é como diagnosticar quais caracteres são **apomórficos**. É importante que exista um método objetivo e replicável para a determinação destes caracteres, pois estas são necessidades em estudos científicos.

A abordagem de Willi Hennig se diferenciava de outras contemporâneas, pois buscava realizar a inferência histórica de maneira puramente lógica e científica. As principais linhas-guia de seu método são:

1. As relações entre espécies são **estritamente genealógicas**, ou seja, verticais;
2. As **apomorfias** são o único tipo de evidência que identifica; **ancestralidade em comum** e são elas que definem novos agrupamentos;
3. A máxima conformidade com a evidência deve ser determinada utilizando-se o **princípio auxiliar da parcimônia**.

Dentre os três princípios gerais, apenas os dois primeiros são baseados no conhecimento biológico. O terceiro princípio é evocado da disciplina da filosofia lógica para ajudar a resolver questões em que a evidência em mãos aponta várias possibilidades de resposta. O princípio da parcimônia, também conhecido como “a navalha de Occam”, é enunciado da seguinte maneira: “Se em tudo o mais forem idênticas as explicações, a mais simples é a melhor”. Em termos filogenéticos, a explicação mais simples de relação filogenética entre os organismos é aquela que assume o menor número de **passos evolutivos**, ou seja, o menor número de mudanças de estado dos caracteres. Desta maneira, os cientistas se

utilizam de métodos especiais para calcular qual árvore possui o menor número de passos. Estes métodos são em geral executados por programas de computador, mas em alguns casos, pode-se executar até mesmo de maneira manual.



Agora é com você:
Realizar **atividade online 2**

Interpretando uma árvore filogenética

As árvores filogenéticas são representadas seguindo-se uma série de convenções. Observe na **Figura 4.1** uma árvore generalizada.

Os **terminais** (A, B, C, D) representam as entidades de estudo. A entidade de estudo pode ser um indivíduo, populações ou até mesmo espécies inteiras em um único terminal. As linhas que saem dos terminais representam os **ramos**. Os ramos podem conectar um terminal à outro por um **nó** (como D a C pelo nó x), ou um nó à outro mais abaixo (como x a y para os terminais B, C e D).

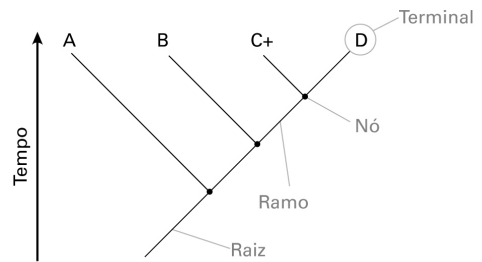


Figura 4.1: Representação de uma árvore filogenética do tipo cladograma, com as convenções indicadas / Fonte: Cepa

Os nós representam ancestrais hipotéticos. No exemplo, o nó x representa o ancestral hipotético de C e D. O nó y representa o ancestral hipotético da linhagem representada por C + D e da linhagem representada por B.

É muito importante frisar que os nós são sempre hipotéticos, e não representam fósseis. Um fóssil, quando encontrado, é um elemento conhecido na análise. Será representado por um terminal na convenção de estudos filogenéticos (note que C possui uma adaga, representando que é um indivíduo ou linhagem fóssil).

O último nó (z) designa a inserção da raiz, que é uma representação hipotética da mais antiga linhagem do grupo.

O termo técnico utilizado para denominar todo tipo de árvore filogenética é **dendrograma**. Um elemento, nem sempre representado, porém sempre implícito, é a direção do tempo. Em todo tipo de representação em forma de árvore filogenética, **a ordem de ramificação** representa o momento de isolamento das linhagens, umas em relação às outras.

Na raiz está representado o evento evolutivo mais antigo, e os terminais representam eventos recentes.

Em dendogramas do tipo **cladograma**, os comprimentos relativos dos ramos não representam unidades de tempo (veja as representações A, B e C na **Figura 4.2**).

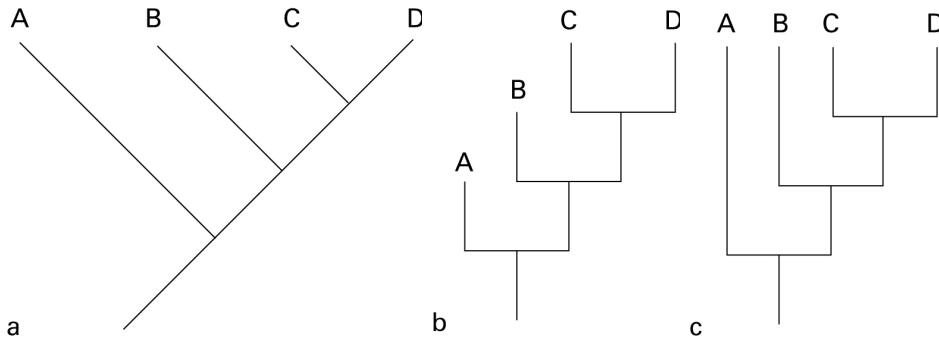


Figura 4.2: Representações de cladogramas

As representações A, B e C nos dizem a mesma coisa em questão de relação entre os terminais. No caso da representação B, no entanto, os ramos estão mais quadrados (uma modificação puramente estética do que está em A), e na C, a única diferença é que os ramos foram alongados para alinhar os terminais no topo (novamente, uma modificação estética de A e B), o que pode ser feito já que os comprimentos relativos não possuem significado temporal.

Analise agora outro tipo de representação:

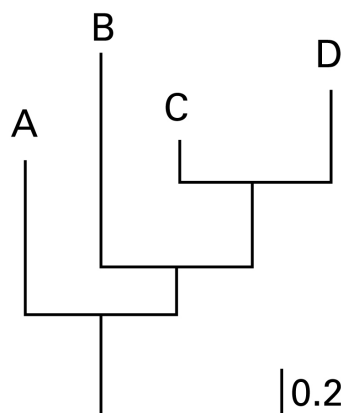


Figura 4.3: Representação de um filograma

Na representação acima, há uma escala de tempo indicada por um pequeno traço ao lado da imagem. No caso, o comprimento do traço equivale a 0,2 milhão de anos. Sendo assim, cada ramo tem um comprimento diferente, pois leva em consideração o tempo, ou seja, cada ramo representa a quantidade de modificação evolutiva que ocorreu desde que as linhagens se separaram no nó. Este tipo de árvore filogenética é conhecido como **filograma**. Outra maneira de reconhecer este tipo de árvore é que, em geral, diferentes ramos têm diferentes tamanhos, enquanto o cladograma é sempre mais simétrico.

Todas as árvores filogenéticas podem ser de dois tipos: enraizadas ou não-enraizadas. As mostradas anteriormente são todas enraizadas, pois possuem uma raiz, uma origem.

Árvores não enraizadas são muito comuns em análises baseadas somente em dados moleculares, mas são usadas em outros casos também. Veja um exemplo desse tipo de árvore e note que nela não há uma raiz: todos os ramos partem de um ponto central.

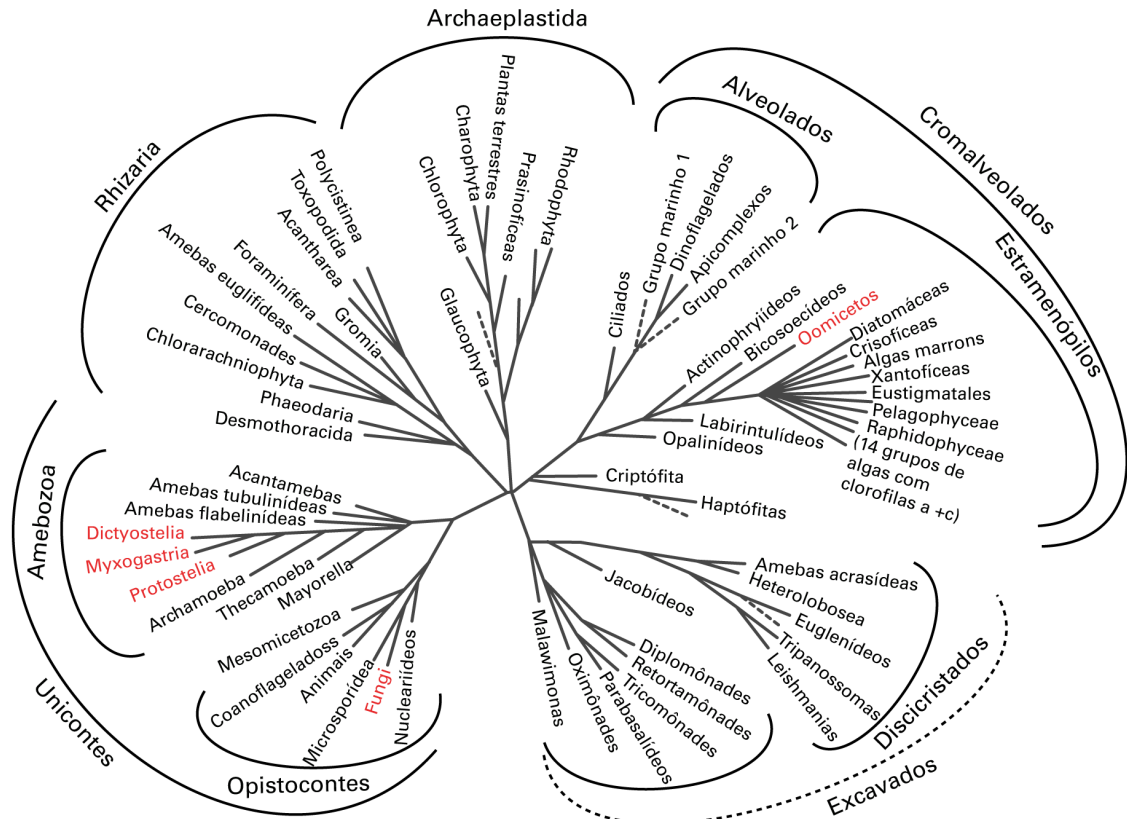


Figura 4.4: Exemplo de árvore não enraizada

Na construção de árvores enraizadas, deve-se definir o posicionamento da raiz com base na determinação de pelo menos um **grupo externo**, ou seja, de pelo menos uma linhagem que não faça parte das linhagens de interesse, chamadas **grupos internos**. A escolha do grupo externo não é simples. Em geral é um grupo aparentado dos grupos internos de interesse, mas que surgiu antes na história evolutiva. Assim, conseguimos definir homologias e comparar os estados de cada caráter para definir as apomorfias.



Agora é com você:
Realizar **atividade online 3**

As apomorfias, principalmente em estudos morfológicos, podem estar indicadas em uma árvore. Veja no exemplo abaixo que o grupo C+D é unificado pela apomorfia 1, o grupo B+C+D é unificado pela apomorfia 2 e assim por diante.

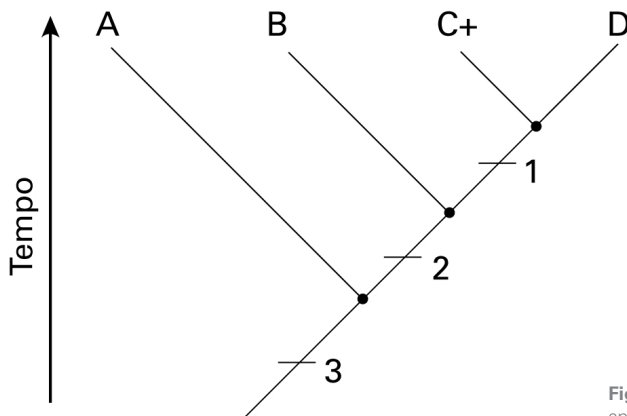


Figura 4.5: Cladograma com as apomorfias indicadas

Os grupos unificados por apomorfias são denominados **monofiléticos** (do grego *monos* = único; *filético* = refere-se à linhagem), pois contêm em uma única linhagem um ancestral e inclui todos os seus descendentes. Desta maneira, C+D compõem um grupo monofilético definido pela apomorfia 1. Do mesmo modo, B + o ramos que inclui (C+D) é monofilético, definindo pela apomorfia 2.

No entanto, agrupamentos que excluem uma das linhagens descendentes, são chamados **parafiléticos** (do grego *para* = ao lado de; *filético* = refere-se à linhagem). Por exemplo, se considerarmos apenas as linhagens B+C, elas são parafiléticas, pois excluímos a linhagem D. Se o grupo excluir mais

de uma das linhagens descendentes, como por exemplo, o grupamento A+C exclui tanto B quanto D, é denominado **polifilético** (do grego *polýs* = muito; *filético* = refere-se à linhagem).

No paradigma estabelecido por Willi Hennig, somente os grupos **monofiléticos** devem ser considerados naturais.

Grupos que compartilham o mesmo nó exclusivo são chamados grupos-irmãos. Assim, C e D são grupos irmãos, pois compartilham exclusivamente um nó. Por sua vez, B é grupo irmão do conjunto formado pelos grupos (B+C), pois também compartilham um nó exclusivo.



Agora é com você:
Realizar **atividade online 4**

4.4 Tipos de dados utilizados em inferência filogenética

Tradicionalmente, dados provenientes de estudos morfológicos e anatômicos são utilizados para a reconstrução filogenética. Tipicamente, cientistas observam uma grande quantidade de espécimes representando diferentes linhagens (linhagens podem ser espécies ou populações). Em cada espécime, são observados vários caracteres morfológicos e anatômicos, e estes são registrados em uma grande tabela chamada **matriz de caracteres**. A matriz de caracteres é então analisada de modo a tentar definir os estados derivados de cada caráter (apomorfias) e transpor isso com parcimônia para um cladograma. Esta análise é feita por um **algoritmo** que tenta reconciliar as mudanças de estado observadas na matriz com uma árvore filogenética que explique tais mudanças da maneira mais simples possível.

Um **algoritmo** é uma série de instruções dada à um computador ou a um ser humano de maneira que permita resolver problemas de um certo tipo. Pode-se fazer uma analogia com uma receita de bolo: o problema é fazer o bolo, o meio de resolvê-lo é seguindo a receita. Algoritmos são, então, protocolos seguidos passo-a-passo, em geral por um computador, para completar o problema em questão, no caso, encontrar a árvore filogenética mais simples.

Atualmente, os cientistas analisam também caracteres moleculares, ou seja, sequências de genes.

Os dados moleculares possuem um poder comparativo muito abrangente, pois todos os organismos vivos compartilham da mesma organização do material genético. Neste caso, cada sítio observado em uma dada sequência é considerado um caráter.

Sequências genéticas (nucleotídeos e aminoácidos) de diferentes organismos devem primeiramente ser alinhadas utilizando programas (*software*) específicos. Após o alinhamento, e tratando cada sítio como um caráter homólogo, utilizam-se algoritmos de reconstrução filogenética de maneira bastante semelhante aos utilizados em reconstruções filogenéticas baseadas em dados morfológicos.

Com o advento de métodos de sequenciamento cada vez mais baratos, hoje em dia existe enorme quantidade de organismos com seus genomas completamente sequenciados (**Tabela 4.1**). Os filogeneticistas atualmente conseguem alinhar genomas inteiros em busca de sítios homólogos, em um novo tipo de estudo chamado **filogenômico**, ou seja, combinando técnicas da filogenética com a área da genômica. Estudos filogenômicos analisam milhões de sítios, para centenas de espécies e levam meses para ser completados, mesmo utilizando os mais avançados computadores.

Outro tipo muito comum de árvore filogenética que se desenvolve hoje em dia são as baseadas em evidência total, ou seja, combinando dados moleculares e morfológicos em uma mesma reconstrução. Este tipo de reconstrução é o mais interessante do ponto de vista teórico, no entanto, apresenta grande desafio computacional.

| Grupo | Completos | Em andamento |
|--------------|-----------|--------------|
| Eucariontes | 37 | 1170 |
| Archaea | 119 | 90 |
| Bacteria | 1673 | 5140 |
| | | |
| Total | 1829 | 6400 |

Tabela 4.1: Síntese dos projetos de sequenciamento de genoma, organizados por domínio da vida /
Fonte: NCBI (Centro Nacional de Informação Tecnológica, do inglês *National Center for Biotechnology Information*).

Reconstruções filogenéticas em si representam um dos problemas mais difíceis de serem resolvidos por computadores, pois o número de árvores possíveis é literalmente astronômico. Verifique na **Tabela 4.2** como o número de árvores possíveis aumenta rapidamente conforme aumenta o número de terminais. Para se ter uma idéia comparativa destes números, os astrônomos calculam que o número total de corpos estelares (planetas, estrelas e asteróides) em todo o universo é por

volta 10^{24} objetos. Note na tabela que com cerca de 30 terminais já existe um número maior de árvores possíveis. De maneira mais dramática ainda, os astrônomos calculam que existem cerca de 10^{80} átomos em todo o Universo. Com somente 50 terminais já existem mais possibilidades de árvores enraizadas do que existem átomos no Universo. Hoje em dia, a maior parte das árvores construídas tem centenas de terminais, e, algumas, até mesmo milhares.

| Terminais | Árvores Possíveis | |
|-----------|------------------------|------------------------|
| | Não-enraizadas | Enraizadas |
| 3 | 1 | 3 |
| 4 | 3 | 15 |
| 5 | 15 | 105 |
| 6 | 105 | 945 |
| 7 | 945 | 10.395 |
| 8 | 10.395 | 135.135 |
| 9 | 135.135 | 2.027.025 |
| 10 | 2.027.025 | 34.495.425 |
| 15 | $7,91 \times 10^{12}$ | $2,13 \times 10^{14}$ |
| 20 | $2,22 \times 10^{20}$ | $8,20 \times 10^{21}$ |
| 30 | $8,69 \times 10^{36}$ | $4,95 \times 10^{38}$ |
| 50 | $2,84 \times 10^{74}$ | $2,75 \times 10^{76}$ |
| 80 | $2,18 \times 10^{137}$ | $3,43 \times 10^{139}$ |

Pessoas vivendo em Londres 1810 (10^6)

Células no corpo humano (10^{11})

Corpos celestes no Universo (10^{24})

Átomos no Universo (10^{80})

Tabela 4.2: Quadro comparativo entre a quantidade possível de árvores diferentes (assumindo sempre árvores bifurcantes completamente resolvidas) ao aumentar o número de terminais. À direita, alguns exemplos comparativos.

Os cientistas não esperam, no entanto, que os computadores construam cada uma das árvores possíveis, para depois definir qual está correta. Na verdade, isso seria impossível – no mínimo, levaria uma eternidade mesmo com os mais avançados dos computadores. O que se tem feito para resolver esse problema é aplicar métodos computacionais especiais que exploram o espaço virtual de possibilidades de maneira mais eficiente, de modo a se chegar a uma melhor solução aproximada. Esses programas, portanto, ao invés de construir todas as árvores possíveis para posteriormente decidir qual delas é a melhor, procuram encontrar padrões lógicos de maneira **heurística**. É na busca de desenvolver estas estratégias de exploração que se concentra a maior parte da pesquisa computacional no momento.

Buscas **heurísticas** baseiam-se em aprendizado para encontrar uma solução. Elas se contrapõem às buscas exaustivas, ou seja, que exploram todo o universo possível para encontrar uma solução. Um bom exemplo de busca heurística é o que chamamos de senso comum: ao se deparar com uma situação nova, tendemos a nos apoiar em experiências prévias para decidir qual a melhor reação. Como não temos como prever o futuro (o que seria equivalente à uma busca exaustiva), tentamos imaginar que outras situações semelhantes passamos para tomar uma decisão. Computadores utilizam uma estratégia semelhante ao procurar árvores filogenéticas, quando se deparam com uma situação que só os levou a construir árvores menos parcimoniosas. Eles podem mudar de direção e explorar outra estratégia mais interessante.

4.5 Utilização da metodologia e principais limitações

Os métodos desenvolvidos pela sistemática filogenética são muito poderosos, e hoje são utilizados em diversas áreas do conhecimento, tanto áreas de ciência básica como de ciência aplicada. As disciplinas da biologia molecular, genética, evolução, desenvolvimento, comportamento, epidemiologia, ecologia, biologia da conservação e ciências forense são todas conceitualmente unificadas pela aplicação da metodologia da sistemática filogenética. Por exemplo, alguns cientistas utilizam métodos baseados em sistemática filogenética para entender melhor como a biodiversidade se desenvolve, aprimorando estimativas de números de espécies em certas localidades e permitindo melhor conhecimento de causa ao criar legislação visando conservação de espécies. Outro uso muito comum é na área de epidemiologia, para criação de vacinas mais eficientes. Anualmente, a Organização Mundial da Saúde reformula a vacina contra a gripe, pois o vírus tem alta taxa de mutação. Para determinar quais linhagens devem ser utilizadas na fabricação de novas vacinas, são feitas análises filogenéticas que auxiliam no processo de determinação de quais linhagens são mais comuns no ano em particular e, portanto, devem ser mais representadas na mistura viral utilizada na vacina.

Ainda assim a sistemática filogenética possui algumas limitações.

Como já discutimos, a metodologia cladística assume que a herança dos caracteres (e genes) ocorre de maneira vertical. Hoje em dia sabemos que isto não é necessariamente verdade para uma série de genes, especialmente em bactérias, que realizam muita transferência lateral

de genes. Eventos de hibridização (como por exemplo, a endossimbiose que deu origem às mitocôndrias) também não podem ser computados utilizando-se métodos cladísticos. Estas limitações tendem a ser incorporadas eventualmente, porém por enquanto, devem ser levadas em conta ao realizar estudos filogenéticos.

Fechamento o assunto

A sistemática filogenética é uma disciplina que se desenvolveu da curiosidade humana e, com o tempo, se tornou uma metodologia unificadora em toda a biologia. Atualmente, alguns dos programas de computador mais potentes são desenvolvidos para resolver problemas filogenéticos, e a metodologia já começa a ser usada em outras disciplinas. A interpretação de árvores filogenéticas é um conceito básico para compreensão de teorias biológicas contemporâneas, assim como um entendimento qualitativo sobre os métodos utilizados para a reconstrução histórica. Apesar dos grandes avanços, a sistemática filogenética (assim como toda metodologia científica) possui limitações, em especial em relação à transferência lateral de genes, e deverá ser modificada nos próximos anos para incorporar novas descobertas biológicas. A síntese evolutiva moderna, que divide a vida em três grandes domínios, não seria possível sem o desenvolvimento da sistemática filogenética em conjunto com a biologia molecular.

Referências Bibliográficas

- FELSENSTEIN, J. (2003) **Inferring phylogenies**.
- GRAUR, D. and LI W. H. (2000) **Fundamentals of Molecular Evolution**.
- HALL, B. (2007) **Phylogenetic trees made easy: a how-to manual**.
- HENNIG, W, DWIGHT-DAVIS D, ZANGERL R. (1999) **Phylogenetic systematics**.
- WILEY, E. O., LIEBERMAN B. S. (2011) **Phylogenetics: Theory and Practice of phylogenetic systematics**.