Bases e Métodos de Datação Molecular

Definição de Biogeografia

"Biogeografia é o estudo das distribuições geográficas no **tempo** e no **espaço**"

(Brown & Lomolino, 1998)

Estimar de tempos de divergência

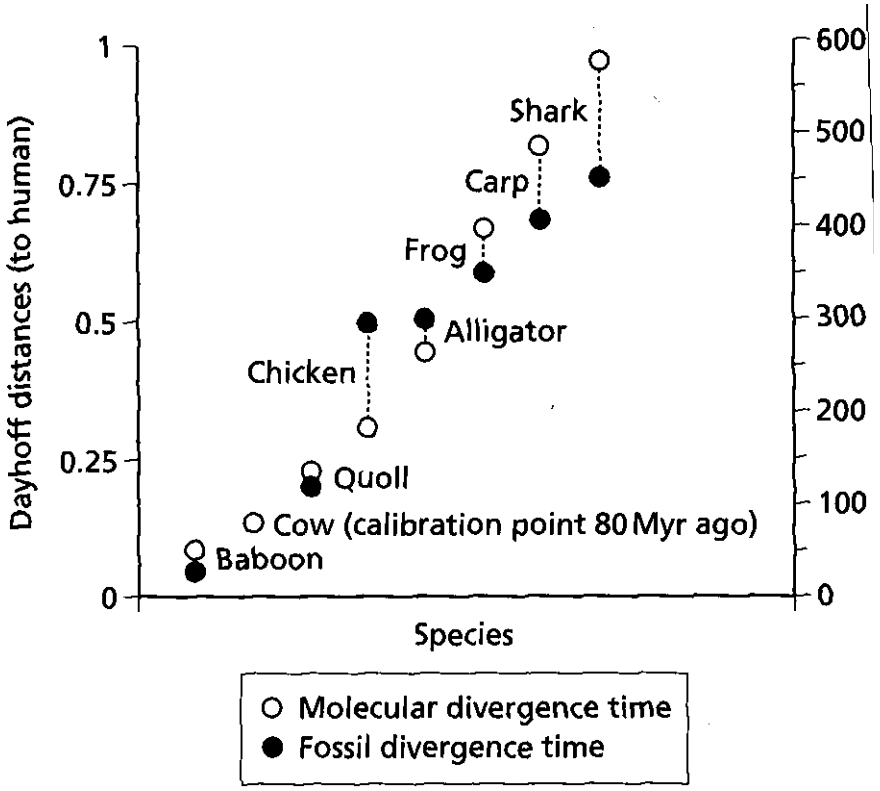
Dimensão temporal do processo evolutivo

Relógio Molecular

(Zuckerland & Pauling, 1962, 1965)

- Taxa de evolução de sequências moleculares (substituição de aminoácidos, nucleotídeos) é estocasticamente constante ao longo do tempo em diferentes linhagens
- Número de substituições de DNA entre duas linhagens é proporcional ao tempo de divergência entre elas
- O grau de divergência das sequências (o qual podemos estimar) é então proporcional ao tempo e pode ser usado para estimar a divergência das linhagens

Constância na taxa de substituição do DNA



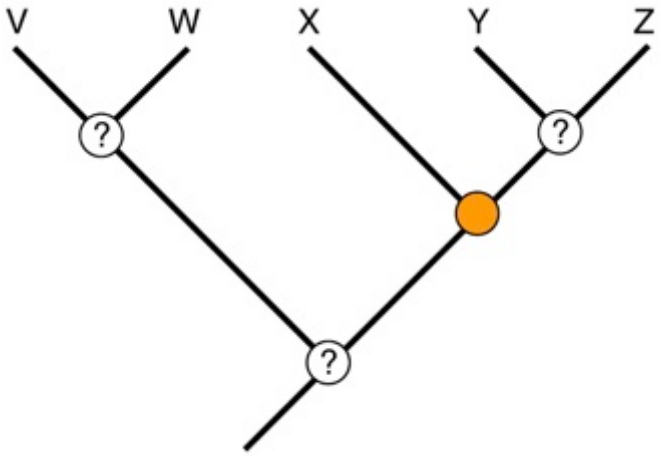
Como funciona o relógio molecular?

Assumindo:

(1) o relógio molecular,

(2) uma árvore com ramos proporcionais à taxa de mutação;

(3) pelo menos um nó com idade conhecida, pode-se estimar a idade dos outros nós.



Neste contexto, o relógio molecular será perfeito se:

(1) Taxas de substituição forem realmente constantes;

(2) A árvore e o comprimento dos seus ramos estiverem corretos;

(3) A idade da calibração estiver correta.

No entanto, nós sabemos que:

(1) As taxas de substituição são heterogêneas;

(2) Existem erros associados à reconstrução das filogenias e respectivos comprimentos de ramos;

(3) Existem vários erros associados às idades estimadas pelos pontos de calibração (fósseis ou eventos geológicos).

(1) Heterogeneidade nas taxas de substituição de DNA

- Taxas de evolução molecular variam de acordo com:

(a) Genoma (mtDNA tem taxa de mutação 10x maior que o nDNA)

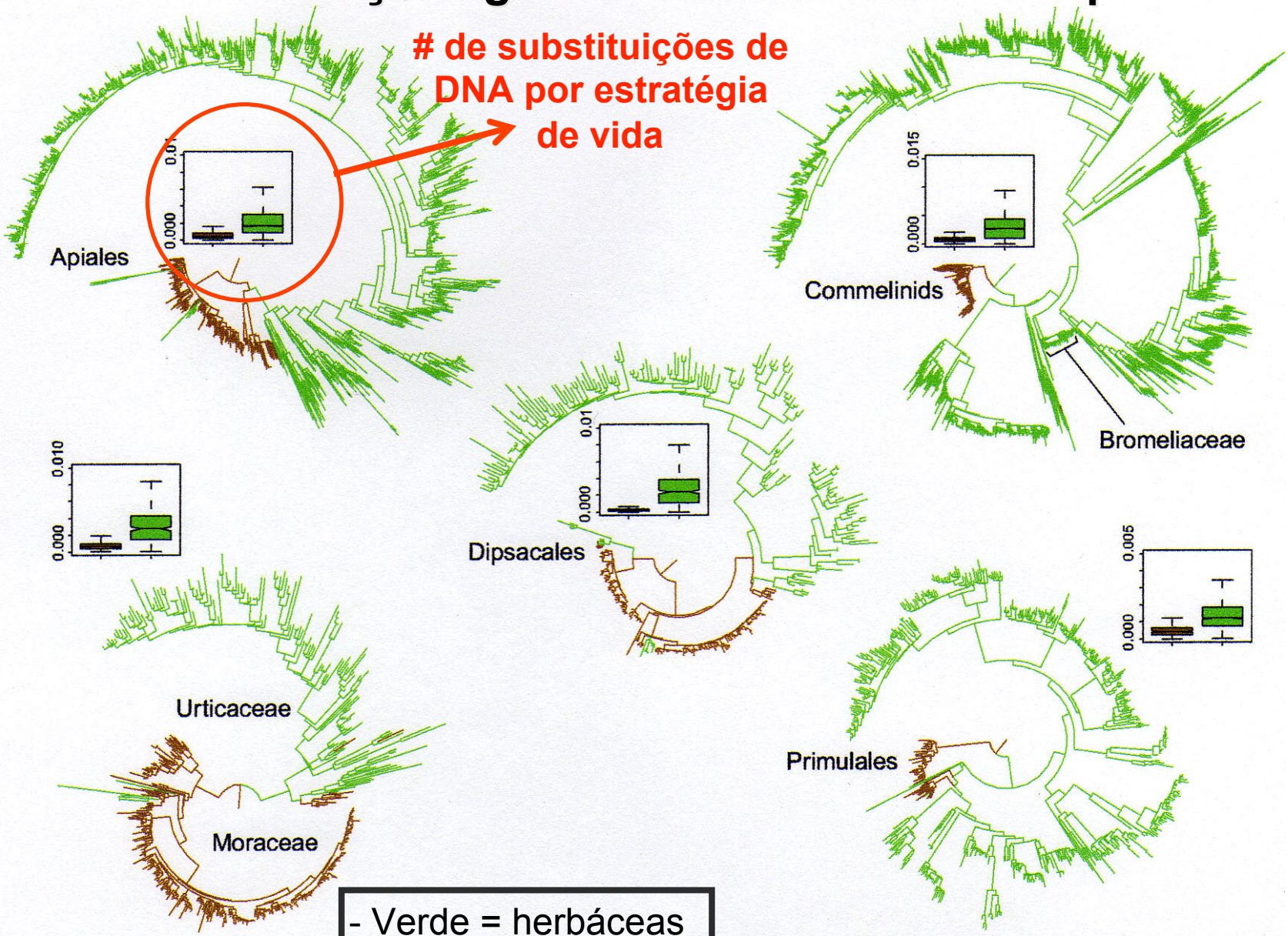
(b) Região do genoma

(c) Posição nucleotídica (códon)

(d) Tempo

(e) Espécie (e.g., história de vida)

Taxas de evolução ligada à história de vida em plantas



(Smith & Donoghue, 2008)

No entanto, nós sabemos que:

(1) As taxas de substituição são heterogêneas;

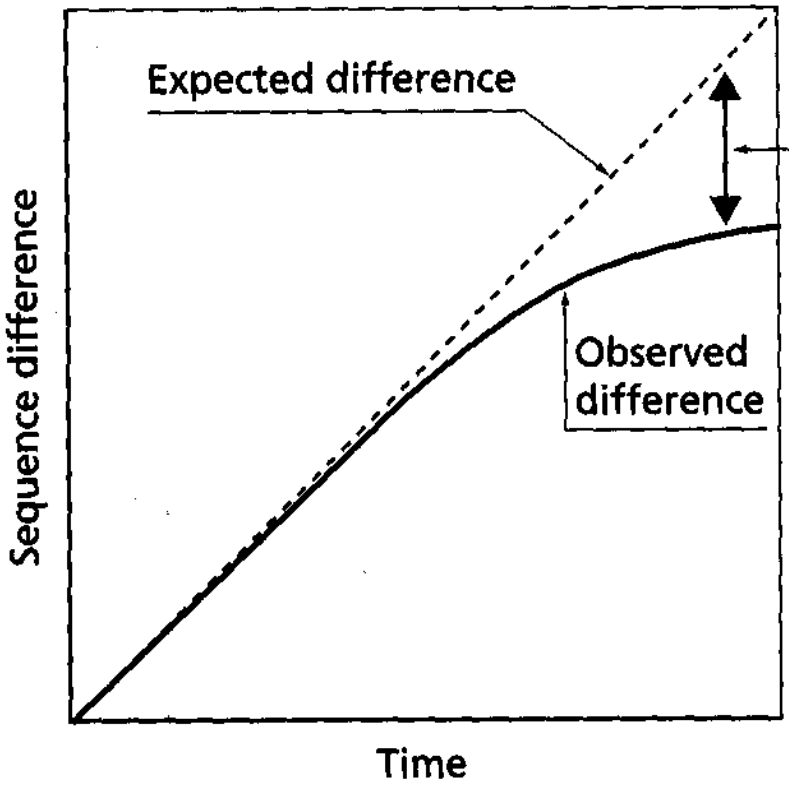
(2) Existem erros associados à reconstrução das filogenias e respectivos comprimentos de ramos;

(3) Existem vários erros associados às idades estimadas pelos pontos de calibração (fósseis ou eventos geológicos).

(2) Erros na reconstrução da filogenia

- Tanto **erros** na topologia como erros na estimativa dos comprimentos dos ramos **são comuns** nas reconstruções filogenéticas.

- Erros na reconstrução das filogenias e taxas de mutação estimadas **afetam** diretamente as **idades** de divergência **estimadas**.



(Page & Holmes, 1998)

No entanto, nós sabemos que:

(1) As taxas de substituição são heterogêneas;

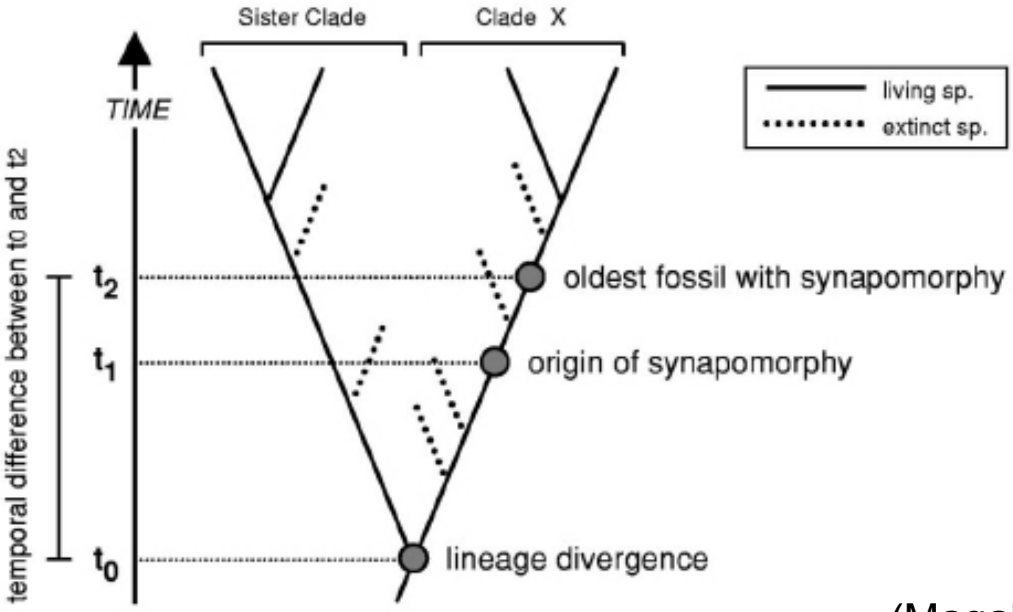
(2) Existem erros associados à reconstrução das filogenias e respectivos comprimentos de ramos;

(3) Existem vários erros associados às idades estimadas pelos pontos de calibração (fósseis ou eventos geológicos).

(3) Erros associados aos pontos de calibração

- Um ponto de calibração oferece um ponto de referência para transformar os ramos da árvore em ramos proporcionais ao tempo

- Há várias formar de calibrar uma árvore:
  - (a) Fóssil
  - (b) Idade proveniente de outro estudo
  - (c) Evento geológico, idade de formação de uma ilha onde ocorre um taxon endêmico



(Magallón, 2004)

Considerando estes problemas….

(1) As taxas de substituição são heterogêneas;

(2) Existem erros associados à reconstrução das filogenias e respectivos comprimentos de ramos;

(3) Existem vários erros associados às idades estimadas pelos pontos de calibração (fósseis ou eventos geológicos).

Por que tentar estimar a idade de divergência entre linhagens?

- Dados são interpretados à luz de outras evidências geológicas em escalas temporais amplas, permitindo o teste de hipóteses biogeográficas em escala de milhões de anos.

- Análises realizadas com cautela, com erros adequadamente estimados, nos permitem adicionar um componente temporal importante aos estudos de biogeografia.

- Escalas temporais nos permitem realizar análises comparativas entre organismos distantemente aparentados.

Como estimar a idade de divergência entre linhagens?

(1) Métodos que assumem o relógio molecular

(a)Regressão Linear (Nei, 1987)

(b)“Mean Path Length Method” (Bremer & Gustafson, 1997)

(c)“ML clock optimization” (Langley & Fitch, 1974)

(2) Métodos que corrigem a heterogeneidade nas taxas de substituição do DNA

(3) Métodos que incorporam heterogeneidade nas taxas de substituição de DNA

Como estimar a idade de divergência entre linhagens?

(1) Métodos que assumem o relógio molecular

(a)Regressão Linear (Nei, 1987)

(b)“Mean Path Length Method” (Bremer & Gustafson, 1997)

(c)“ML clock optimization” (Langley & Fitch, 1974)

(2) Métodos que corrigem a heterogeneidade nas taxas de substituição do DNA

(3) Métodos que incorporam heterogeneidade nas taxas de substituição de DNA

Métodos que incorporam heterogeneidade nas taxas de substituição de DNA

(1) “Node-age Calibrations”

(a) NPRS & Verossimilhança Penalizada (r8s; Sanderson, 1997, 2002)

(b) Multidivtime (Thorne & Kishino, 2002; Shiguno, 2005)

(c) BEAST (Drummond & Rambout, 2005)

(d) Use of multiple calibration points (Marshall, 2008)

(2) “Tip-dating” (Pyron, 2011; Ronquist et al., 2012)

(a) “Total-Evidence Approach” (Ronquist et al. 2012)

(b) “Fossilized Birth-Death Model” (Heath et al. 2014)

Métodos que incorporam heterogeneidade nas taxas de substituição de DNA

(1) “Node-age Calibrations”

- (a) NPRS & Verossimilhança Penalizada (r8s; Sanderson, 1997, 2002)
- (b) Multidivtime (Thorne & Kishino, 2002; Shiguno, 2005)
- (c) BEAST (Drummond & Rambout, 2005)
- (d) Use of multiple calibration points (Marshall, 2008)

(2) “Tip-dating” (Pyron, 2011; Ronquist et al., 2012)

- (a) “Total-Evidence Approach” (Ronquist et al. 2012)
- (b) “Fossilized Birth-Death Model” (Heath et al. 2014)

“Node-Age Calibrations”

(1) Obter uma árvore filogenética com comprimento dos ramos;

(2) Checar se a árvore apresenta taxas de evolução constantes;

(3) Estabelecer os pontos de calibração com base nas evidências disponíveis;

(4) Transformar ramos da árvore de forma que sejam proporcionais ao tempo utilizando um método adequado.

(1) Obter uma árvore

- É importante obter a **melhor** topologia e estimativa de comprimento de ramos **possível**;

- Árvores reconstruídas com parcimônia são insuficientes pois não incluem estimativas acuradas dos comprimentos dos ramos (i.e., comprimento de ramos representam apenas as mudanças observadas);

- Árvores reconstruídas com metodologias Bayesianas ou Verossimilhança são preferidas pois utilizam modelos de evolução que incorporam também as mudanças esperadas (“multiple-hits”)

“Node-Age Calibrations”

(1) Obter uma árvore filogenética com comprimento dos ramos;

(2) Checar se a árvore apresenta taxas de evolução constantes;

(3) Estabelecer os pontos de calibração com base nas evidências disponíveis;

(4) Transformar ramos da árvore de forma que sejam proporcionais ao tempo utilizando um método adequado.

(2) Testar se a taxa de evolução é constante

- Existem vários métodos para testar se uma linhagem evolui de acordo com o relógio molecular:

(a) **Relative Rate Test**: Compara as taxas de substituição do DNA usando três espécies por vez;

(b) **Tajima Test**: Examina se o número de sítios compartilhados entre o grupo externo e cada um dos grupos internos é igual;

(c) **Branch Length Test**: Testa se o caminho entre a raíz e cada terminal é maior do que o caminho médio;

(d) **Likelihood Ratio Test**: Utiliza a matriz molecular completa para estimar o desvio do relógio molecular como um todo.

Likelihood Ratio Test é o método mais robusto

- Idéia básica:

(a) Estima a filogenia assumindo o relógio molecular (i.e., distâncias "root-to-tip" precisam ser iguais em todos os caminhos);

(b) Estima a filogenia sem assumir o relógio molecular, utilizando o modelo evolutivo mais adequado para aquele conjunto de dados;

(c) **Compara** as duas topologias e testa se existe uma diferença significativa no valor de verossimilhança obtido utilizando um teste de qui-quadrado;

(d) É um teste conservador (i.e., no caso de dados limitados pode não rejeitar o relógio molecular mesmo quando a taxa de divergência é desigual).

“Node-Age Calibrations”

(1) Obter uma árvore filogenética com comprimento dos ramos;

(2) Checar se a árvore apresenta taxas de evolução constantes;

(3) Estabelecer os pontos de calibração com base nas evidências disponíveis;

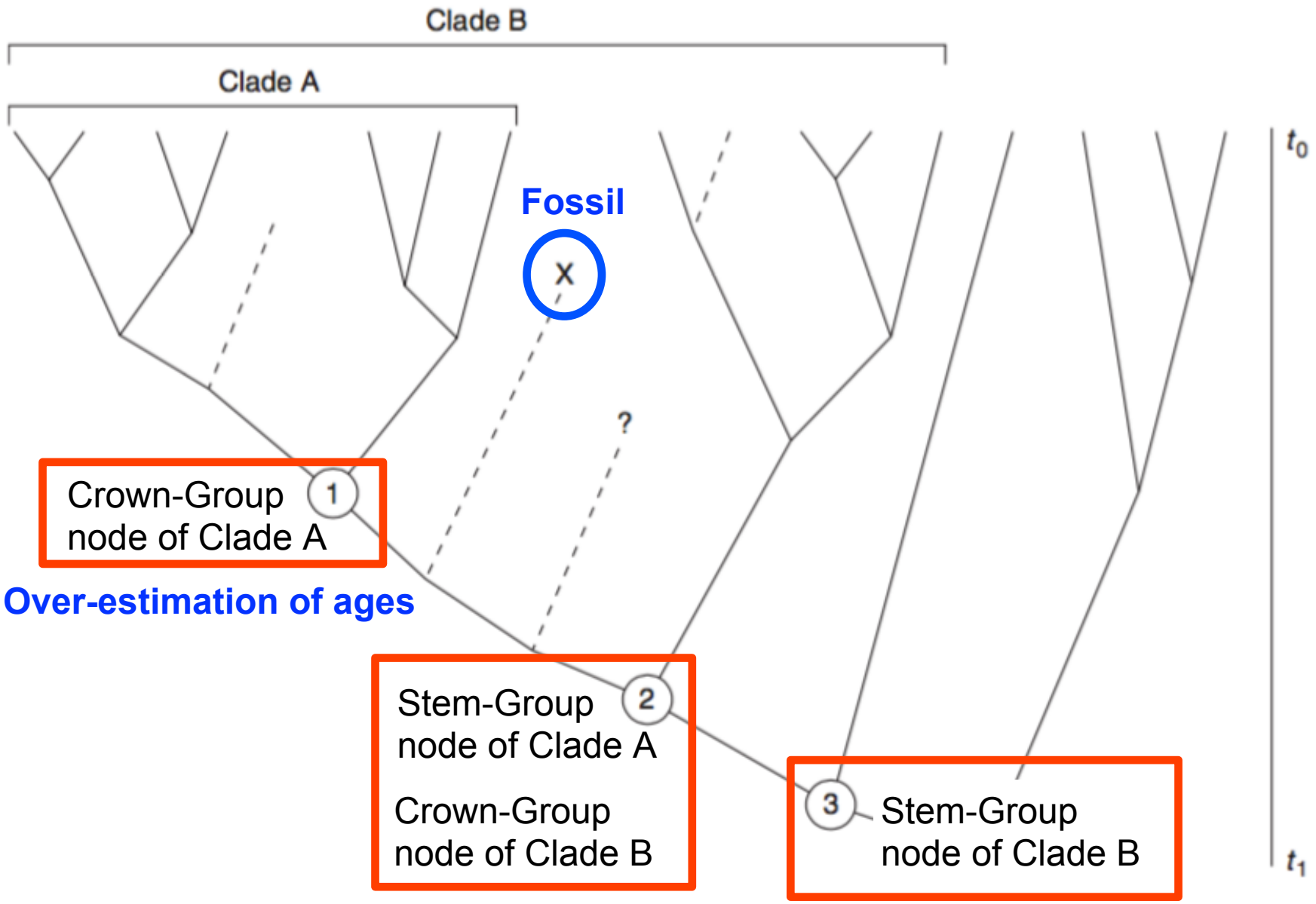
(4) Transformar ramos da árvore de forma que sejam proporcionais ao tempo utilizando um método adequado.

(3) Estabelecer os pontos de calibração

- A **primeira aparição** de um táxon no registro fóssil geralmente indica o momento no qual aquela **espécie** se tornou **abundante** e não quando ela apareceu pela primeira vez.

- Estimativas com base no fóssil tendem a **subestimar a idade** real da linhagem. Isto é, datas mais recentes tendem a ser designadas ao clado onde o fóssil é posicionado (i.e., estimativas mínimas são obtidas).

- Outra fonte de **incerteza** está associada à estimativa da **idade do fóssil** (estimada através de correlações estratigráficas ou datação radiométrica).

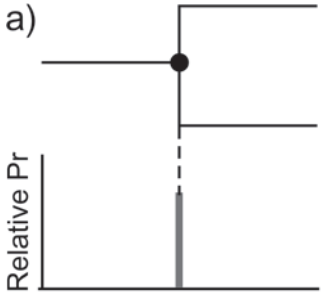


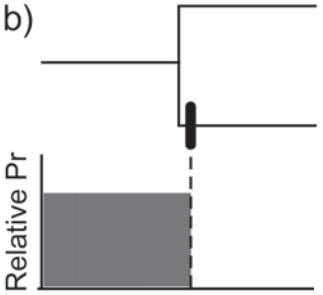
Minimum Constraint
(Takes into account incompleteness of fossil record)

(Forest, 2009)

Métodos para calibração

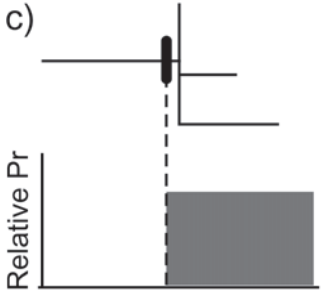
Point-calibration

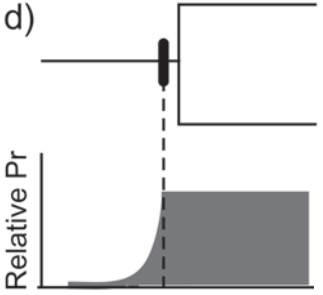




Hard Minimum Bound

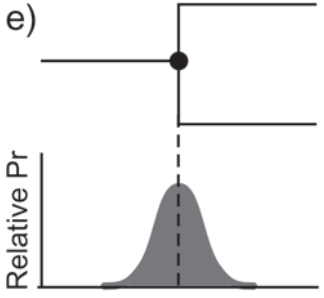
Hard Maximum Bound

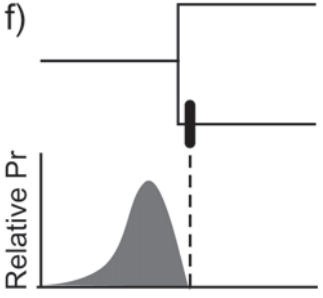




Soft Maximum Bound

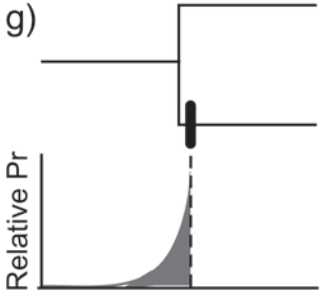
Normal Distribution





Lognormal Distribution

Exponential Distribution



(Ho & Phillips 2009)

“Node-Age Calibrations”

(1) Obter uma árvore filogenética com comprimento dos ramos;

(2) Checar se a árvore apresenta taxas de evolução constantes;

(3) Estabelecer os pontos de calibração com base nas evidências disponíveis;

(4) Transformar ramos da árvore de forma que sejam proporcionais ao tempo utilizando um método adequado.

(4) Transformar os ramos da árvore

- Se os dados passarem o teste do relógio molecular é possível transformar o comprimento de ramos em ramos proporcionais ao tempo diretamente.

- Se a hipótese do relógio molecular for rejeitada (maioria dos casos), é possível implementar um relógio molecular relaxado. O relógio molecular relaxado busca modelar a variação estocástica existente nas taxas de mutação.

Métodos que incorporam heterogeneidade nas taxas de substituição de DNA

(1) “Node-age Calibrations”

(a) NPRS & Verossimilhança Penalizada (r8s; Sanderson, 1997, 2002)

(b) Multidivtime (Thorne & Kishino, 2002; Shiguno, 2005)

(c) BEAST (Drummond & Rambout, 2005)

(d) Use of multiple calibration points (Marshall, 2008)

(2) “Tip-dating” (Pyron, 2011; Ronquist et al., 2012)

(a) “Total-Evidence Approach” (Ronquist et al. 2012)

(b) “Fossilized Birth-Death Model” (Heath et al. 2014)

Abordagem com Máxima Verossimilhança

“Non-Parametric Rate Smoothing” (NPRS; Sanderson, 1997):

- Estima a heterogeneidade nas taxas de substituição de DNA e cria um parâmetro que **suaviza a heterogeneidade** ao longo da árvore.

- Este parâmetro estabelece limites (**penalidades**) para a variação das **taxas** de substituição de DNA em linhagens **diferentes**.

- A penalidade minimiza as diferenças nas taxas de substituição entre ramos próximos (e.g., a penalidade para a variação na taxa de substituição entre espécies irmãs é maior do que a penalidade para espécies pouco aparentadas).

Abordagem Bayesiana

- Utiliza um **modelo estocástico** para descrever a taxa de mudanças evolutivas.

- Considera informações “a priori” e MCMC (Markov-Chain Monte-Carlo) para determinar a **distribuição das taxas de evolução**.

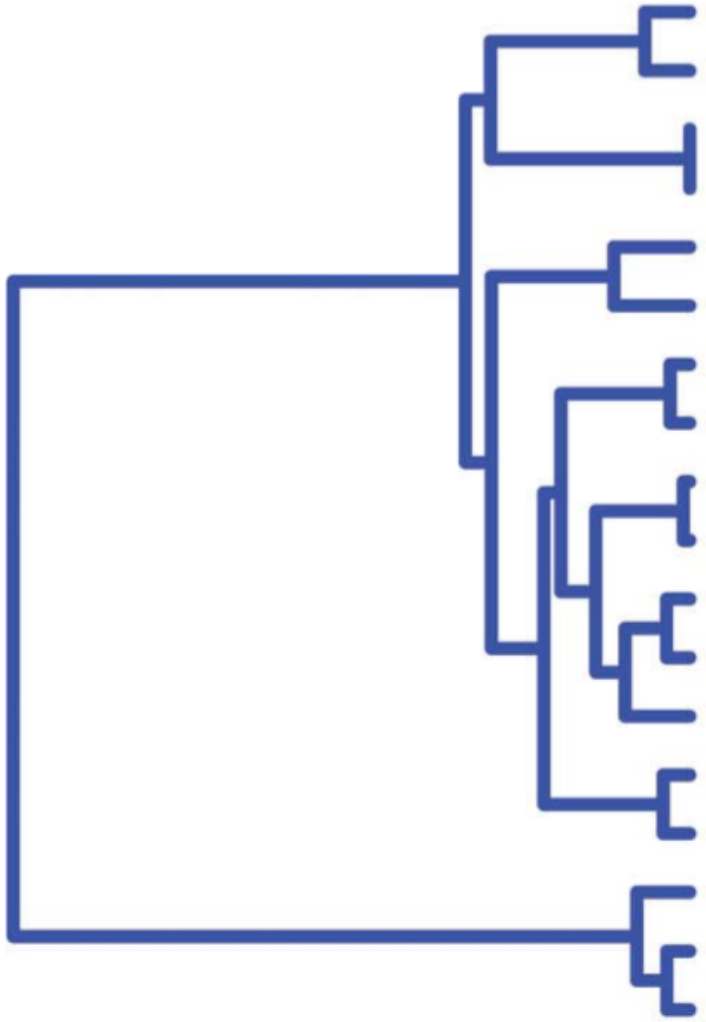
- Estima as taxas de substituição, parâmetros de substituição, e idades de divergência entre linhagens de forma simultânea.

- Representa uma forma flexível para modelar a variação na taxa de substituição e obter estimativas confiáveis para os eventos de especiação dado um modelo com pressupostos adequados.

- Ao invés de buscar uma árvore ótima, busca o **conjunto de árvores** com maior probabilidade.

Exemplos de modelos evolutivos

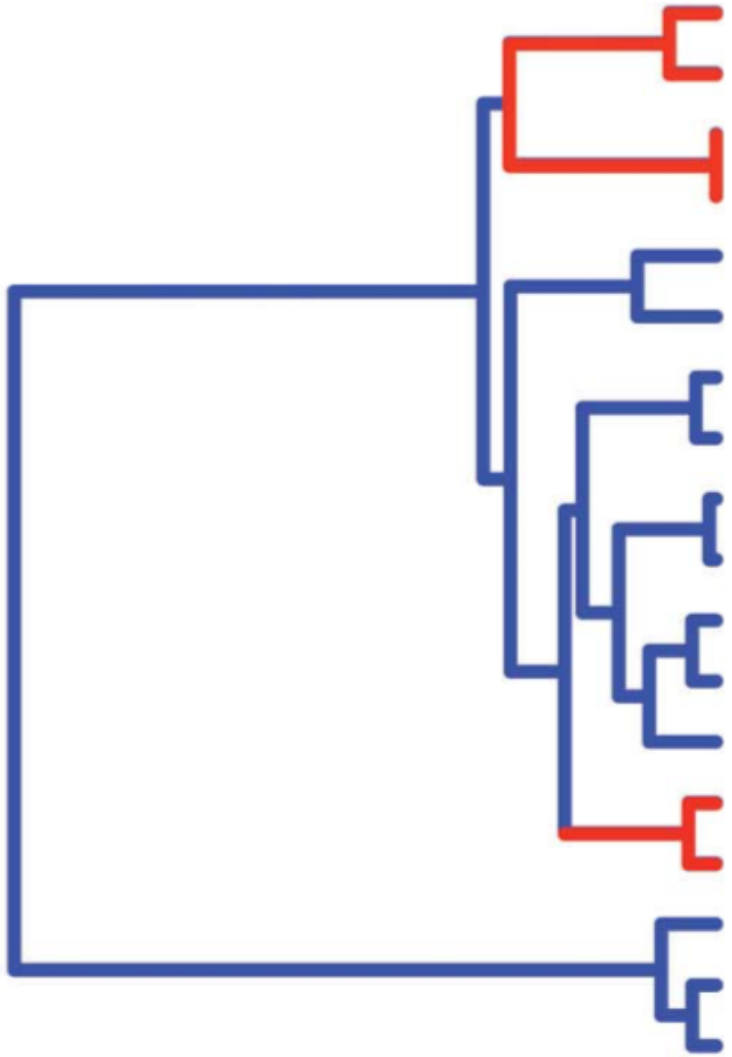
(1) **Relógio Molecular Global** (Zuckerland & Pauling, 1962): Utiliza uma regressão linear para **estimar taxas constantes** de substituição ao longo do tempo para todas as linhagens (e.g., “Hawaiian Silversword” Asteraceae).



Global Clock

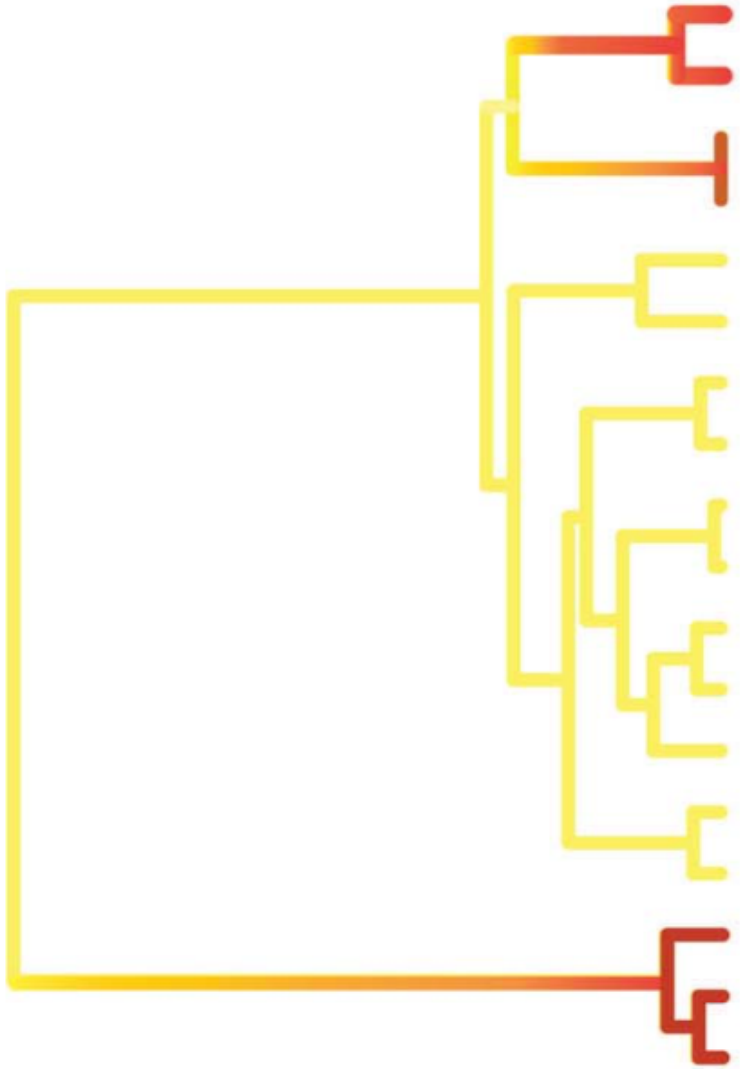
Exemplos de modelos evolutivos

(2) **Relógio Molecular Local** (Yoder & Tang, 2000): Espécies proximamente aparentadas compartilham a mesma taxa de substituição mas, espécies distantemente aparentadas não. **Taxas de substituição são constantes dentro de sub-clados** (i.e., clados vermelhos apresentam uma taxa de substituição maior que os clados azuis).



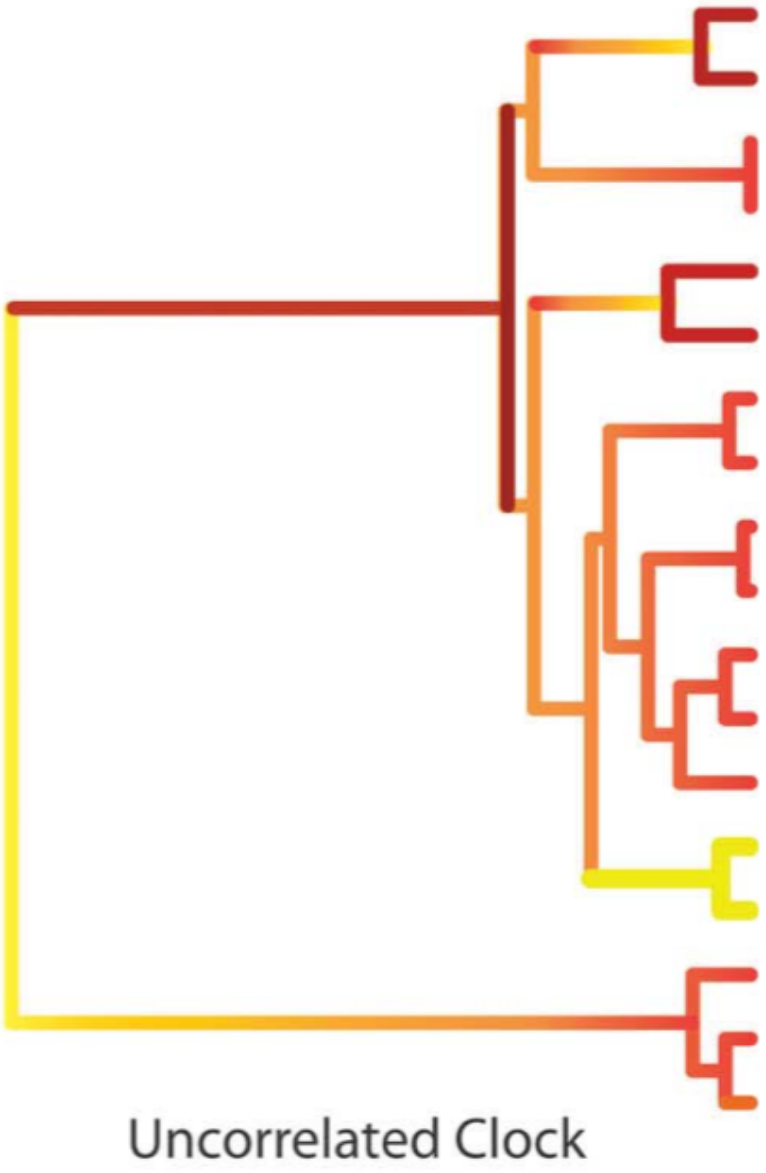
Exemplos de modelos evolutivos

(3) **Taxas de substituição auto-correlacionadas** (Sanderson, 1997, 2002; Thorne et al., 1998): Taxa de substituição varia na árvore como um todo, “evoluindo” de forma gradual e auto-correlacionada ao longo da árvore. **Cada ramo da árvore herda a taxa de variação do seu ancestral imediato**.



Exemplos de modelos evolutivos

(4) **Taxas de substituição não-correlacionadas** (Drummond et al., 2006): A taxa de substituição de cada ramo é estimada a partir de uma distribuição paramétrica ("normal") de forma **independente** (não correlacionada à taxa de substituição dos ancestrais).



Abordagem Bayesiana

BEAST “Bayesian Evolutionary Analysis Sampling Trees”

- “Node-Age Calibration”: Método que utiliza fósseis (majoritariamente) como pontos de calibração.

- Assume um modelo estocástico e taxas de substituição não correlacionadas (“uncorrelated relaxed clock”); mas podemos empregar outros.

- Estima a filogenia e taxa de substituição ao mesmo tempo em que estima as idades (não depende de uma árvore inicial), dado um modelo evolutivo e um modelo de relógio.

- Gera um grande número de árvores e calcula os intervalos de confiança diretamente.

- Considera incerteza: (i) na idade do fóssil; (ii) no posicionamento do fóssil; (iii) na árvore filogenética; (iv) nas taxas de mudanças evolutivas.

Métodos que incorporam heterogeneidade nas taxas de substituição de DNA

(1) “Node-age Calibrations”

(a) NPRS & Verossimilhança Penalizada (r8s; Sanderson, 1997, 2002)

(b) Multidivtime (Thorne & Kishino, 2002; Shiguno, 2005)

(c) BEAST (Drummond & Rambout, 2005)

(d) Use of multiple calibration points (Marshall, 2008)

(2) “Tip-dating” (Pyron, 2011; Ronquist et al., 2012)

(a) “Total-Evidence Approach” (Ronquist et al. 2012)

(b) “Fossilized Birth-Death Model” (Heath et al. 2014)

Métodos que incorporam heterogeneidade nas taxas de substituição de DNA

(1) “Node-age Calibrations”

(a) NPRS & Verossimilhança Penalizada (r8s; Sanderson, 1997, 2002)

(b) Multidivtime (Thorne & Kishino, 2002; Shiguno, 2005)

(c) BEAST (Drummond & Rambout, 2005)

(d) Use of multiple calibration points (Marshall, 2008)

(2) “Tip-dating” (Pyron, 2011; Ronquist et al., 2012)

(a) “Total-Evidence Approach” (Ronquist et al. 2012)

(b) “Fossilized Birth-Death Model” (Heath et al. 2014)

“Tip-Dating”

- Ao invés de posicionar fósseis na filogenia, incorpora os táxons **fósseis como terminais** (juntamente com as espécies viventes) na matriz utilizada para a reconstrução filogenética.

- Dados morfológicos e moleculares são geralmente combinados em uma única matriz. Outra alternativa seria a inclusão de DNA fóssil, o que pode ser difícil em alguns casos.

- O fóssil é automaticamente posicionado na filogenia e utilizado como ponto de calibração.

- O posicionamento exato dos fósseis pode ser complicado no caso de matrizes com grande quantidades de dados faltantes.

“Fossilized Birth-Death Model”

PNAS

PNAS PLUS

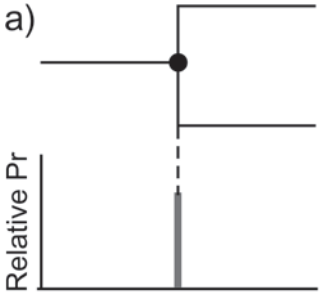
The fossilized birth–death process for coherent calibration of divergence-time estimates

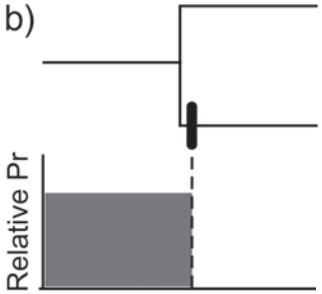
Tracy A. Heath[a,b], John P. Huelsenbeck[a,c], and Tanja Stadler[d,e,1]

[a]Department of Integrative Biology, University of California, Berkeley, CA 94720; [b]Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045; [c]Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia; [d]Department of Environmental Systems Science, Eidgenössische Technische Hochschule Zürich, 8092 Zurich, Switzerland; and [e]Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich, 4058 Basel, Switzerland

Edited by Joseph Felsenstein, University of Washington, Seattle, WA, and approved May 30, 2014 (received for review October 10, 2013)

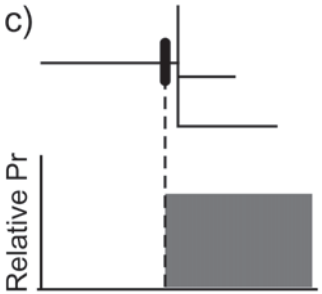
Point-calibration

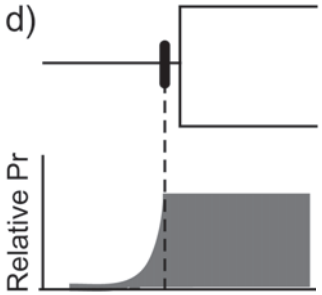




Hard Minimum Bound

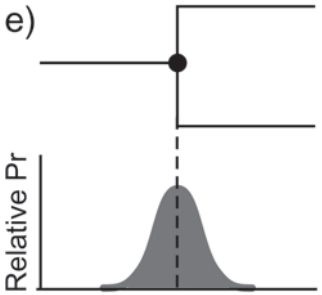
Hard Maximum Bound

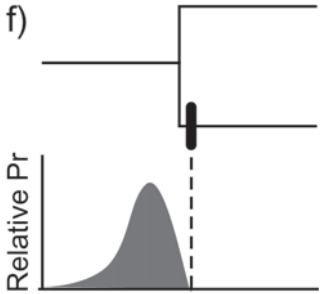




Soft Maximum Bound

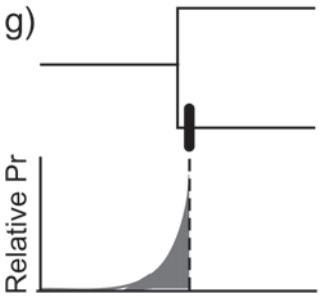
Normal Distribution





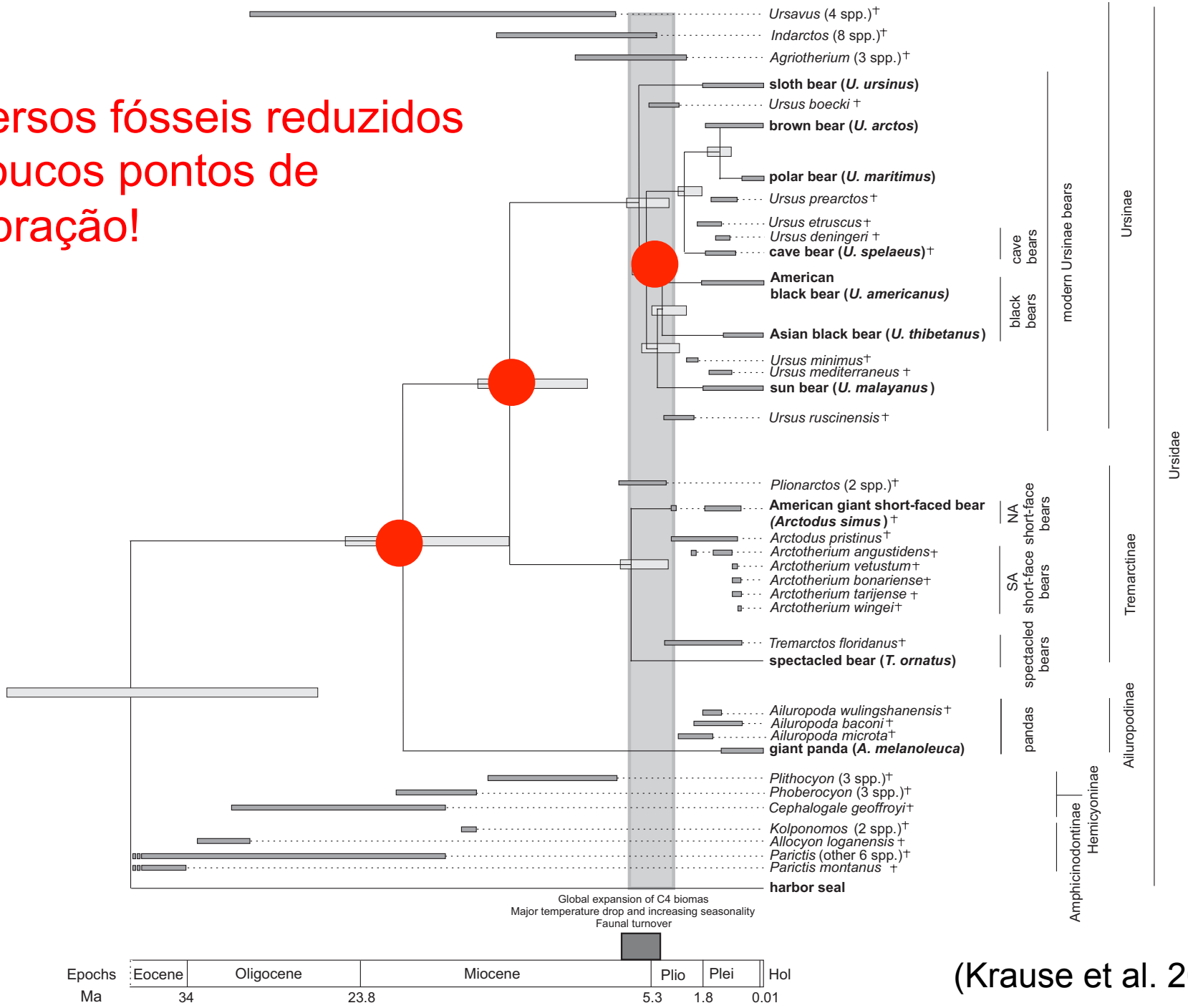
Lognormal Distribution

Exponential Distribution



Especificar qual o modelo a ser usado é um desafio!

Diversos fósseis reduzidos a poucos pontos de calibração!

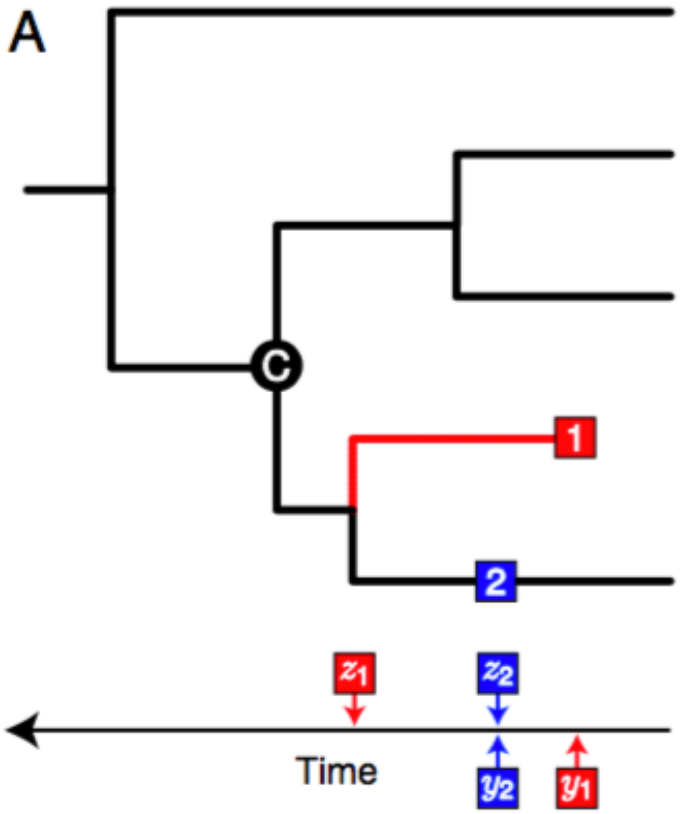


(Krause et al. 2008)

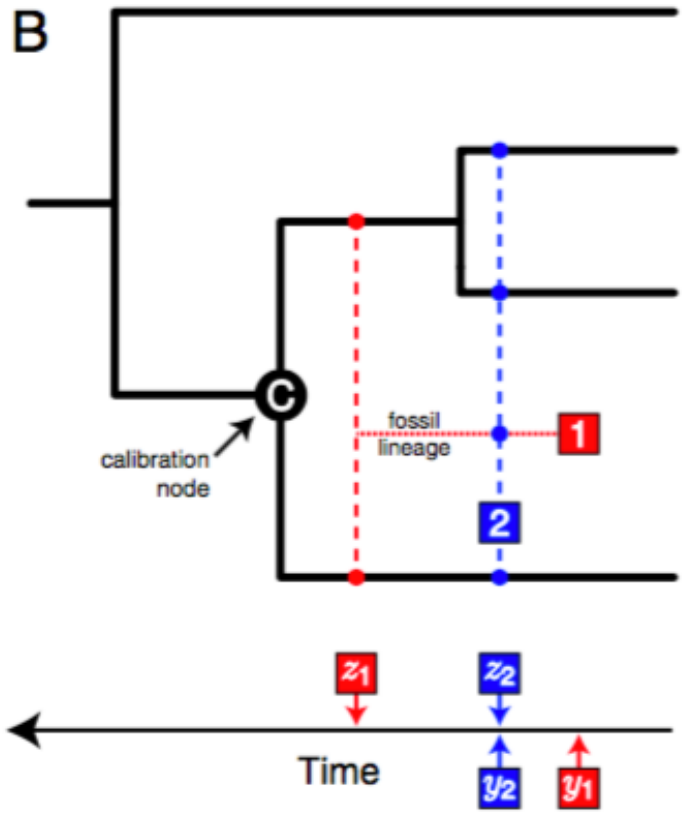
“Fossilized Birth-Death Model” (Heath et al., 2014)

- Este modelo (FBD) se baseia no fato de que a **diversificação de espécies viventes e fósseis são parte do mesmo processo** evolutivo e propõe o uso de um único modelo de diversificação para os táxons.
- O modelo calibra as idades de diversificação na árvore inteira utilizando um único modelo de diversificação para espécies viventes e fósseis de um dado clado.
- **Elimina a necessidade de designar fósseis para um único nó**, com uma mesma idade mínima fixada. Múltiplos fósseis podem ser designados para um mesmo intervalo, com idades variáveis.
- Supostamente captura melhor a incerteza nas idades de divergência estimadas.

Árvore FBD



Parentesco entre táxons viventes e extintos apresentado



Parentesco com táxons extintos ignorado; dois fósseis utilizados para calibrar o mesmo nó.

y = idade do fóssil; z = momento no qual o fóssil é "conectado" à árvore

Comparação entre métodos

An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae)

Dahiana Arcila [a,b,*,1], R. Alexander Pyron [a], James C. Tyler [b], Guillermo Ortí [a], Ricardo Betancur-R. [b,c,1]

[a] Department of Biological Sciences, The George Washington University, 2023 G St. NW, Washington, DC 20052, United States
[b] Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, P.O. Box 37012, MRC 159, Washington, DC 20013, United States
[c] Department of Biology, University of Puerto Rico – Río Piedras, P.O. Box 23360, San Juan 00931, Puerto Rico

- Idades provenientes de idades calibradas nos nós ("**node-age calibrations**") estimaram **idades** consistentemente **mais jovens** do que as idades geradas através de metodologias de "tip-dating";
- Idades provenientes de "**node-age calibrations**" são mais **consistentes** com o **registro estratigráfico** (diferença em ca. 100 My nas idades estimadas!)
- Grau de precisão das idades estimadas com "tip-dating" aumenta com o número de fósseis analisados e com a proximidade entre o fóssil e nó sendo avaliado.

Formas para aumentar a confiança nas idades estimadas

(1) Utilizar múltiplos conjuntos de dados e modelos evolutivos

"Trying to estimate time of divergence from one protein is like trying to estimate the average height of humans by measuring one human" (Hillis, 1996)

(2) Utilizar múltiplos pontos de calibração e explorar metodologias diferentes;

(3) Assegurar-se que a identificação e idade dos fósseis estão corretos;

(4) Estimar intervalos de confiança para as estimativas.