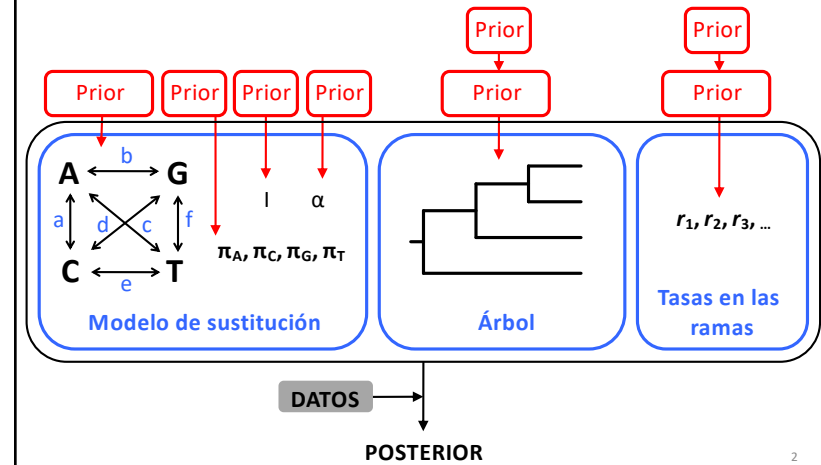


Clase 6

Modelos y priors**Modelos jerárquicos Bayesianos****Distribuciones para priors****Priors**

- Reflejan nuestra expectativa (e incertidumbre) sobre los valores de parámetros (sin conocimiento sobre los datos)
 - Observaciones previas
 - Opinion personal
 - Uso de un modelo biológico

Distribuciones continuas

- Uniforme
- Normal

Usadas para definir las distribuciones del prior de parámetros continuos

- Exponencial
- Lognormal
- Gama

Usadas para definir las distribuciones del prior de parámetros continuos que no pueden tomar valores negativos

- Beta
- Dirichlet

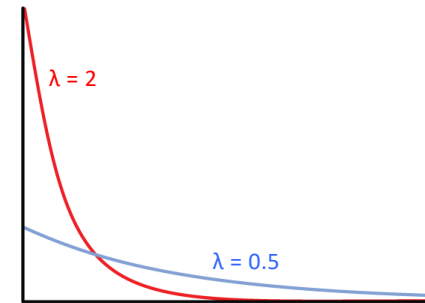
5

Distribuciones continuas

- Uniforme
- Normal
- **Exponencial**
- Lognormal
- Gama
- Beta
- Dirichlet

Parámetros

- λ = tasa de decaimiento



6

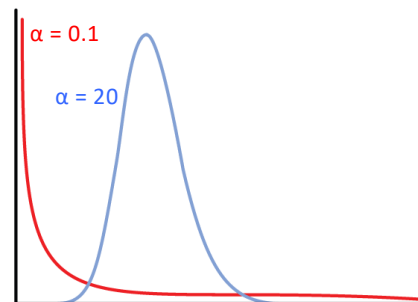
Continuous distributions

- Uniforme
- Normal
- Exponencial
- Lognormal
- **Gama**
- Beta
- Dirichlet

Parámetros

- α = forma
- β = escala inversa

Gama escalada:
 $\alpha = \beta$



7

Continuous distributions

- Uniforme
- Normal
- Exponencial
- Lognormal
- Gama

- Beta
- Dirichlet

Usados para definir las distribuciones del prior para parámetros simplex

- Las frecuencias de las bases
- Las tasas relativas del modelo GTR

8

Distribución Dirichlet

- Uniforme
- Normal
- Exponencial
- Lognormal
- Gama
- Beta
- **Dirichlet**

Parámetros

- $\alpha_1, \alpha_2, \dots$ = parámetros de forma

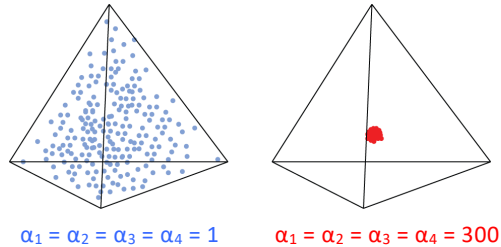


Imagen modificada de las clases de Paul Lewis 9

Distribuciones discretas

- Distribución Bernoulli
- Binomial
- Multinomial
- Poisson

10

Default priors

	<i>BEAST2</i>	<i>MrBayes</i>
Parámetros de la matriz de tasas	Gama(0.05,10)	Dirichlet(1,1,1,1,1,1)
Frecuencias de las bases	Uniforme(0,1)	Dirichlet(1,1,1,1)
Parámetro para la forma (α)	Exponencial(1)	Exponencial(2)
Proporción de sitios invariables	Uniforme(0,1)	Uniforme(0,1)

11

Priors no informativos

- **Prior plano o uniforme**
- **Prior de Jeffrey**
 - No varía al reparametrizar
 - Solo funciona bien para modelos de un solo parámetro
- **Priors de referencia**
 - Maximiza la distancia esperada entre el prior y el posterior
 - Maximiza el efecto de los datos al posterior

12

Prior del árbol

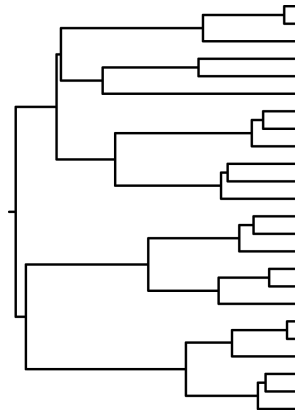
Prior del árbol

1. Podemos usar un **prior plano** (*MrBayes*)
 - Todos los árboles tiene una probabilidad igual
 - También necesita un prior para las longitudes de las ramas o las edades de los nodos
 2. Usa un **modelo biológico** (*BEAST* y *MrBayes*)
 - Entre varias especies: modelo de especiación
 - En una especie: modelo coalescente
- } **Priors para árboles enraizados**

14

Modelo de especiación

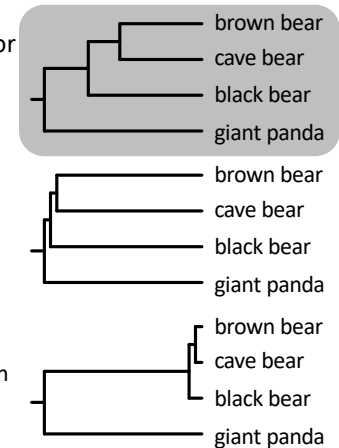
- La forma del árbol esta descrita por un proceso estocástico
- **Proceso Yule**
 - La raíz se divide en dos
 - Los linajes se dividen a una tasa constante
 - Simula al proceso de especiación
- **Proceso Birth-death**
 - Permite que los linajes se extingan



15

Modelo de especiación

- La forma del árbol esta descrita por un proceso estocástico
- **Proceso Yule**
 - La raíz se divide en dos
 - Los linajes se dividen a una tasa constante
 - Simula al proceso de especiación
- **Proceso Birth-death**
 - Permite que los linajes se extingan

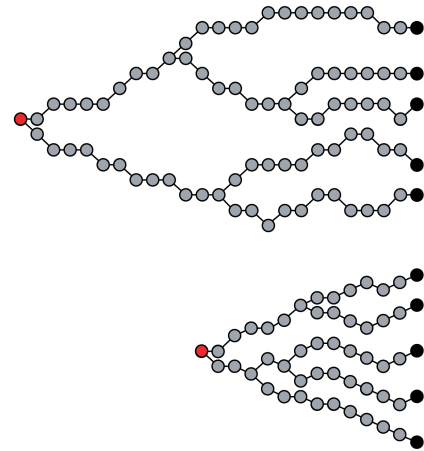


16

Coalescent model

Tamaño
constante

Crecimiento
exponencial



17

¿Cómo elegir un prior para el árbol?

- Se puede probar si los resultados son robustos a la elección
- Datos mixtos: múltiples secuencias de cada especie
 - El prior Birth-death generalmente funciona bien
- Se pueden comparar los priors del árbol usando selección de modelos Bayesiana

18

Selección de modelos Bayesiana

Bayesian model selection

- La selección de modelos Bayesiana se basa en la probabilidad marginal de los datos condicionados en el modelo:

$$\Pr(D|M)$$

- Esto es un promedio de la verosimilitud, pesada por el prior

La verosimilitud marginal
del modelo
Marginal likelihood

Probabilidad	Modelo 1
Árbol 1	0.1
Árbol 2	0.7
Árbol 3	0.15
Árbol 4	0.05
Suma	1

20

Bayesian model selection

- La selección de modelos Bayesiana se basa en la probabilidad marginal de los datos condicionados en el modelo:

$$\frac{\Pr(\theta) \Pr(D | \theta)}{\Pr(D)}$$

$\Pr(D | M)$

- Esto es un promedio de la verosimilitud, pesada por el prior

La verosimilitud marginal del modelo
Marginal likelihood

Probabilidad	Modelo 1
Árbol 1	0.1
Árbol 2	0.7
Árbol 3	0.15
Árbol 4	0.05
Suma	1

21

Selección de modelos Bayesiana

- Comparamos la verosimilitud marginal de modelos candidatos
- La proporción de verosimilitudes marginales es el **Factor de Bayes**

$$BF = \frac{\Pr(D | M_1)}{\Pr(D | M_2)}$$

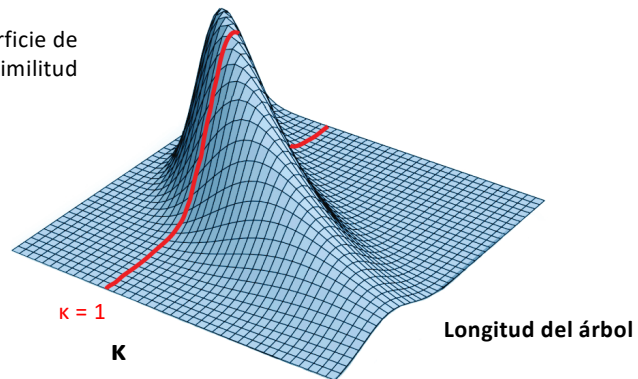
$$\log BF = \log \Pr(D | M_1) - \log \Pr(D | M_2)$$

- Los modelos no tienen que estar anidados
- No es necesario corregir por el número de parámetros

22

Selección de modelos Bayesiana

Superficie de verosimilitud



23

Bayesian model selection

- Interpretando el Factor de Bayes (*Bayes Factor*)

BF	logBF	Evidencia contra M_2
1 – 3	0 – 1	No merece mención
3 – 20	1 – 3	Positiva
20 – 150	3 – 5	Contundente
> 150	> 5	Muy contundente

Kass & Raftery (1995) *J Am Stat Assoc* ²⁴

Ventajas y problemas

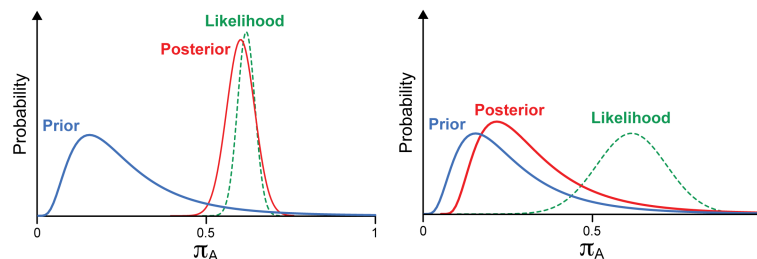
Ventajas

- Puede implementar **modelos altamente parametrizados**
- **Estimar la incertidumbre en el estimativo de árbol es fácil**
 - Mientras en máxima verosimilitud solo se puede hacer indirectamente (usando el bootstrap)
- **Las probabilidades a posterior** tienen una interpretación intuitiva
- Puede incorporar información independiente (en el prior)

26

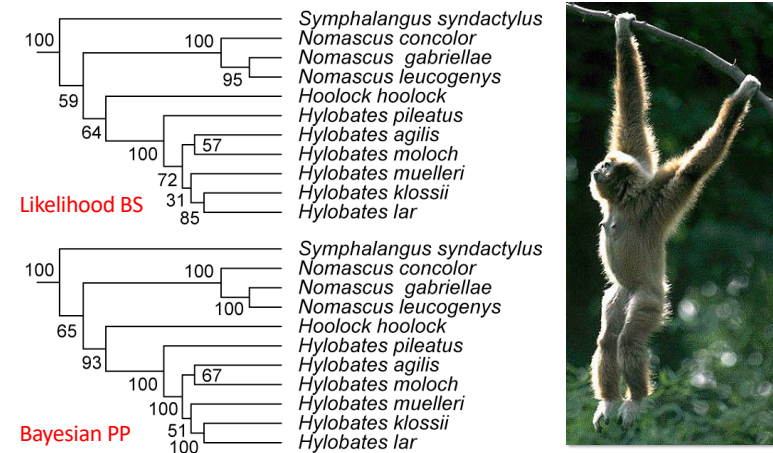
La influencia del prior

- El posterior es sensible al prior
- Este problema ocurre cuando los datos son poco informativos, o el prior es muy informativo, o ambas



27

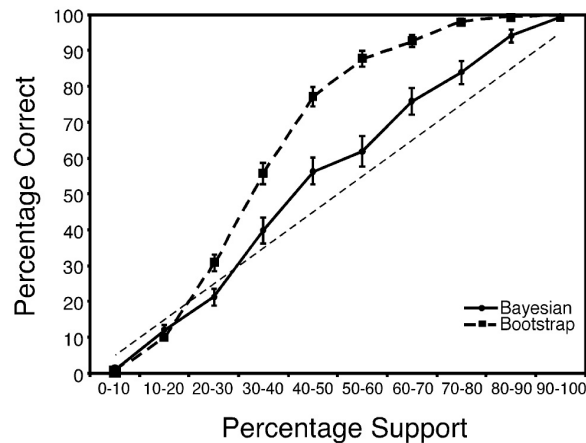
Soporte de los nodos



Chatterjee et al. (2009) BMC Evol Biol

28

Soporte de los nodos



Wilcox et al. (2002) *Mol Phylogenet Evol* 29

BEAST 1



- **Bayesian Evolutionary Analysis by Sampling Trees**
- Analiza datos a nivel poblacional o de multiples especies
- Estima simultáneamente el árbol y las fechas de los nodos
- Implementa un amplio rango de relojes moleculares
- Implementa un amplio rango de modelos demográficos y de especiación

30



Bayesian evolutionary analysis by sampling trees

- Una renovación de *BEAST* para hacerlo más modular
- Los usuarios pueden extender *BEAST* adicionando paquetes
- Tiene algunos priors que no estan en *BEAST 1*
- Se puede usar para hacer simulación de datos

Para una comparación entre *BEAST 1* and *2*:
www.beast2.org/beast-features



31

MrBayes



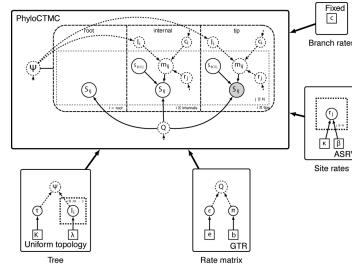
- Diseñado para análisis de datos de múltiples especies
- Estima simultáneamente el árbol y las fechas de los nodos
- Implementa un amplio rango de relojes moleculares
- Implementa un amplio rango de modelos demográficos y de especiación
- Múltiples cadenas y diagnósticos de MCMC

32

RevBayes

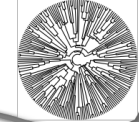


- Usa su propio lenguaje de codificación, Rev
- Permite construir modelos graficos interactivamente
- Flexible para uso en estudios de simulaciones e inferencia
- Tiene Desarrollo continuo



33

EXABAYES



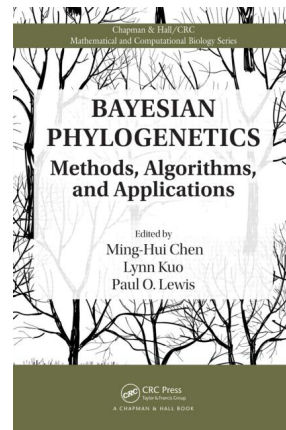
XXXXXXXXXX

- Para análisis de cantidades grandes de datos en computadores de alto rendimiento
- Tiene priors disponibles parecidos a aquellos en las primeras versiones de *MrBayes*
- Opciones limitadas, sin estimación de tasas moleculares
- Una adaptación de verosimilitud de *RAXML*

34

Referencias útiles

- **The impact of the tree prior on molecular dating of data sets containing a mixture of inter- and intraspecies sampling**
Ritchie, Lo, & Ho (2017) *Syst Biol*, 66: 413–425.



35