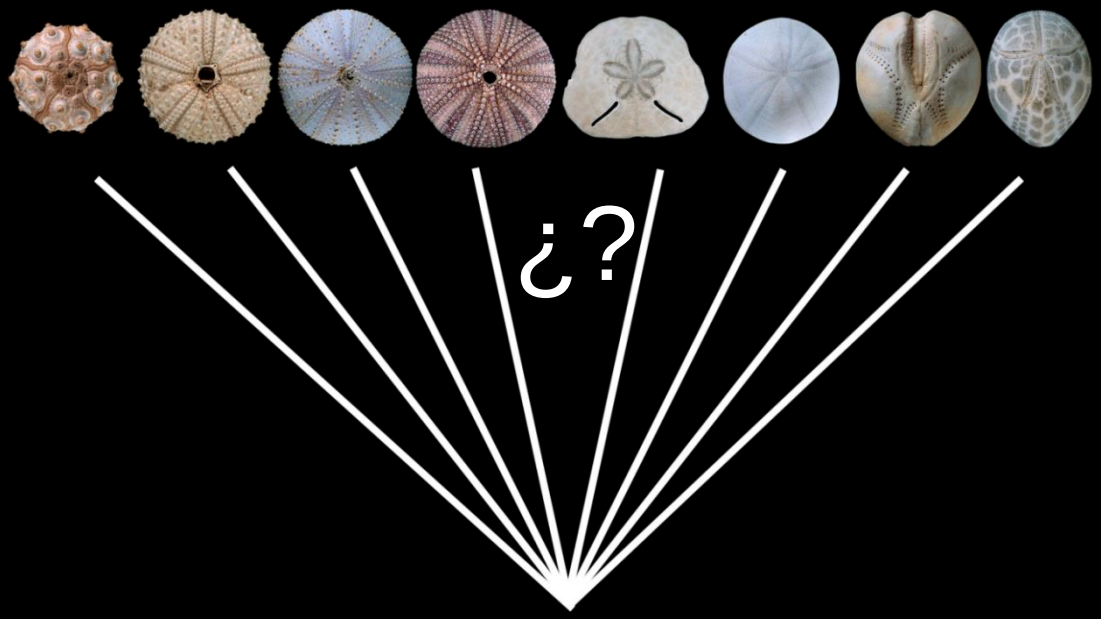
Inferencia bayesiana

En esta práctica aprenderemos a resolver la filogenia de nuestro grupo modelo aplicando inferencia bayesiana

a. Nociones básicas de probabilidad bayesiana aplicada a la filogenia molecular
b. Reconstrucción de la filogenia de nuestra selección de erizos empleando MrBayes
c. Comparación general de los métodos de reconstrucción filogenética tratados en prácticas

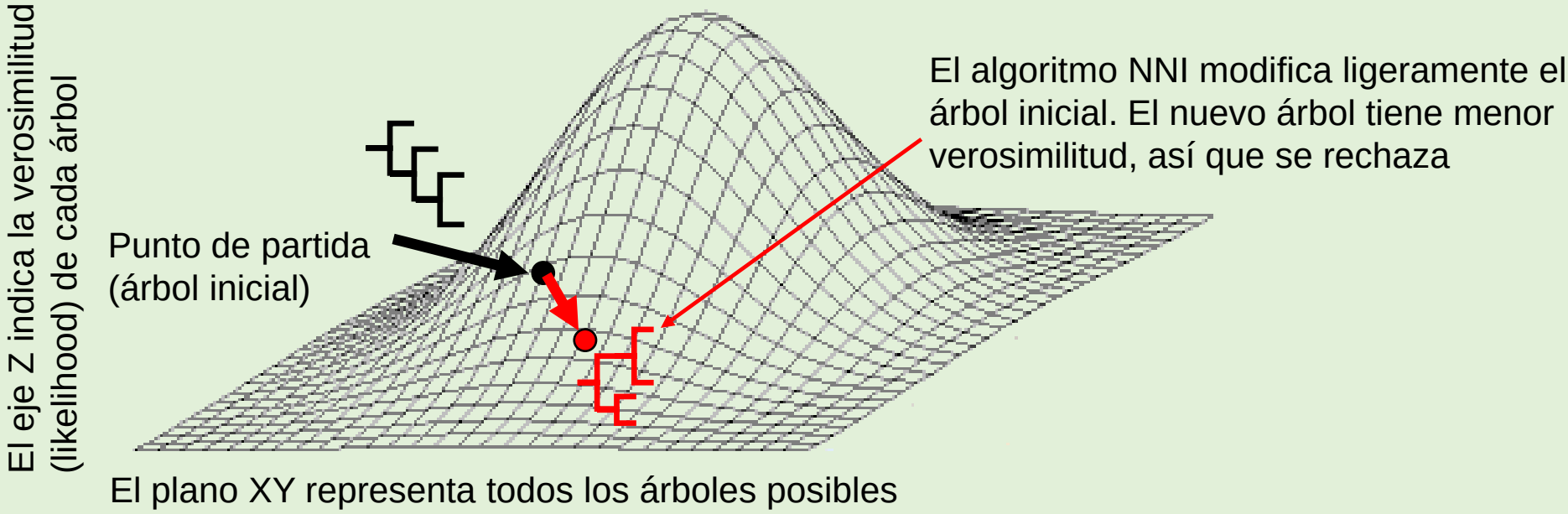


Concluimos la práctica anterior reconstruyendo la evolución de nuestros erizos empleando una aproximación de **máxima verosimilitud** (*maximum likelihood*, ML o MV)

Como punto de partida usamos:

- Los datos (el alineamiento de las secuencias 18S)
- Un modelo de sustitución adecuado (K2+G)

El software MEGA exploró el inabarcable conjunto de todos los árboles posibles ayudándose de un algoritmo heurístico (en nuestro caso, NNI)

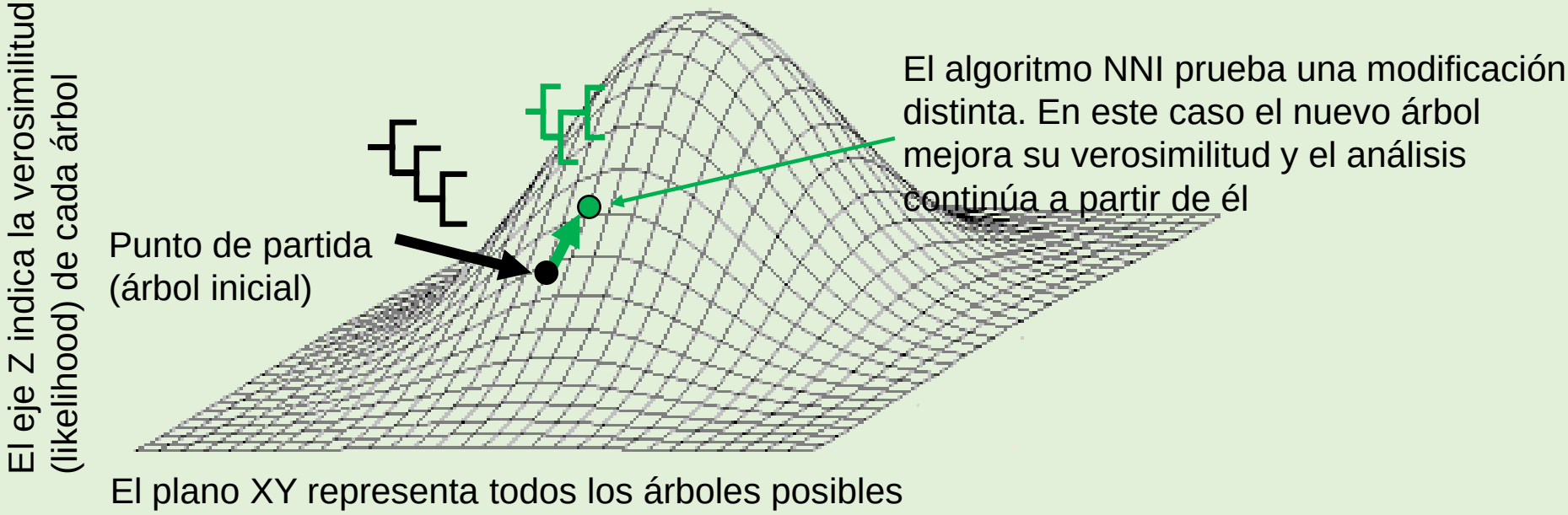


Concluimos la práctica anterior reconstruyendo la evolución de nuestros erizos empleando una aproximación de **máxima verosimilitud** (*maximum likelihood*, ML o MV)

Como punto de partida usamos:

- Los datos (el alineamiento de las secuencias 18S)
- Un modelo de sustitución adecuado (K2+G)

El software MEGA exploró el inabarcable conjunto de todos los árboles posibles ayudándose de un algoritmo heurístico (en nuestro caso, NNI)

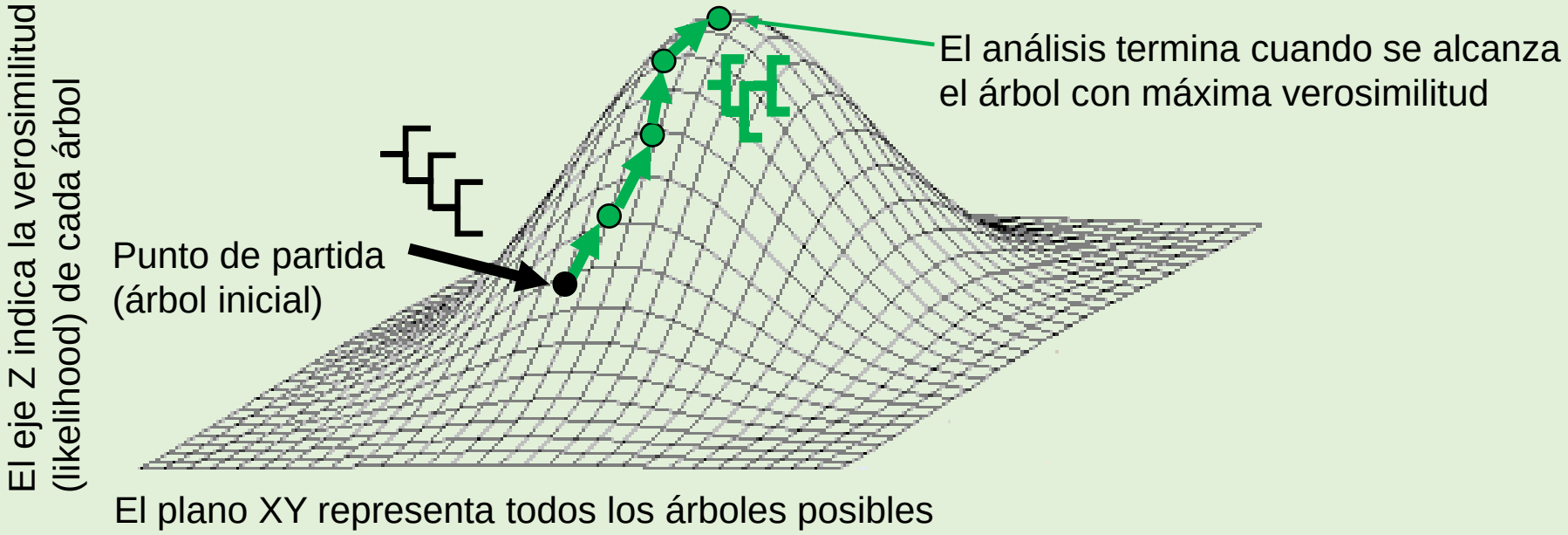


Concluimos la práctica anterior reconstruyendo la evolución de nuestros erizos empleando una aproximación de **máxima verosimilitud** (*maximum likelihood*, ML o MV)

Como punto de partida usamos:

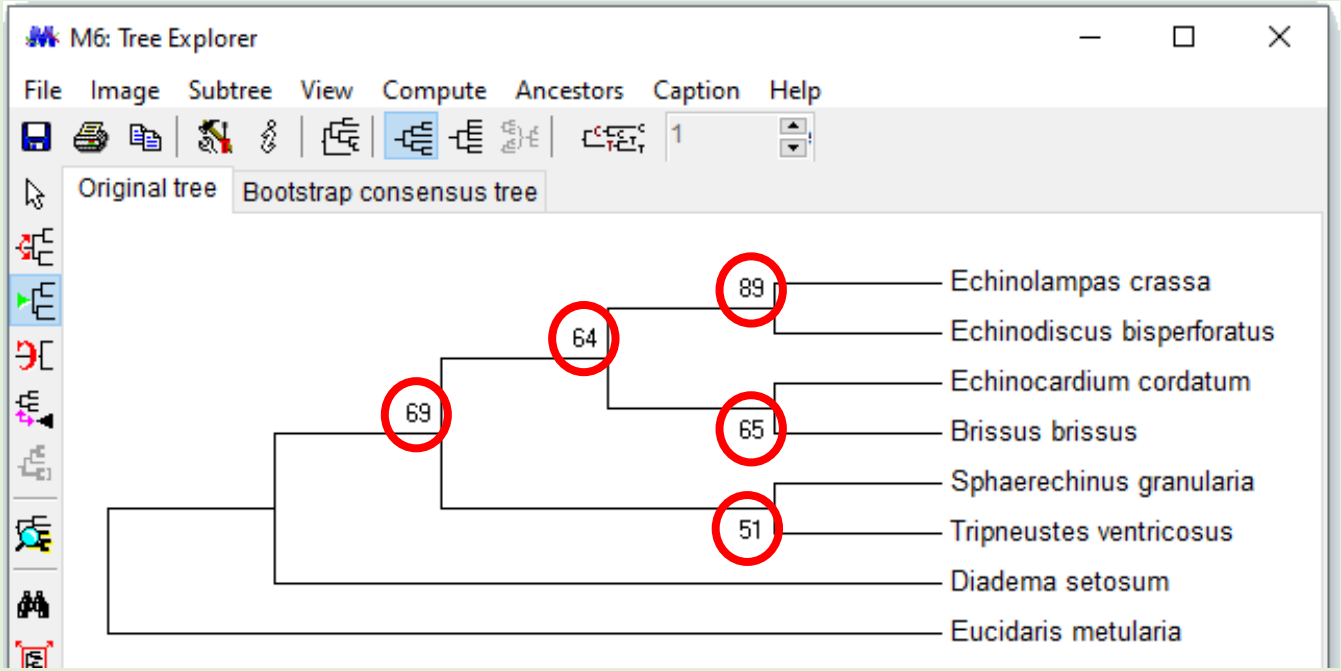
- Los datos (el alineamiento de las secuencias 18S)
- Un modelo de sustitución adecuado (K2+G)

El software MEGA exploró el inabarcable conjunto de todos los árboles posibles ayudándose de un algoritmo heurístico (en nuestro caso NNI)



Concluimos la práctica anterior reconstruyendo la evolución de nuestros erizos empleando una aproximación de **máxima verosimilitud** (*maximum likelihood*, ML o MV)

El análisis termina cuando se alcanza el árbol con máxima verosimilitud, es decir, el árbol que maximiza la probabilidad de obtener nuestro alineamiento



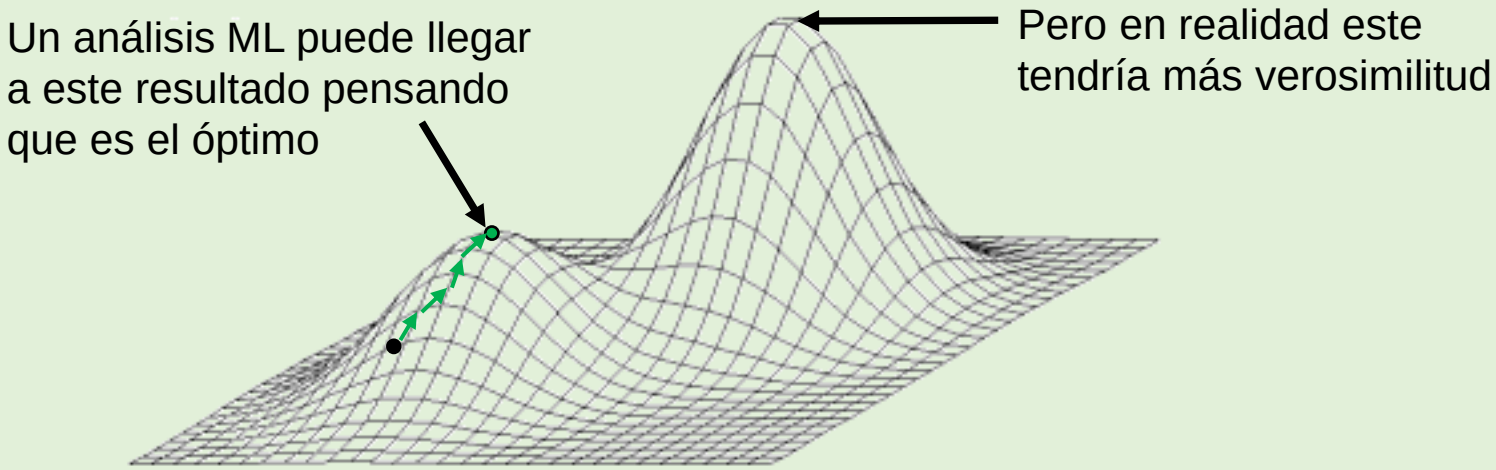
Finalmente, para evaluar cómo de robusto era el resultado, MEGA ejecutó un análisis de bootstrap con 100 pseudorréplicas. Los valores de bootstrap en cada rama nos dan una idea del apoyo de cada una de ellas

El criterio de ML para la reconstrucción filogenética es muy potente y sofisticado, pero también adolece de algunas debilidades

- No puede calcular el árbol más probable dados nuestros datos (calcula el árbol que haría más probable producir nuestros datos, de haber sido cierto, que no es lo mismo)

- El análisis es muy costoso de computar (esto es un ejemplo sencillo, imagina analizar cientos de especies y miles de pares de bases)

- Se corre el riesgo de que el análisis se quede estancado en un máximo local



El criterio de inferencia bayesiana resuelve en parte estos problemas al aproximarse a la reconstrucción filogenética desde una óptica un poco distinta de en qué consiste la probabilidad

Probabilidad frecuentista. A base de repetir muchos sucesos o experimentos se alcanza una probabilidad basada en la frecuencia pasada de un evento concreto

¿Lloverá el 1 de abril? Durante los últimos 150 años, el día 1 de abril en Madrid ha sido lluvioso 65 veces*, así que hay una probabilidad del 43% (65/150=0.43) de lluvia el 1 de abril de 2022

* Invent

El criterio de inferencia bayesiana resuelve en parte estos problemas al aproximarse a la reconstrucción filogenética desde una óptica un poco distinta de en qué consiste la probabilidad

Probabilidad bayesiana. Concepto más abstracto basado en la incertidumbre a partir de cierta información previa (*prior*), aunque sea vaga

La probabilidad bayesiana nos permite calcular la **probabilidad posterior** de un suceso despejándolo gracias al teorema de Bayes de donde obtiene el nombre.

La probabilidad posterior es un tipo de **probabilidad condicional**:
probabilidad de que ocurra un suceso X sabiendo que ha ocurrido Y P(X/Y)

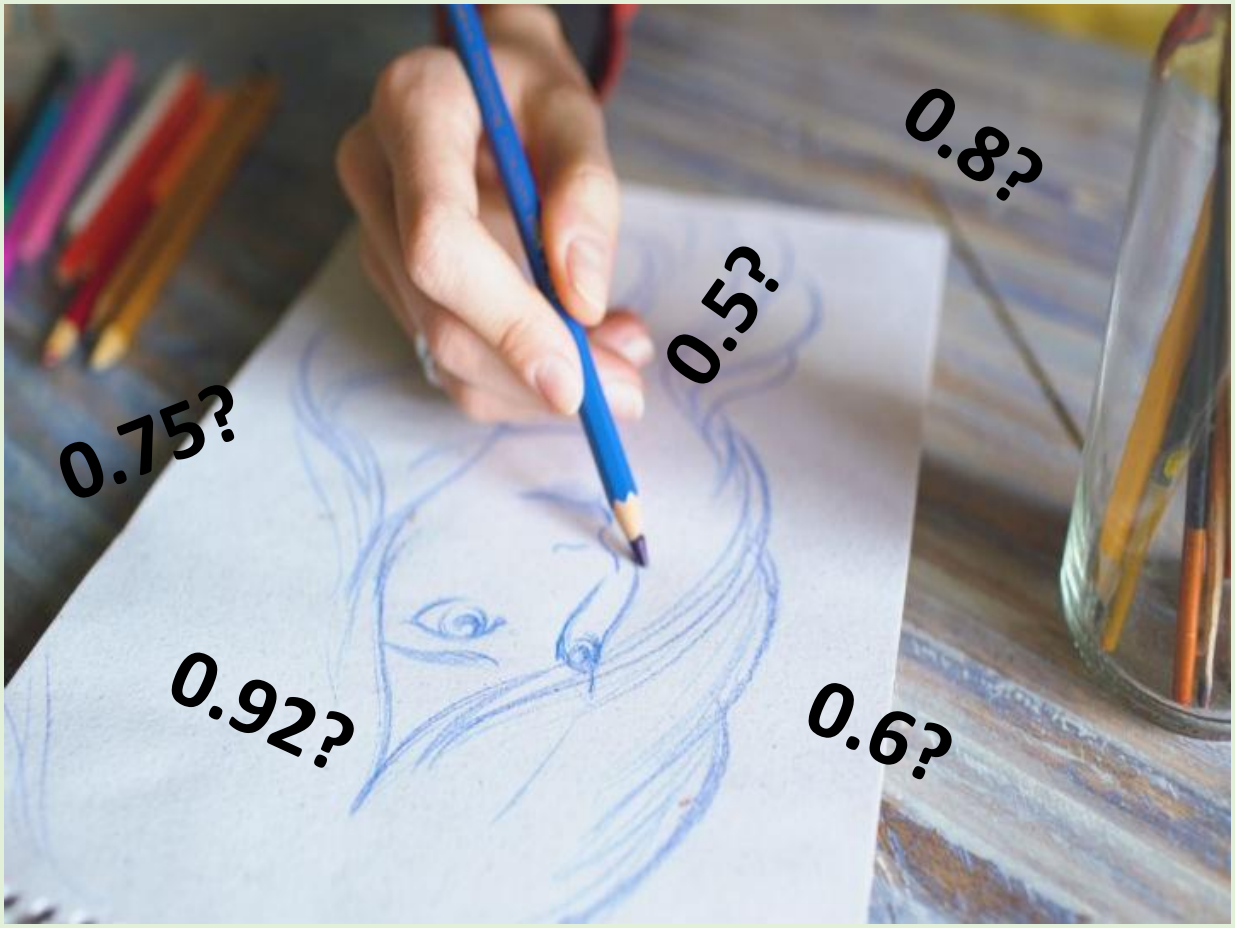
$$P\ (A/B) = \frac{P(B/A)\ *\ P(A)}{P(B)}$$

El teorema de Bayes nos permite “invertir” y despejar probabilidades condicionales

* Este señor en realidad no fue Bayes, sino alguien que vivió 200 años después, pero por algún motivo, se ha convertido en su imagen moderna

La estadística bayesiana puede ser útil para mejorar las estimaciones que hacen con probabilidad frecuentista precisamente por tener en cuenta unos datos de contexto

Imagina que te encuentras a alguien (estudiante de la UCM) dibujando en un lugar cualquiera del campus. Dibuja muy bien. ¿Qué probabilidad dirías que hay de que esta persona sea estudiante de bellas artes?



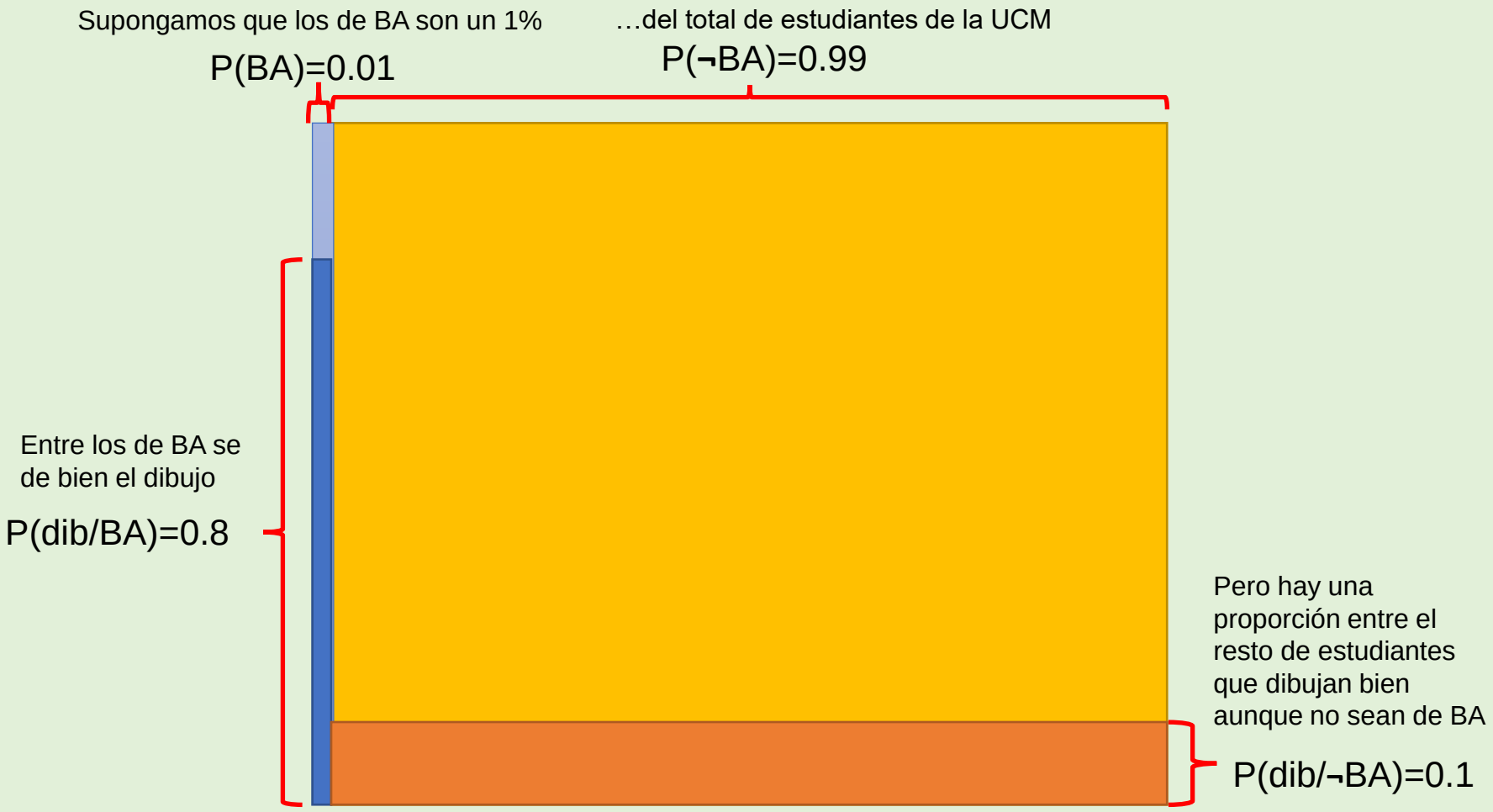
A lo que quieres responder, en realidad, es a una probabilidad condicional:

P(BA/dib)

(probabilidad de que la persona estudie bellas artes sabiendo que sabe dibujar bien)

Lo que a menudo podemos pasar por alto son los datos de contexto (**prior**) que en este caso sería una estimación de qué probabilidad hay de que sea estudiante de BA (un grado entre muchos)

¿Cuál es **P(BA/dib)**?



El teorema de Bayes nos permite resolver esta situación:

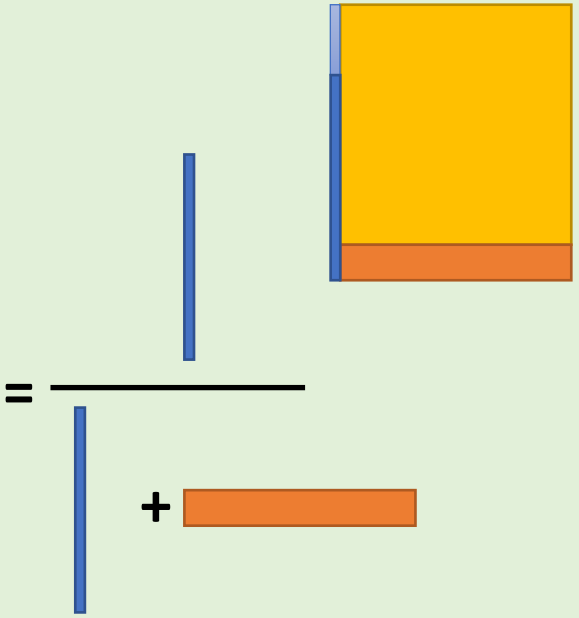
Probabilidad posterior

Verosimilitud

Prior

$$P(BA/dib) = \frac{P(dib/BA) * P(BA)}{P(dib)}$$

Probabilidad marginal



$$P(BA/dib) = \frac{P(dib/BA) * P(BA)}{P(BA)*P(dib/BA) + (\neg BA)*P(dib/\neg BA)} = \frac{0,8 * 0.01}{0.01 * 0.8 + 0,99 * 0,1}$$

$$P(BA/dib) = 0.075$$

En realidad es bastante poco probable que esa persona sea estudiante de BA. Modular nuestra estimación con un prior nos ha ayudado, en este caso, a mejorar nuestra primera impresión

En el contexto de la biología evolutiva, la inferencia bayesiana nos va a permitir aproximarnos al árbol que tiene una mayor probabilidad posterior

Inferencia bayesiana: el mejor árbol maximiza **P(H/D)**
o sea, la probabilidad de observar una Hipótesis (árbol) dados unos Datos (alineamiento)

Compárese con el criterio de máxima verosimilitud

Máxima Verosimilitud: el mejor árbol maximiza **P(D/H)**
o sea, la probabilidad de observar una Datos (alineamiento) dada una Hipótesis (árbol)

Probabilidad posterior
P de que un árbol sea correcto dados unos datos

Verosimilitud
P de que unos datos sean correctos dado un árbol

Prior
P de que un árbol sea correcto -sin saber nada del alineamiento-

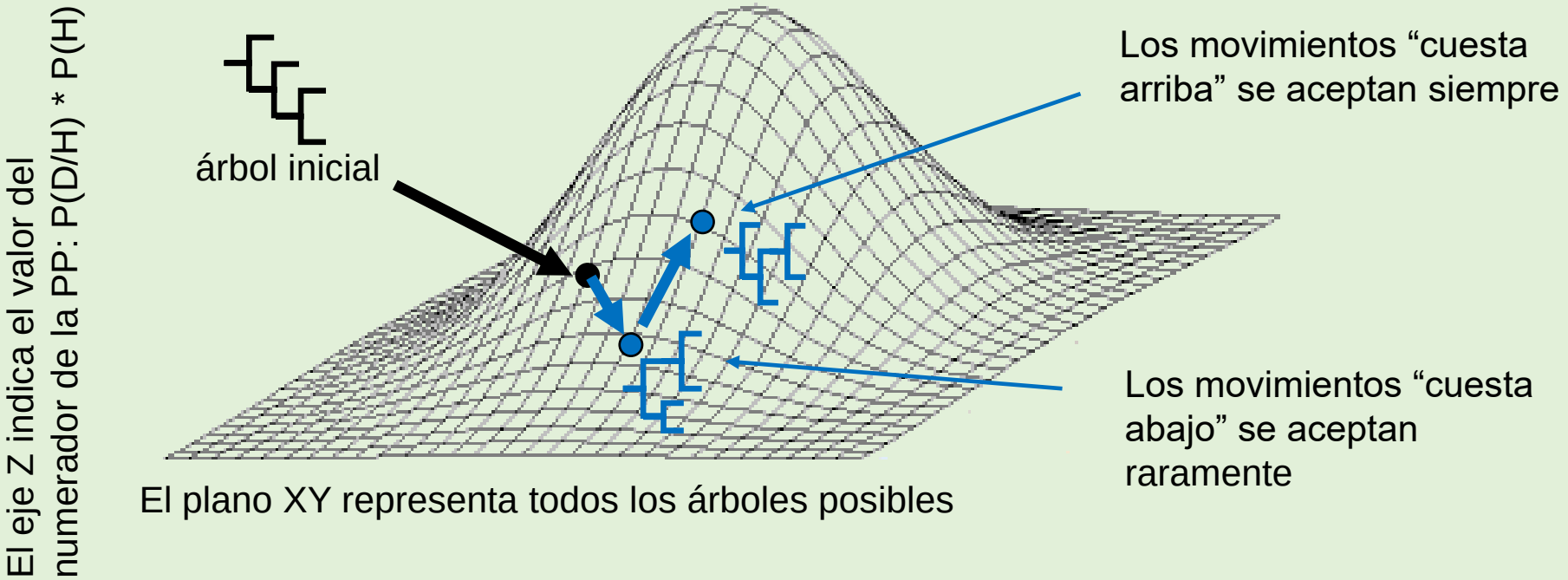
$$P(H/D) = \frac{P(D/H) * P(H)}{P(D)}$$

Probabilidad marginal
P de observar los datos

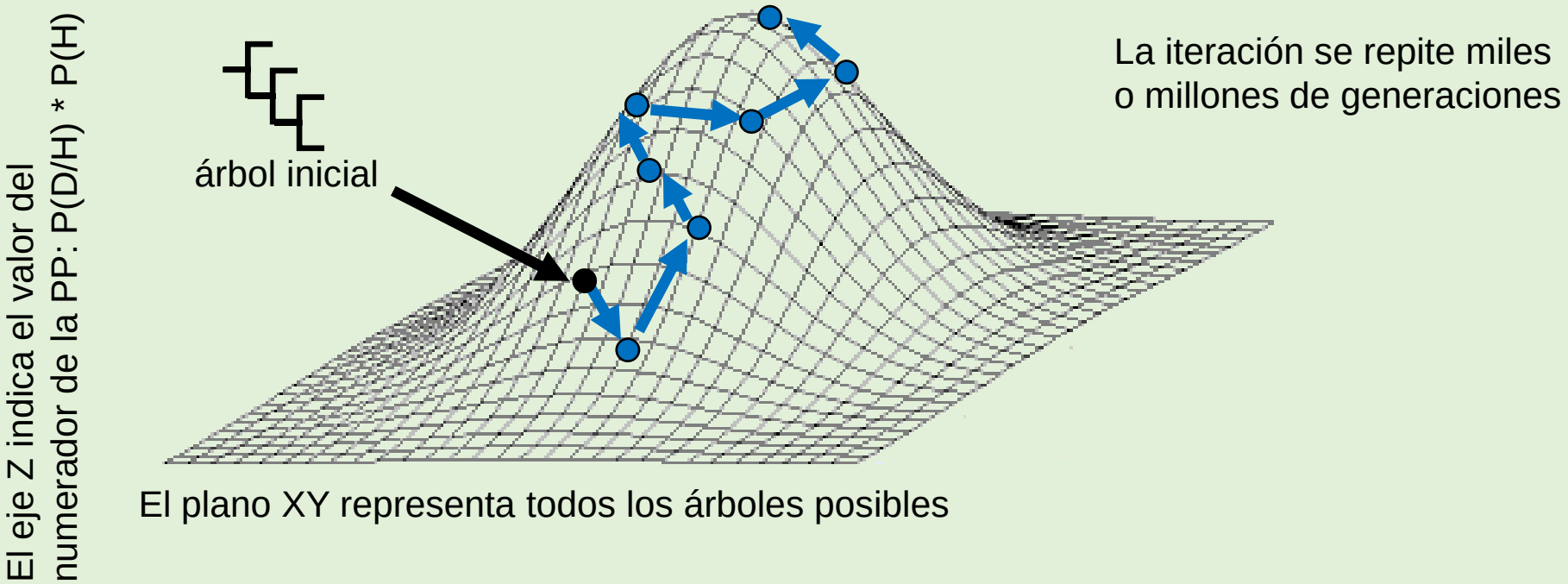
En la práctica, la probabilidad marginal es muy difícil de computar, así que un análisis bayesiano ser sirve de un algoritmo para funcionar: Markov Chain Monte Carlo (**MCMC**)

En lugar de intentar calcular un solo árbol, la inferencia bayesiana mediante MCMC nos dará un conjunto de árboles, una “nube” de certidumbre.

La “carrera” MCMC repite una serie de generaciones o intentos de movimiento desde un árbol inicial. El algoritmo aplica una variación a los parámetros del mismo y decide si aceptar el cambio o no evaluando el numerador de la fórmula de Bayes (Verosimilitud * P Anterior)



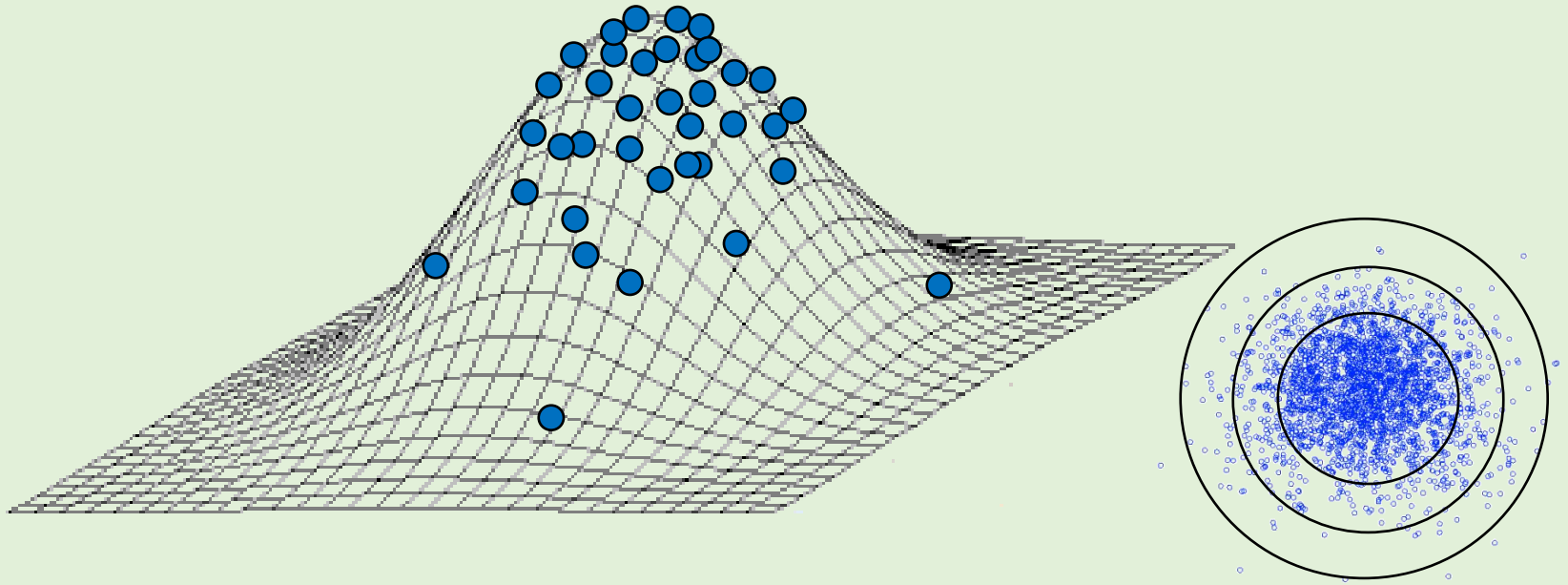
En la práctica, la probabilidad marginal es muy difícil de computar, así que un análisis bayesiano ser sirve de un algoritmo para funcionar: Markov Chain Monte Carlo (**MCMC**)



En la práctica, la probabilidad marginal es muy difícil de computar, así que un análisis bayesiano ser sirve de un algoritmo para funcionar: Markov Chain Monte Carlo (**MCMC**)

La nube de árboles obtenidos nos da una aproximación empírica de la máxima probabilidad posterior (no una certidumbre absoluta)

El eje Z indica el valor del numerador de la PP: P(D/H) * P(H)

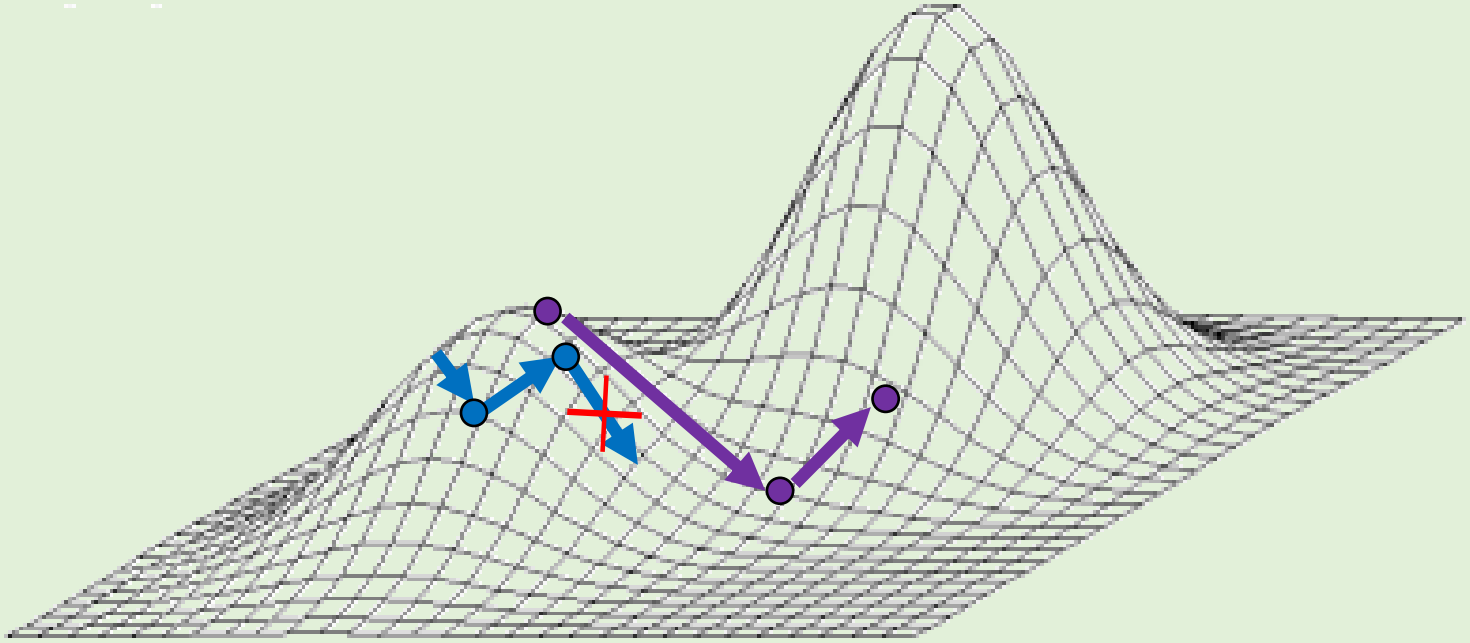


Es decir, la carrera pasa más tiempo cerca del árbol con máxima probabilidad posterior, aunque nunca estemos seguros de cuál es

La MCMC se puede modificar para ser menos propensa a caer en máximos locales. Para ello se ejecutan varias carreras y cadenas secundarias.

Las **cadenas frías** son más conservadoras y nunca aceptan pasos hacia abajo bruscos

Las **cadenas calientes** son más atrevidas y a veces sí que aceptan bajadas bruscas. Esto les permite escapar de máximos locales.



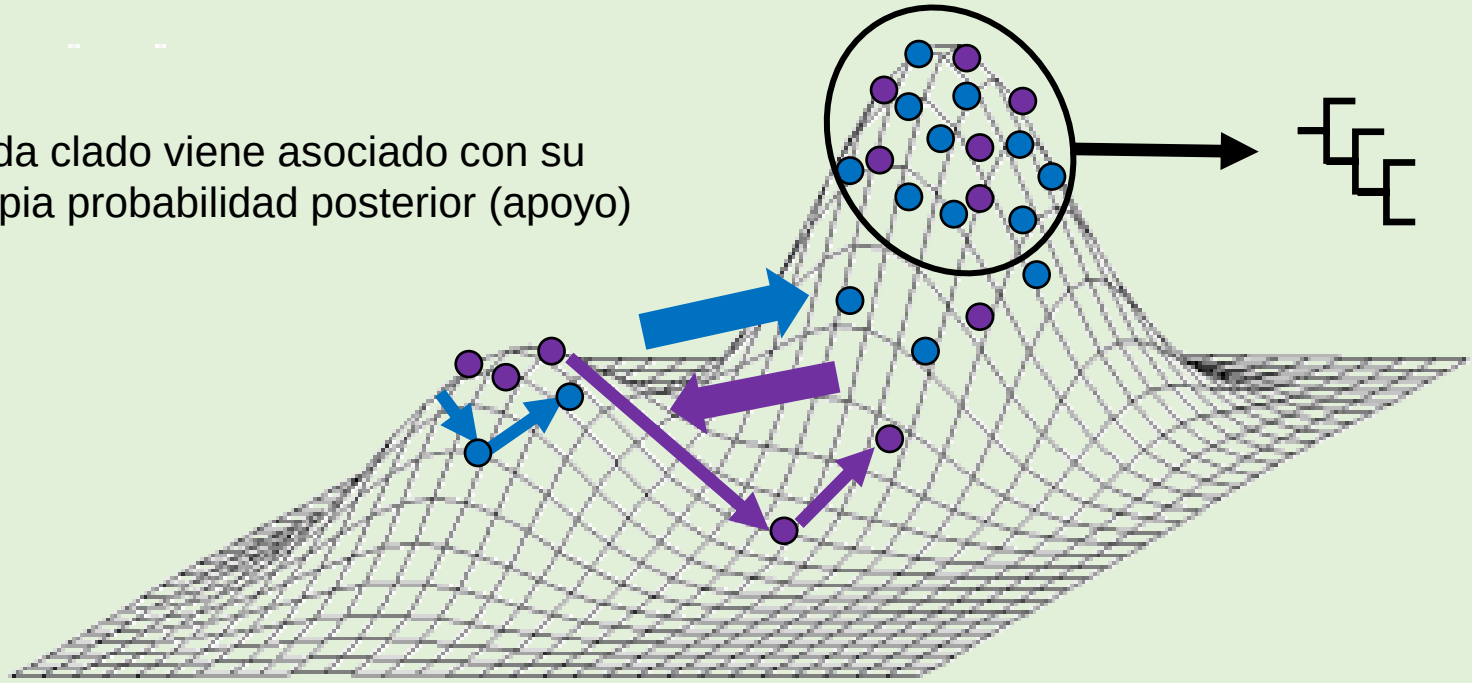
Periódicamente, las cadenas intercambian roles

De esta forma se “barre” el espacio de árboles de forma más eficaz

Cuando pasa el suficiente número de generaciones, todas las carreras y cadenas tienden a converger alrededor de un mismo pico

Finalizamos el análisis y obtenemos el árbol consenso de una muestra de los árboles de la nube

Cada clado viene asociado con su propia probabilidad posterior (apoyo)



¡Vamos a la parte práctica!
¿Cómo reconstruir una filogenia mediante inferencia bayesiana?

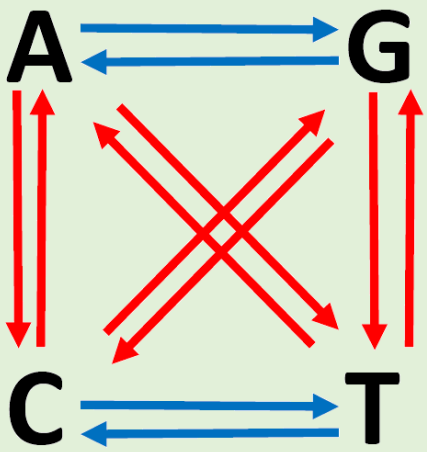
Necesitaremos:

1. Los **Datos**, es decir, el alineamiento erizos en formato nexus (*erizos.nex*)
(Nexus es un tipo de formato distinto a fasta, pero a efectos prácticos con la misma información)

¡Vamos a la parte práctica!
¿Cómo reconstruir una filogenia mediante inferencia bayesiana?

Necesitaremos:

2. Un **modelo de sustitución** de nucleótidos (igual que en el análisis de MV)
   Nuestro análisis previo de la práctica anterior sugirió K2+G como modelo óptimo



Frequencies of each nucleotide:

$\pi_A=\pi_G=\pi_T=\pi_C$

En el K2 las frecuencias se mantienen constantes y hay dos tasas de mutación: una para transiciones y otra para transversiones

El sufijo “+G” (o “+Γ”) indica que además esperamos que las tasas sean heterogéneas

¡Vamos a la parte práctica!
¿Cómo reconstruir una filogenia mediante inferencia bayesiana?

Necesitaremos:

3. (Esto es nuevo) Una **distribución de probabilidades anteriores** (***prior***) de los distintos parámetros que caracterizan el modelo. Es decir, la capacidad de calcular que un árbol sea el "correcto" sin tener en cuenta los datos

$$P(H/D) = \frac{P(D/H) * P(H)}{P(D)}$$

No es necesario que los *prior* sean muy detallados, e incluso se puede ejecutar un análisis con unos *prior* "planos", pero incluso una mínima información plausible los parámetros puede tener un efecto importante en los resultados.

¡Vamos a la parte práctica!
¿Cómo reconstruir una filogenia mediante inferencia bayesiana?

Necesitaremos:

3. (Esto es nuevo) Una **distribución de probabilidades anteriores** (***prior***) de los distintos parámetros que caracterizan el modelo. Es decir, la capacidad de calcular que un árbol sea el "correcto" sin tener en cuenta los datos

¿Esto qué quiere decir?

Calcular la probabilidad de un árbol cualquiera implica una serie de parámetros

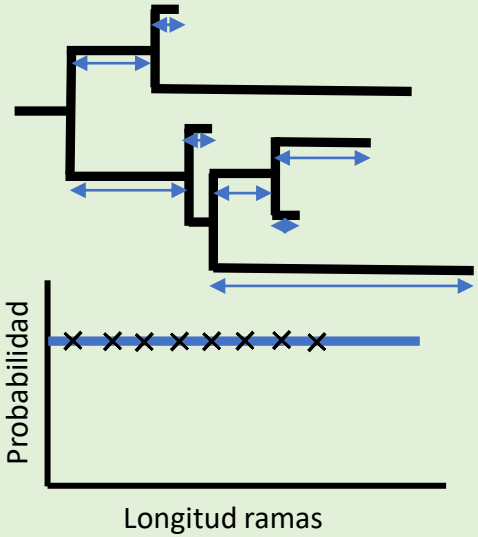
- La topología en sí
- La longitud de cada rama
- Ratio entre transiciones y transversiones
- Tasas de sustitución entre cada una de los posibles cambios nucleotídicos posibles

(…)

Es necesario aportar una expectativa de qué pinta van a tener, aunque sea muy aproximada

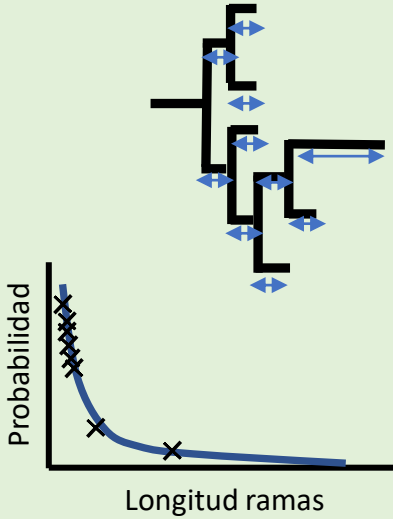
Ejemplo de selección de prior. Parámetro "**longitud de las ramas**". ¿Qué elegiríamos?

Todas las longitudes son equiprobables



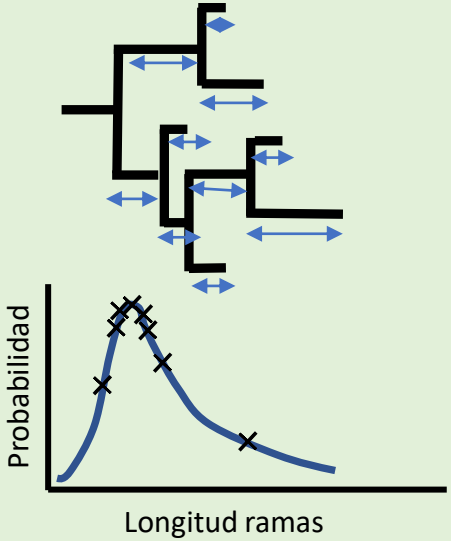
Distribución uniforme

Las ramas cortas son más probables



Distribución exponencial
(un tipo de distribución usado en probabilidad)

Las ramas de alrededor de cierto valor pequeño son más probables



Distribución gamma
(otro tipo de distribución usado en probabilidad)

Aquí entra nuestro criterio biológico
¿Qué es más razonable en el caso de los erizos?
(ve pensándolo)

Vamos a analizar nuestra matriz de datos 18S de erizos de mar usando el programa Mr Bayes, que funciona con línea de comandos bastante sencilla

Abre el programa y carga el alineamiento (asegúrate de que el archivo está en el la misma carpeta que el ejecutable, para simplificar el proceso)

```
>execute erizos.nex
```

MrBayes leerá el bloque de datos del archivo nexus y si todo está en orden quedará así

```
                            MrBayes 3.2.7a x86_64

                      (Bayesian Analysis of Phylogeny)

              Distributed under the GNU General Public License


                Type "help" or "help <command>" for information
                      on the commands that are available.

                    Type "about" for authorship and general
                       information about the program.


MrBayes > execute erizos.nex

   Executing file "erizos.nex"
   UNIX line termination
   Longest line length = 379
   Parsing file
   Expecting NEXUS formatted file
   Reading data block
      Allocated taxon set
      Allocated matrix
      Defining new matrix with 8 taxa and 350 characters
      Data is Dna
      Gaps coded as -
      Missing data coded as ?
      Taxon 1 -> Tripneustes_ventricosus
      Taxon 2 -> Sphaerechinus_granularia
      Taxon 3 -> Echinolampas_crassa
      Taxon 4 -> Echinodiscus_bisperforatus
      Taxon 5 -> Echinocardium_cordatum
      Taxon 6 -> Brissus_brissus
      Taxon 7 -> Diadema_setosum
      Taxon 8 -> Eucidaris_metularia
      Successfully read matrix
      Setting default partition (does not divide up characters)
      Setting model defaults
      Seed (for generating default start values) = 1643799075
      Setting output file names to "erizos.nex.run<i>.<p|t>"
   Exiting data block
   Reached end of file

MrBayes >
```

Para examinar la lista completa de comandos puedes usar el comando **help**

Además, puedes obtener información extra de cómo usar cada comando:

```
>help mcmc
```

Ahora, establezcamos el modelo de sustitución del análisis. Implementaremos el modelo que obtuvimos en la práctica anterior: K2+G

El comando **`lset`** es el que establece el modelo del análisis (utiliza el archivo “implementing models on MrBayes” del CV para ver cómo se implementa cada modelo)

```
>lset nst=2 rates=gamma
```

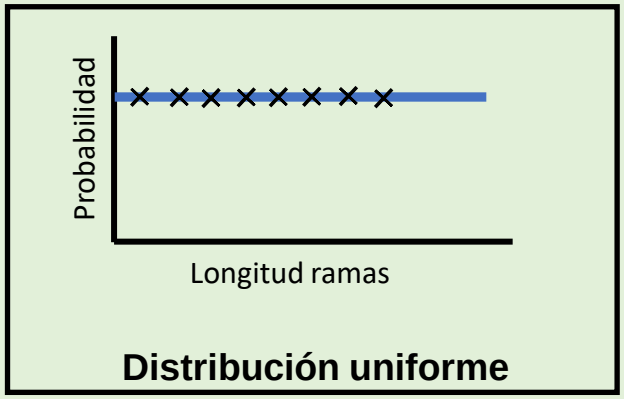
El K2 tiene dos tipo de tasa de sustitución (una para transversiones y otra para transiciones)

Para que el K2 sea “+G”, permitiremos cierta heterogeneidad siguiendo una distribución gamma

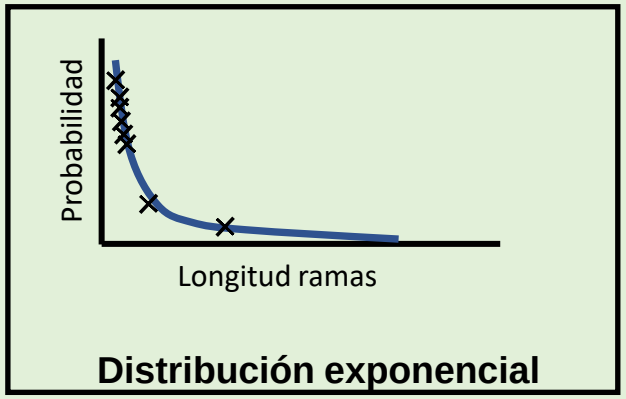
También podemos especificar dónde queremos enraizar el resultado del análisis

```
>outgroup Eucidaris_metularia
```

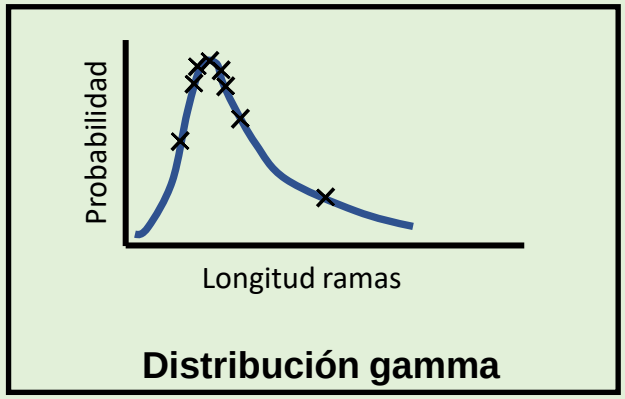
En cuanto al ***prior***, centrémonos en el parámetro en el que pensamos antes. La longitud de las ramas. ¿Tienes claro qué *prior* tiene mayor sentido biológico?



Distribución uniforme



Distribución exponencial



Distribución gamma

Estamos estudiando un grupo con nodos relativamente antiguos, no una radiación de especies recientes. Nos decidimos por una **distribución gamma**

```
>prset brlenspr = unconstrained:gammadir(1,0.1,1,1)
```

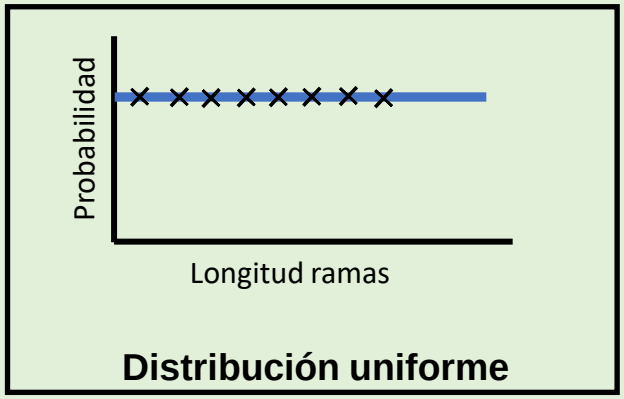
Comando para fijar los priors

De entre todos los parámetros posibles, en esta práctica nos vamos a centrar en longitud de ramas

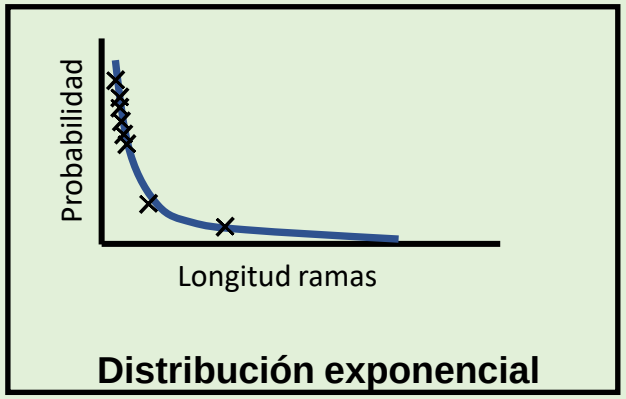
Desconocemos cuáles son los valores de longitud de ramas, pero “sospechamos” que seguirán una distribución gamma

Al igual que “sospechamos” que lloverá si tenemos algunos datos de contexto

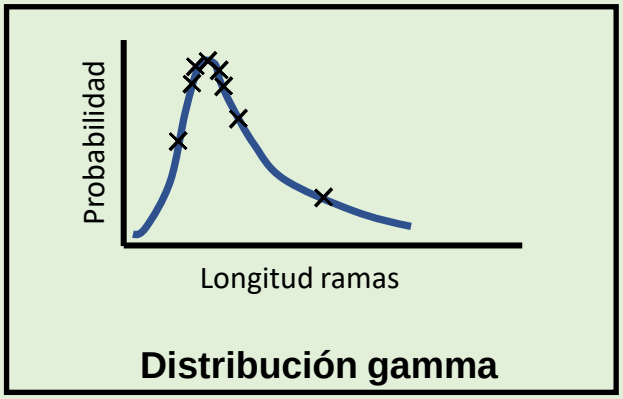
En cuanto al ***prior***, centrémonos en el parámetro en el que pensamos antes. La longitud de las ramas. ¿Tienes claro qué prior tiene mayor sentido biológico?



Distribución uniforme



Distribución exponencial



Distribución gamma

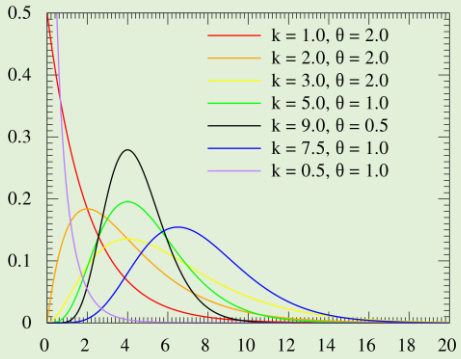
Estamos estudiando un grupo con nodos relativamente antiguos, no una radiación de especies recientes. Nos decidimos por una **distribución gamma**

```
>prset brlenspr = unconstrained:gammadir(1,0.1,1,1)
```

Comando para fijar los priors

Dist gamma

Estos valores modulan la forma de la curva. Usaremos de momento estos valores por defecto, basta con que sepas que cada distribución se podría afinar más si quisiéramos



De entre todos los parámetros posibles, en esta práctica nos vamos a centrar en longitud de ramas

Podemos repasar cómo ha quedado el modelo del análisis pidiéndole a MrBayes que nos lo enseñe con **>showmodel**

```
MrBayes > showmodel

   Model settings:

      Data not partitioned --
         Datatype  = DNA
         Nucmodel  = 4by4
         Nst       = 2
                     Transition and transversion  rates, expressed
                     as proportions of the rate sum, have a
                     Beta(1.00,1.00) prior
         Covarion  = No
         # States  = 4
                     State frequencies are fixed to be equal
         Rates     = Gamma
                     The distribution is approximated using 4 categories.
                     Shape parameter is exponentially
                     distributed with parameter (1.00).
```

Ahora pasemos a establecer las características del MCMC (comando **mcmcp**):

>mcmcp

Ngen=2000000 — MCMC se ejecutará durante dos millones de iteraciones o generaciones

samplefreq=1000
printfreq=1000 — No es necesario salvar todos los árboles, guardaremos uno de cada 1000, y en la pantalla saldrán los mismos

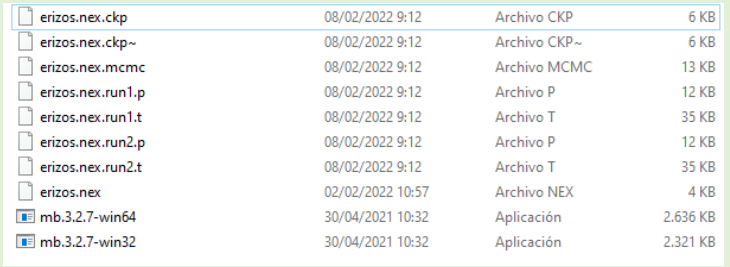
nruns=2
nchains=4 — El análisis constará de dos “carreras” en paralelo, cada una con 4 cadenas (una fría y tres calientes)

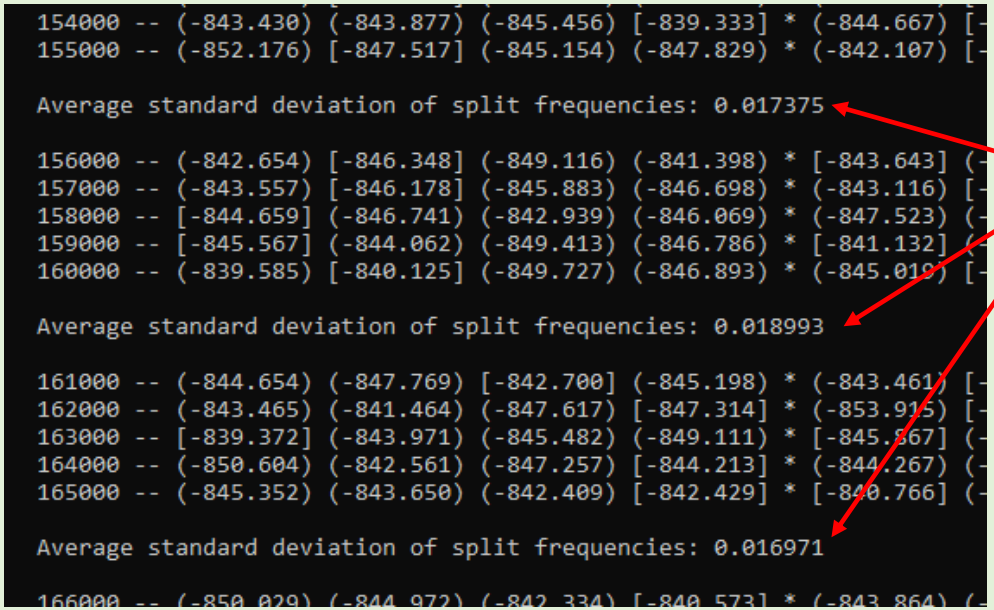
starttree=random — El punto inicial será un árbol completamente al azar

Cuando esté todo listo, ejecuta el análisis: **>mcmc**

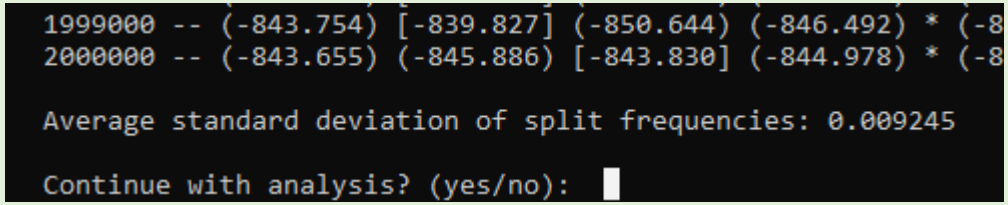
MrBayes ejecuta el mcmc. Tal y como le hemos pedido, reflejará en pantalla algunos datos de cada milésimo árbol, y los irá guardando en los archivos de output del análisis.

Verás un archivo mcmc, más un archivo “t” (trees) por carrera y un archivo “p” (parameters) por carrera





Este valor indica la divergencia entre las distintas cadenas. Cuanto menor sea, más cerca están de converger alrededor del mismo “pico”



Al alcanzar la iteración 2000000, MrBayes nos pregunta si deseamos continuar el análisis

```
1999000 -- (-843.754) [-839.827] (-850.644) (-846.492) * (-8
2000000 -- (-843.655) (-845.886) [-843.830] (-844.978) * (-8

Average standard deviation of split frequencies: 0.009245

Continue with analysis? (yes/no):
```

¿Qué responder?

```
Average standard deviation of split frequencies: 0.009774

1941000 -- (-846.704) (-848.743) (-844.956) [-842.980] * (-
1942000 -- (-848.638) (-847.685) [-849.988] (-849.437) * [-
1943000 -- (-844.323) (-844.714) [-848.920] (-846.670) * (-
1944000 -- (-849.246) (-842.938) (-844.401) [-845.326] * (-
1945000 -- (-852.363) [-847.896] (-851.812) (-853.416) * (-

Average standard deviation of split frequencies: 0.009686

1946000 -- (-842.941) [-844.020] (-845.394) (-845.247) * [-
1947000 -- (-844.353) (-851.706) [-841.348] (-843.903) * (-
1948000 -- (-860.182) (-843.705) [-844.256] (-845.260) * (-
1949000 -- (-846.580) (-848.224) [-844.797] (-843.234) * (-
1950000 -- (-850.515) (-848.562) (-846.836) [-844.430] * (-

Average standard deviation of split frequencies: 0.009539

1951000 -- (-844.939) (-845.918) (-846.231) [-844.718] * (-
1952000 -- (-854.344) (-841.599) [-842.117] (-847.289) * [-
1953000 -- (-845.054) [-843.136] (-847.995) (-850.457) * (-
1954000 -- [-842.535] (-851.986) (-843.068) (-843.636) * (-
1955000 -- (-842.297) (-841.279) (-852.619) (-842.843) * (-

Average standard deviation of split frequencies: 0.009459

1956000 -- (-840.717) [-843.287] (-847.584) (-847.378) * [-
1957000 -- (-844.293) (-842.090) (-848.051) [-846.806] * [-
1958000 -- (-841.943) (-845.551) (-842.458) [-845.731] * [-
1959000 -- (-845.009) (-840.638) (-845.701) [-844.699] * (-
1960000 -- (-842.553) [-845.335] (-841.401) (-845.210) * [-

Average standard deviation of split frequencies: 0.009314

1961000 -- [-848.181] (-844.958) (-842.780) (-842.609) * [-
1962000 -- [-843.367] (-844.876) (-841.707) (-848.766) * (-
1963000 -- (-849.437) [-842.842] (-840.004) (-849.178) * (-
1964000 -- (-849.007) [-851.635] (-858.412) (-844.127) * (-
1965000 -- (-847.308) [-842.802] (-847.522) (-844.883) * (-

Average standard deviation of split frequencies: 0.009228

1966000 -- [-843.221] (-852.680) (-856.417) (-844.573) * [-
1967000 -- (-844.817) (-840.952) (-842.415) [-841.427] * (-
1968000 -- (-845.473) (-847.722) [-848.921] (-849.203) * (-
1969000 -- [-844.028] (-849.325) (-845.334) (-852.006) * (-
1970000 -- (-842.589) (-844.715) (-841.191) [-845.251] * (-

Average standard deviation of split frequencies: 0.009861

1971000 -- [-843.479] (-843.477) (-848.059) (-844.519) * [-
1972000 -- (-847.622) (-846.696) (-845.596) [-849.382] * (-
1973000 -- (-851.583) (-841.100) (-841.467) [-841.766] * [-
1974000 -- [-846.970] (-844.723) (-846.556) (-844.524) * (-
```

En general, podemos decidir que el análisis ha alcanzado cierta convergencia si los valores de frecuencia se mantienen por debajo de 0.01 durante las últimas 10000 generaciones.

Si consideras que las cadenas aún no han convergido lo suficiente, puedes pedirle a MrBayes que continúe, diciéndole cuántas generaciones más debe realizar.

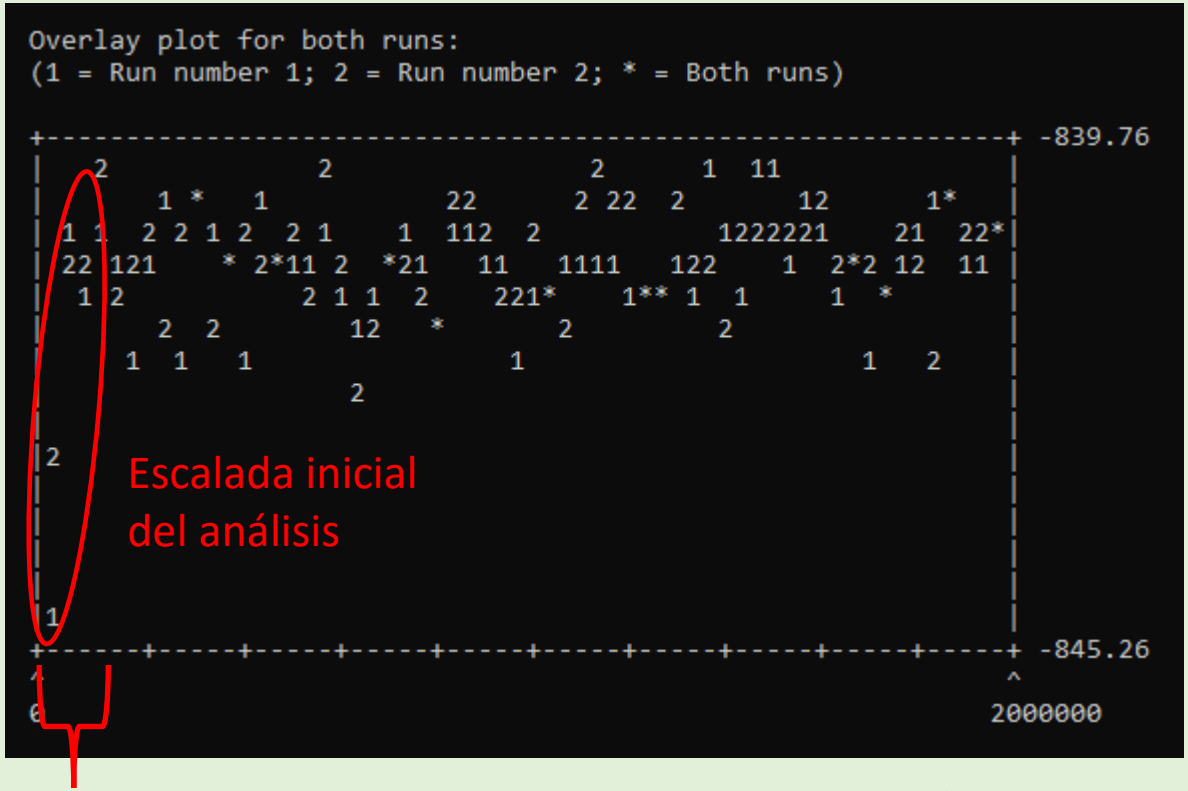
```
  2000000 -- (-846.445) (-849.052) [-845.186] (-841.695) * (-845.531) [-842.25
(-843.504) (-847.067) -- 0:00:00

  Average standard deviation of split frequencies: 0.003533

  Continue with analysis? (yes/no): y
     Additional number of generations: 10000 (las que quieras)
```

Por el contrario, para finalizar el análisis, responde “no”

Desde MrBayes hay una opción rápida con un comando que nos da un resumen de la evolución de los parámetros más importantes del análisis, en concreto de la verosimilitud:

`>sump relburnin=no`

El eje X es el número de generaciones, el eje Y el logaritmo de la P(D/H) -verosimilitud-. 1 y 2 son los valores que han dado las cadenas frías de cada carrera (* cuando coinciden)



Decidimos deshacernos del primer 10% de los árboles

Aquí lo importante es verificar que no hay una tendencia marcada, sino que el recorrido de las carreras oscila sin patrón

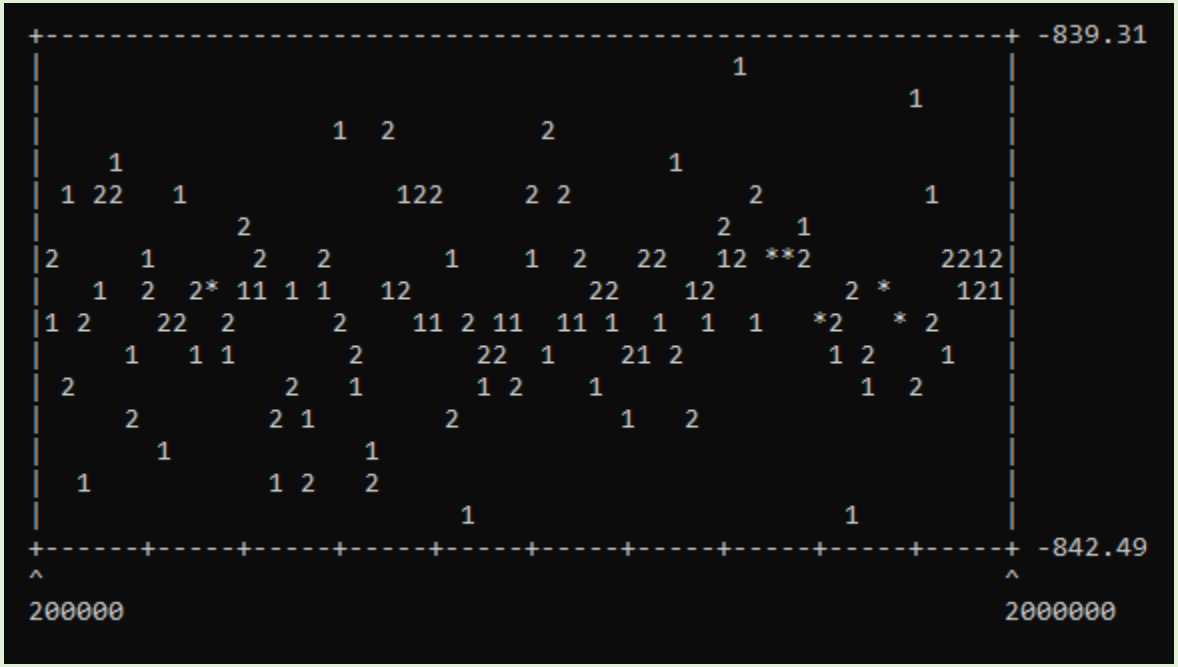
Es posible que observes que las generaciones iniciales sí que muestran una tendencia (la primera “escalada” desde un árbol aleatorio a zonas con picos de verosimilitud)

Estima de qué cantidad de árboles vas a querer desprenderte

Para corregirlo, empleamos el comando **burnin**

```
>sump relburnin=no burnin=200
```

Si hemos “guardado” uno de cada mil árboles hay salvados 2000 tras los dos millones de generaciones. Con el comando burnin nos deshacemos del 10% inicial



Visualizamos que ahora el resumen de las carreras no muestra ninguna tendencia, así que este valor de burnin nos vale

Calcula el árbol consenso de los árboles que has almacenado en la memoria (en nuestro caso, uno de cada 1000, o sea, 2000 en total). Recuerda además que queremos deshacernos del 10% inicial de los árboles

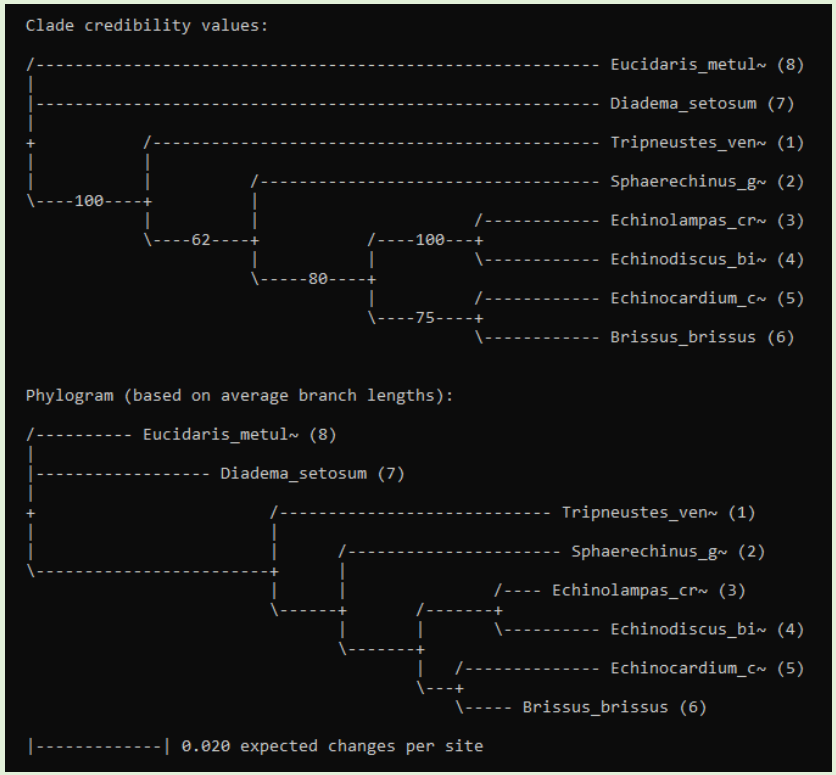
```
>sumt contype=halfcompat burnin=200
```

Sumt pide calcular un árbol

Que sea de consenso

De la mayoría (50%)

Ignorando los 200 árboles iniciales salvados en los archivos ".t" (el 10% de 2000)



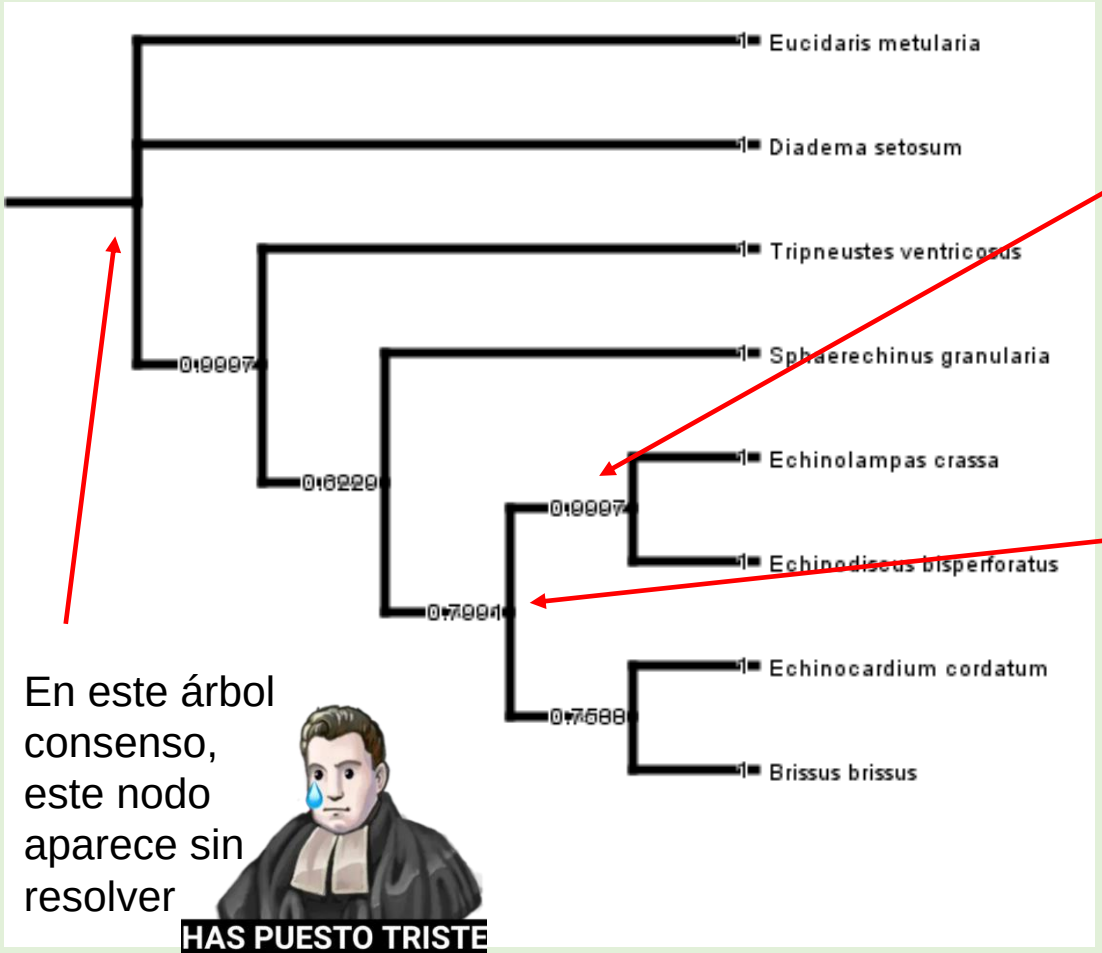
El programa nos dará el árbol consenso:

- Primero como cladograma con los valores de apoyo de cada rama (son probabilidades posteriores)
- Luego como filograma con una aproximación de las longitudes de cada rama

Más importante todavía: lo salvará en un archivo ".con"

Este árbol consenso lo podemos visualizar en Mesquite:
File>Open Other>Special NEXUS (y se selecciona el árbol consenso)
Después selecciona **Import MrBayes Consensus Tree File**

Para mostrar las probabilidades posteriores de cada nodo:
Display>Node-Associated Values>Choose Values to Show (elije "prob")

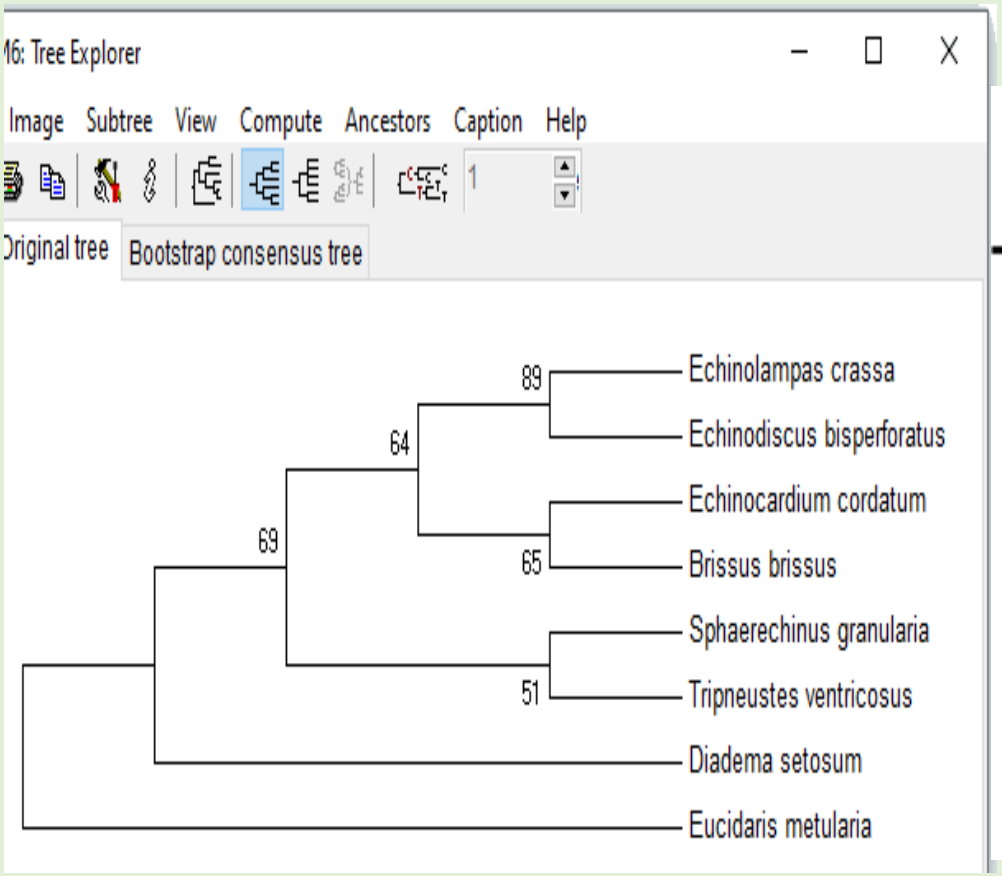


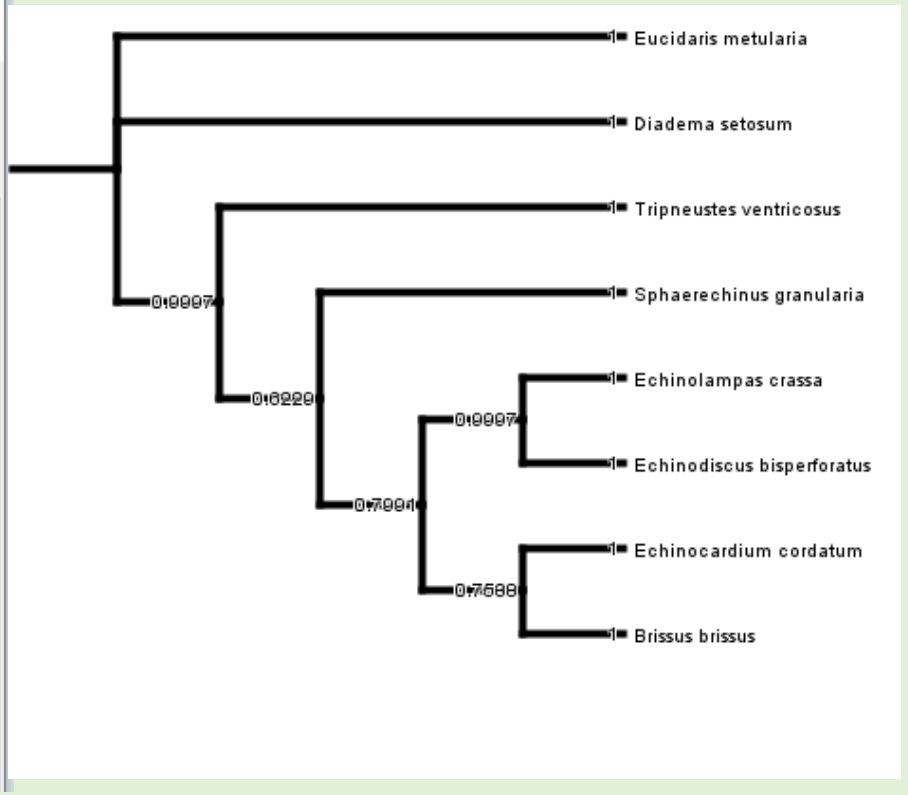
Los valores de probabilidad posterior superiores a 0.95 se consideran una certidumbre elevada

Cuanto menor sea la PP, menor certidumbre bayesiana

En este árbol consenso, este nodo aparece sin resolver

Ahora puede ser un buen momento para comparar los resultados obtenidos por MV y por IB





¿Qué nodos tienen en común y en cuáles se diferencian?

En aquellos que son diferentes, ¿Cuáles parecen tener un apoyo más robusto?

¿Qué valor añadido proporciona resolver árboles con más de un criterio?

Por último, a nivel metodológico, ¿Puedes resumir los requisitos y fundamentos de los tres criterios explorados hasta ahora? (MP, MV, IB)

	¿Cómo se define al "árbol óptimo"?	Elementos necesarios para realizar análisis	Ventajas	Debilidades
Máxima Parsimonia con MEGA				
Máxima verosimilitud con MEGA				
Inferencia Bayesiana con MrBayes				

Por último, a nivel metodológico, ¿Puedes resumir los requisitos y fundamentos de los tres criterios explorados hasta ahora? (MP, MV, IB)

	¿Cómo se define al "árbol óptimo"?	Elementos necesarios para realizar análisis	Ventajas	Debilidades
Máxima Parsimonia con MEGA	El de menor longitud (menor número de cambios de carácter)	Basta con la matriz de datos	-El más rápido de computar -No requiere modelo de sustitución	-Planteamiento reduccionista -Muy susceptible a la saturación (atracción de ramas largas) - Requiere bootstrap
Máxima verosimilitud con MEGA	El que maximiza la probabilidad de observar los datos si asumimos que es verdadero	-Matriz de datos -Modelo de sustitución	- Evita el problema de atracción de ramas largas - Con sentido biológico al usar modelo	- Muy costoso de computar - Susceptible a máximos locales - Requiere bootstrap
Inferencia Bayesiana con MrBayes	El que tiene una mayor probabilidad de ser verdadero dados nuestros datos	- Matriz de datos - Modelo de sustitución - "Prior"	- Las mismas que MV - Gracias al MCMC se evitan máximos locales y su computación es eficiente - No requiere bootstrap	- Aunque es eficaz, sigue siendo costoso de computar - Es el que más elementos necesita (modelo y prior)

Ten en cuenta que has aprendido a usar algunos programas introductorios. Hoy en día existe una gama de herramientas filogenéticas muy amplia y muchas de las limitaciones de MEGA o MrBayes que aparecen en este cuadro están superadas