Proyecto de tesis Valdivia, diciembre de 2024

**Historia evolutiva de las señales de comunicación acústica de las aves modernas: Integración de métodos filogenéticos y bioacústicos para la reconstrucción de estados ancestrales.**

Gabriel Morales Tello

Doctorado en Ciencias Mención Ecología y Evolución

Universidad Austral de Chile

1. **Material Suplementario**

**Simulación 01**

Mediante las librerías *R* *phytools* y *geiger* se realizó una simulación de una reconstrucción ancestral del carácter discreto «tipo de siringe» categorizado en cuatro estados:

1. Siringe traqueal
2. Siringe bronquial
3. Siringe traqueo-bronquial
4. Siringe bronquio-traqueal

**Paso 1:** Simular una filogenia de 16 especies con topología aleatoria y de nacimiento puro y asignar aleatoriamente un estado del carácter a cada terminal *Sp*:



**Figura S1.** Simulación de una filogenia de 16 especies hipotéticas con topología de árbol aleatoria y de nacimiento puro. A la vez, se asignan aleatoriamente los estados de carácter.

**Paso 3:** Ajustar 3 modelos *Mk* mediante máxima verosimilitud con la función *fitMk* de *phytools*. A continuación se describen los modelos escogidos:

1. Modelo *Equal Rates (ER):* El modelo de tasas iguales *ER* es una variante del modelo Mk y se ajusta a una única tasa de transición entre todos los estados para nuestro caracter discreto (Revell & Harmon, 2022).
2. Modelo *Symmetric Transition (SYM):* El modelo de tasas simétricas *SYM* asume que la tasa de transición de cada estado de carácter *i* a cada estado *j* es igual a la tasa de cambio del estado *j* al estado *i*, pero que cada par de estados de carácter puede tener una tasa diferente (Revell & Harmon, 2022).
3. Modelo *All Rates Different (ARD):* El modelo «todas las tasas diferentes» *ARD* es la variante *Mk* más compleja al permitir que cada tipo de transición entre estados ocurra con una tasa diferente (Revell & Harmon, 2022).

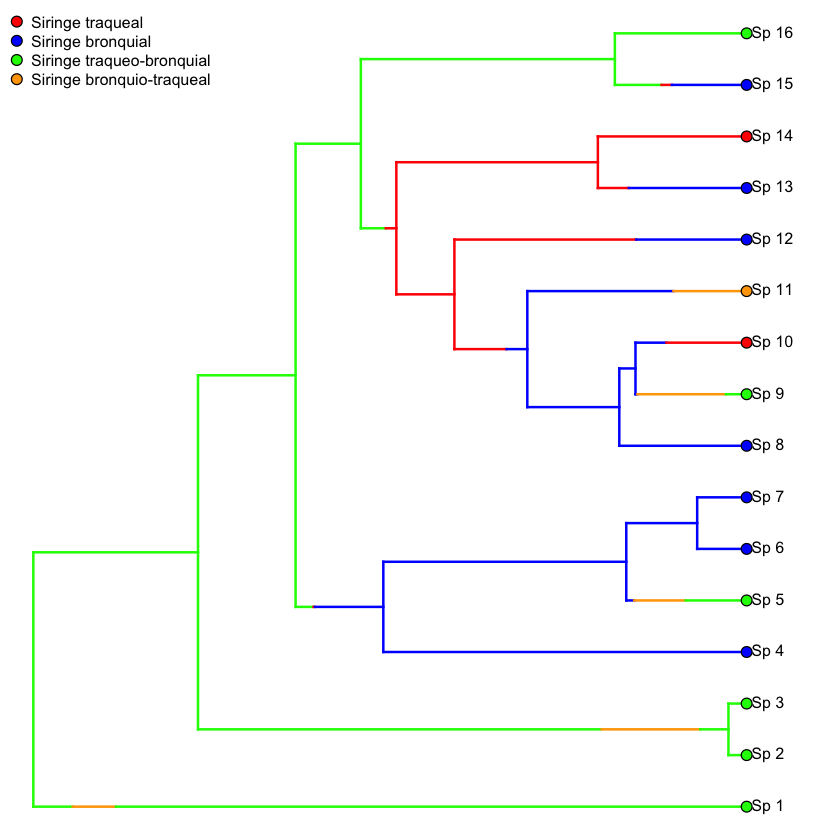
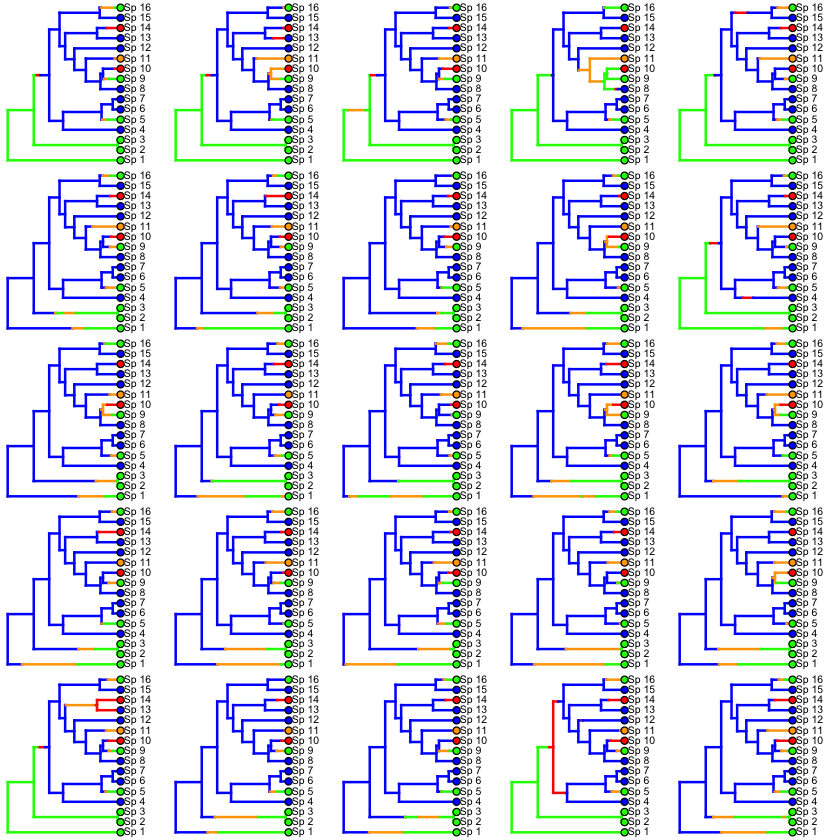
**Paso 4:** Comparar los modelos utilizando *ANOVA*.

**Tabla S1.** La tabla proporciona los nombres de cada uno de los modelos ajustados, sus verosimilitudes logarítmicas (log(L)), el número de parámetros estimados (d.f), un valor del criterio de información de Akaike (*AIC*) y las ponderaciones o «pesos» del modelo. Valores más pequeños de *AIC* indican un mejor soporte considerando la complejidad de sus parámetros (Revell, 2024).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | log(L) | d.f. | AIC | weight |
| ***ER*** | **-20.7762** | **1** | **43.5525** | **0.1285029** |
| *ARD* | -17.6850 | 12 | 59.3700 | 0.0000472 |
| *SYM* | -19.5703 | 6 | 51.1407 | 0.0028916 |

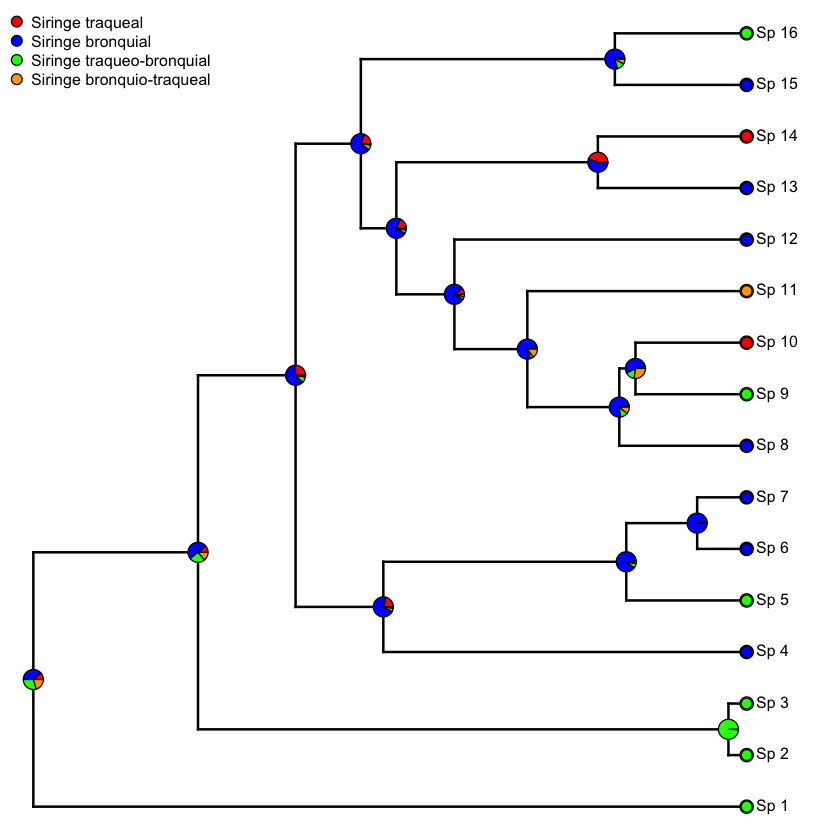
Según el análisis *ANOVA*, el modelo *ER* es el mejor respaldado. Sin embargo, podría haber ocurrido un resultado más equilibrado con similitud de puntajes. Con esto en mente y a modo de ejercicio, realizaremos el siguiente paso considerando todos los modelos y sus pesos.

**Paso 5:** Generar 1000 mapas estocásticos («*entre 100 y 1000 pueden ser suficientes*» (Revell, 2024)) con el método *simmap* de *phytools.* El mapeo implica muestrear aleatoriamente historias de caracteres que son consistentes con los datos en las terminales. Los mapas son creados bajo cada uno de los tres modelos en proporción relativa al peso que respalda a cada modelo.



**Figura S2.** (Izquierda): Muestra aleatoria de 25 mapas estocásticos. (Derecha): Muestra de uno de los árboles con la menor cantidad de cambios (criterio de Máxima Parsimonia).

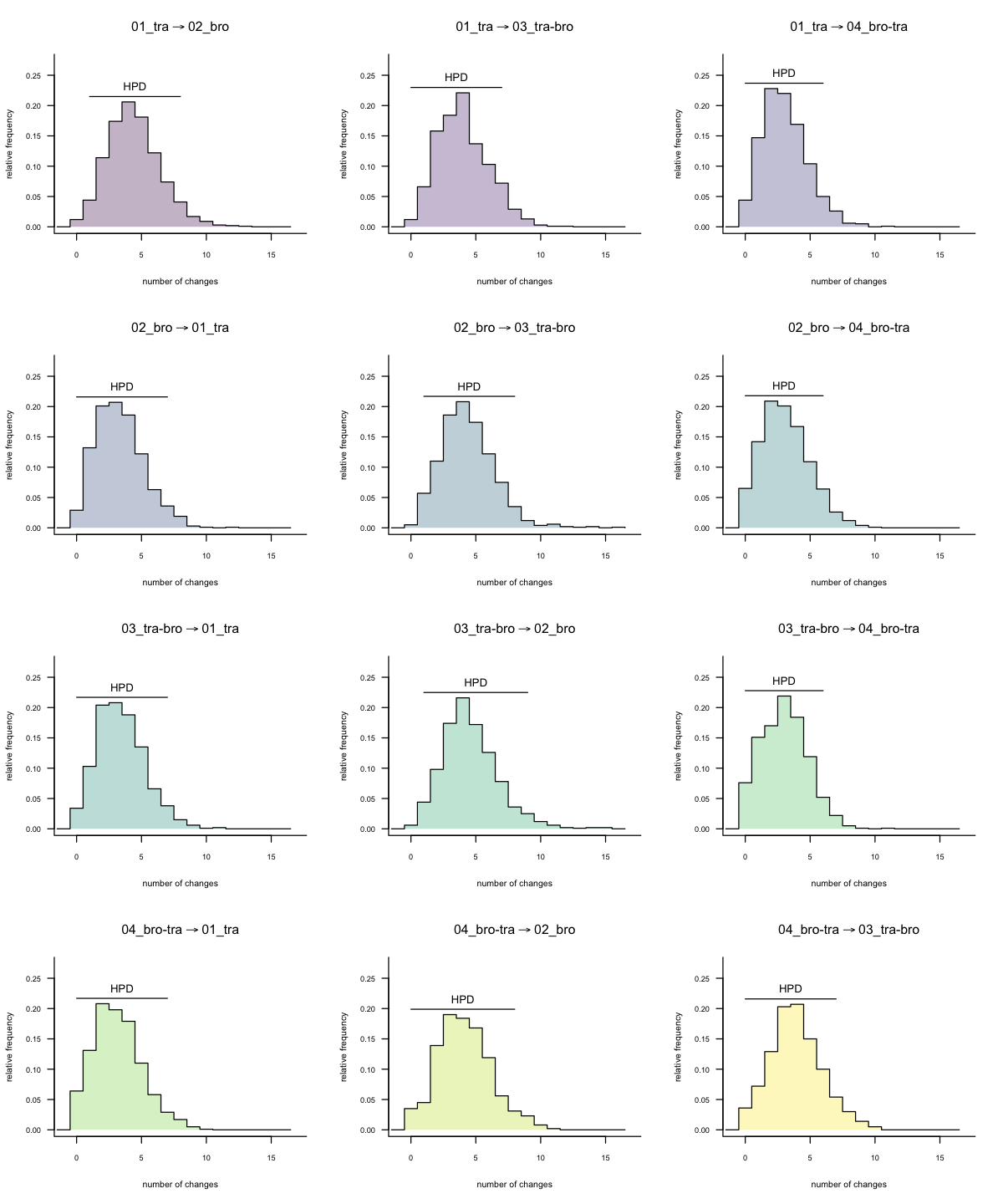
**Paso 6.** Calcular, utilizando el método de resumen genérico *summary,* las probabilidades posteriores de cada estado de carácter en cada nodo interno y a lo largo de las ramas del árbol.



**Figura S3.** Probabilidades posteriores en cada nodo ancestral del árbol de la Fig. S1 a partir del mapeo de caracteres estocásticos utilizando pesos de modelo para muestrear tres modelos Mk.

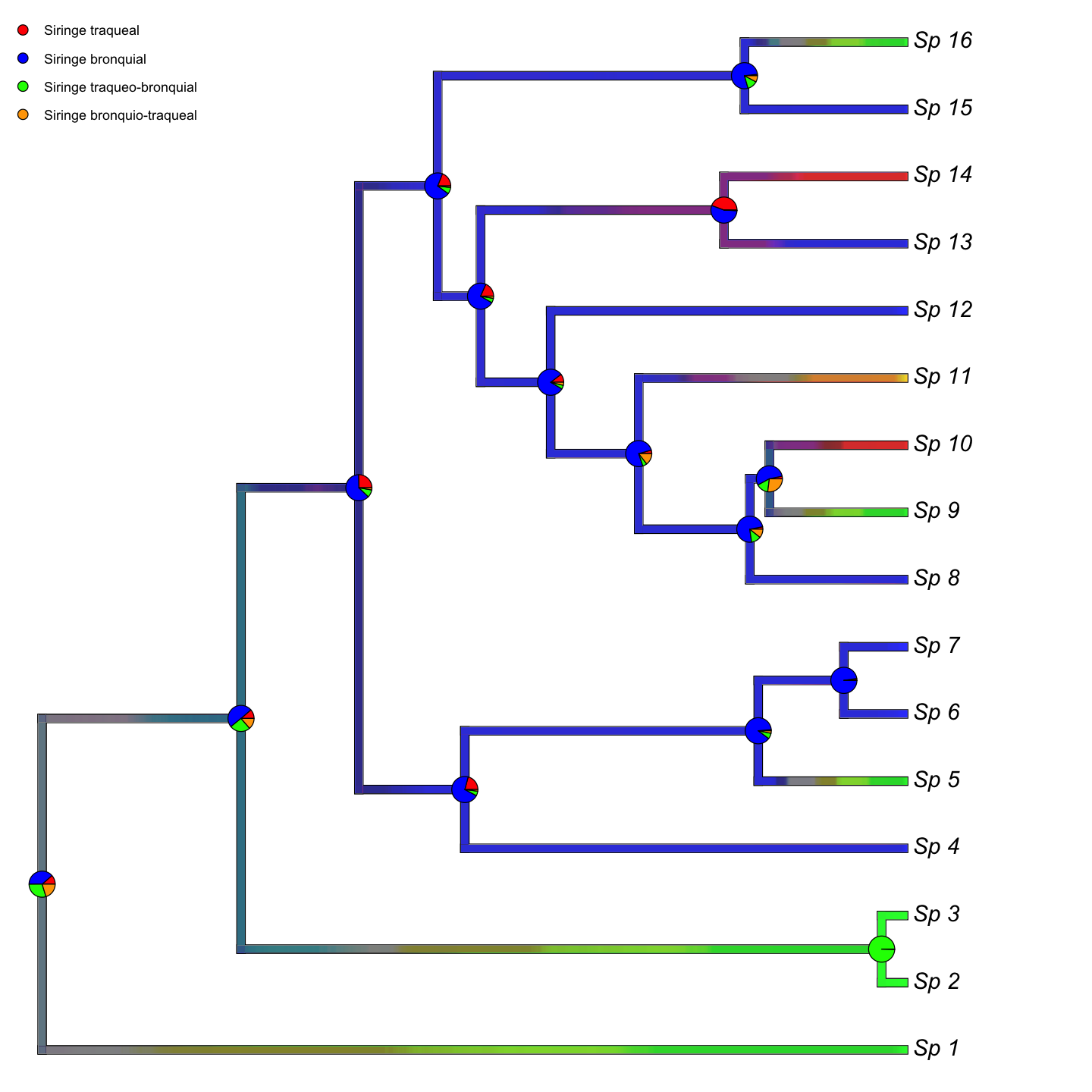
Las probabilidades posteriores ilustradas en los nodos de la filogenia de la Fig. S3, corresponden a una forma de reconstrucción del estado ancestral. En otras palabras, estos valores representan la probabilidad de que cada ancestro hipotético, en cada nodo, estuviera en cada uno de los cuatro estados conocidos del caracter.

**Paso 7:** Calcular la distribución de frecuencias relativas de las transiciones de cada tipo en el conjunto de historias mapeadas y a lo largo de todo el árbol, así como los intervalos bayesianos de densidad de alta probabilidad (HPD) del 95%. Esto se hace con la función *density.*



**Figura S4.** Distribuciones de probabilidad posterior de tipos de transición entre los estados del carácter a lo largo de todo el árbol, obtenidas a partir de un análisis de mapeo estocástico. HPD indica el intervalo de densidad de alta probabilidad del 95% para cada tipo de transición.

**Paso 8:** Visualizar las probabilidades posteriores de cada uno de los estados a lo largo de las ramas de la filogenia. Esto se logra graficando simultáneamente todos los mapas estocásticos utilizando colores transparentes. De esta manera es posible visualizar la incertidumbre en la reconstrucción del estado ancestral en un carácter multiestado.

**

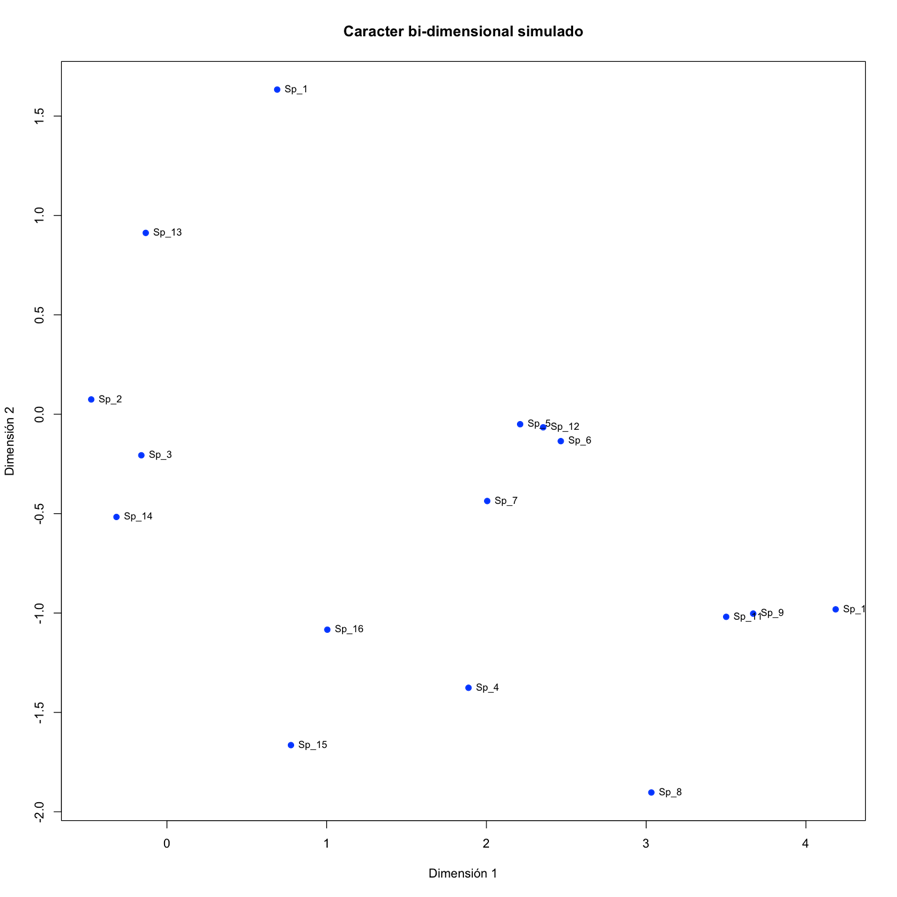
**Figura S5.** Simulación de la incertidumbre en la reconstrucción ancestral de los cuatro estados imprimiendo los 1000 mapas estocásticos y colores transparentes. (técnica original publicada en el [blog de phytools](http://blog.phytools.org/2020/06/mapping-multi-state-discrete-character.html), (Liam Revell, 2020)).

**Simulación 02**

Mediante las librerías *R* *phytools* se realizó una simulación de un análisis de evolución de caracteres continuos multidimensionales respecto a restricciones del tipo de siringe:

**Paso 1:** Escoger una filogenia previamente mapeada con restricciones del carácter multiestado «tipo de siringe». Para simplicidad de esta simulación se escogió uno de los mapas estocásticos con la menor cantidad de cambios ([Figura S2 (derecha)](#FigS2der)). Cade recordar que este mapa es obtenido a partir de una filogenia y datos generados aleatoriamente.

**Paso 2:** Simular datos de caracteres continuos según movimiento browniano, similar a lo que sería una extracción de características de alto nivel, reducidas en un espacio *UMAP* de dos dimensiones.



|  |  |  |
| --- | --- | --- |
|  | **Dimensión 1** | **Dimensión 2** |
| **Sp\_1** | 0.69013177539786 | 1.63328767997123 |
| **Sp\_2** | -0.474030428609983 | 0.0743433733013926 |
| **Sp\_3** | -0.160306988672232 | -0.20656700749657 |
| **Sp\_4** | 1.88771978003635 | -1.37618613274501 |
| **Sp\_5** | 2.2110391872523 | -0.0502882484923551 |
| **Sp\_6** | 2.46569662422801 | -0.135328317399678 |
| **Sp\_7** | 2.00495924547358 | -0.436557765227439 |
| **Sp\_8** | 3.032531225208 | -1.90287465243649 |
| **Sp\_9** | 3.67008522208175 | -1.00297194471207 |
| **Sp\_10** | 4.18628652668175 | -0.981545183690777 |
| **Sp\_11** | 3.50102028077165 | -1.01910939910399 |
| **Sp\_12** | 2.35528214661549 | -0.0658552104206803 |
| **Sp\_13** | -0.132787320740133 | 0.912240181412787 |
| **Sp\_14** | -0.316194498942599 | -0.516784759046936 |
| **Sp\_15** | 0.776342194202296 | -1.66464872830956 |
| **Sp\_16** | 1.00390452615405 | -1.08380106360011 |

**Figura S6.** Tabla y gráfico de los datos generados aleatoriamente según movimiento browniano para un carácter bi-dimensional hipotético de 16 especies.

**Paso 3:** Ajustar modelos con *evolvc.lite* de la librería *phytools*.

La función *evolvcv.lite* ajusta un conjunto jerárquico de modelos para las tasas evolutivas (de cada dimensión) y las correlaciones evolutivas entre ellas, respecto a un carácter discreto previamente mapeado en la filogenia. Estos modelos son: (1) tasas y correlaciones comunes; (2) tasas diferentes y correlación común; (3) tasas comunes y correlación diferente; y (4) tasas y correlaciones diferentes.

Resultados de este análisis:

Fitting model 1: common rates, common correlation...

Best log(L) from model 1: -39.8424.

Fitting model 2: different rates, common correlation...

Best log(L) from model 2: -35.7058.

Fitting model 3: common rates, different correlation...

Best log(L) from model 3: -37.9525.

Fitting model 4: no common structure...

Best log(L) from model 4: -33.3078.

**Paso 4:** Comparar los modelos utilizando *ANOVA*.

**Tabla S2.** La tabla proporciona los nombres de cada uno de los modelos ajustados, sus verosimilitudes logarítmicas (log(L)), el número de parámetros estimados (d.f), un valor del criterio de información de Akaike (*AIC*) y las ponderaciones o «pesos» del modelo. Valores más pequeños de *AIC* indican un mejor soporte considerando la complejidad de sus parámetros (Revell, 2024).

log(L) d.f. AIC weight

**model 1 -39.84240 5 89.68479 0.63708689**

model 2 -35.70579 11 93.41159 0.09884046

model 3 -37.95249 8 91.90498 0.20993765

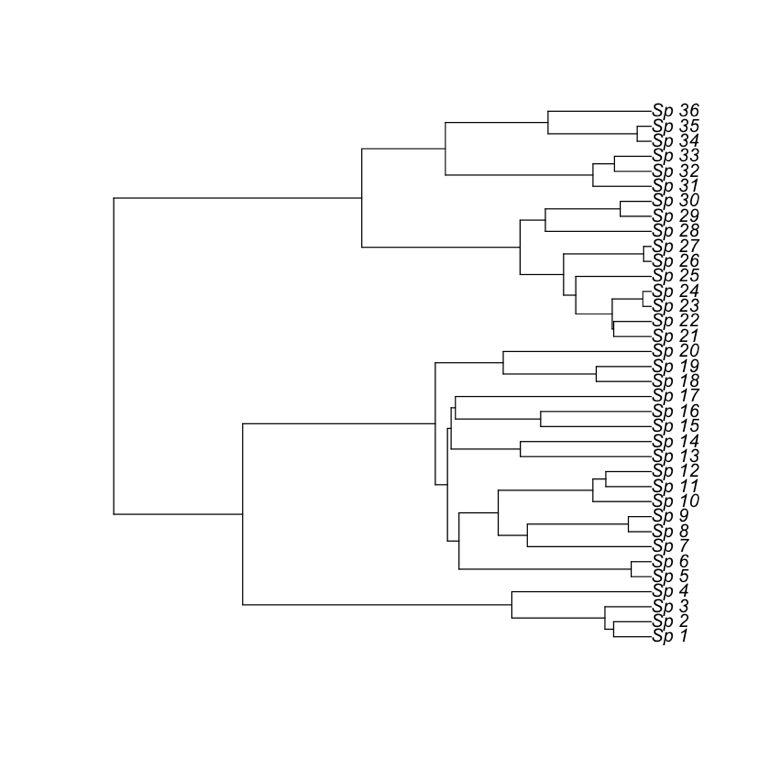
model 4 -33.30782 14 94.61564 0.05413500

Según el análisis *ANOVA*, el modelo 1 es el mejor respaldado. En un eventual análisis con datos reales este resultado le quitaría apoyo a nuestra hipótesis, puesto que la evolución de nuestro carácter continuo bi-dimensional no se vería afectado por el tipo de siringe en cuanto a tasas de evolución y cambios en la correlación entre las dimensiones del carácter.

**Simulación 03**

Utilizando *phytools*, se llevó a cabo una simulación de un análisis de tasas de evolución y correlación de caracteres continuos de bajo nivel.

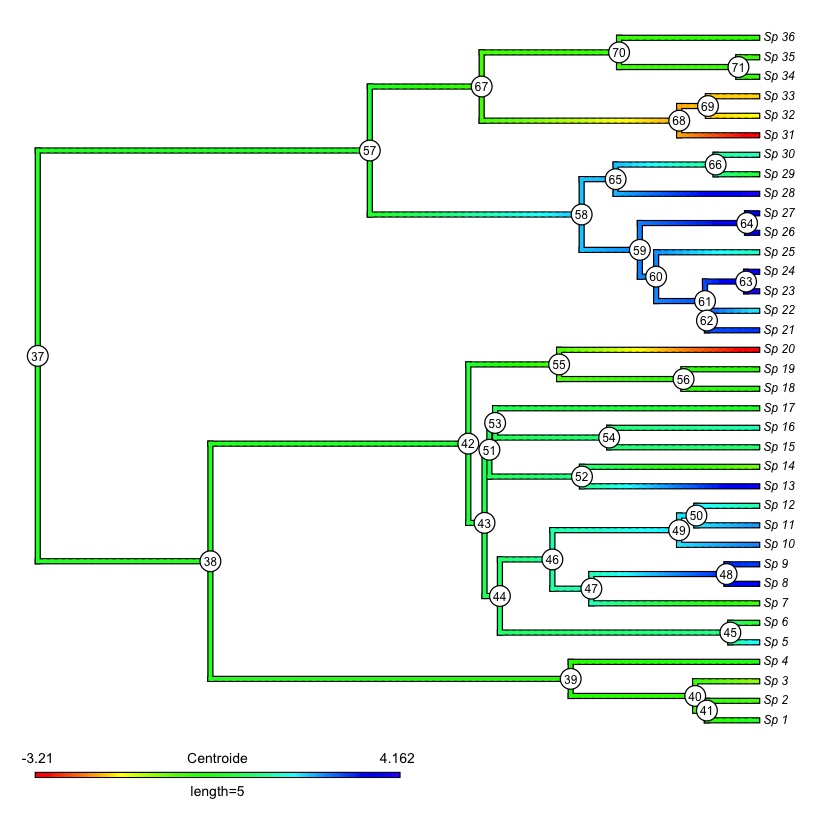
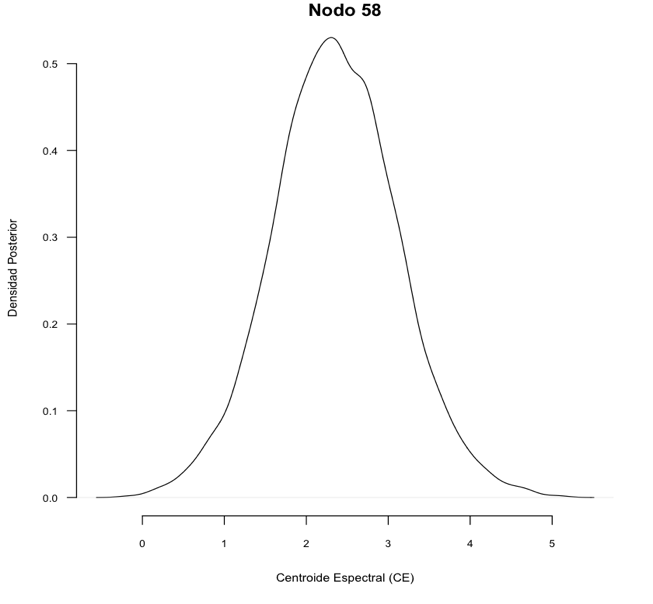
**Paso 1:** Teniendo en mente a los tinamúes (no aprendices vocales) e ictéridos (aprendices vocales) se generó una simulación de filogenias de tamaños proporcionales a estas dos familias de aves:



**Figura S7.** Filogenias de nacimiento puro, generadas de forma aleatoria y con la misma calibración temporal relativa. (Izquierda): Filogenia de 36 especies hipotéticas de un clado de aves aprendices vocales. (Derecha): Filogenia de 12 especies hipotéticas de un clado de aves no aprendices vocales.

**Paso 2:** Se simularon dos conjuntos de datos acústicos para los 5 caracteres de bajo nivel (), teniendo en cuenta el árbol y un modelo evolutivo de movimiento browniano.

**Paso 3:** Con la función *anc.Bayes,* se ejecutó una reconstrucción de estados ancestrales para cada carácter en ambas filogenias, bajo un movimiento browniano. Esta función utiliza *MCMC* bayesiana para tomar muestras de la distribución posterior de los estados en los nodos internos del árbol. Se ajustó un total de 500 mil generaciones y un 20% de *burn-in* para garantizar la convergencia.

**Figura S8.** (Izquierda): Reconstrucción de estados ancestrales del carácter continuo Centroide Espectral () en la filogenia simulada de un grupo de aves aprendices vocales. (Derecha): A modo de ejemplo se muestra la densidad de probabilidad posterior en uno de los nodos: el nodo 58.

**Paso 4:** Con la función *summary* se calculó la media de la probabilidad posterior de cada nodo reconstruido

**Paso 5:** Comparar tasas de evolución entre caracteres inter e intra filogenias.

Se combinaron los valores simulados en los nodos terminales y los valores reconstruidos en los nodos internos para cada carácter y para cada filogenia. Luego calculamos la varianza (normalizada por la altura total del árbol, es decir la suma de longitudes de rama) como una estimación de la tasa de evolución para cada carácter y para cada filogenia. Para comparar las tasas de cada carácter entre filogenias calculamos la razón de varianzas entre los resultados para cada filogenia, con el objetivo de identificar si un caracter evoluciona más rápido en aprendices o no aprendices.

**Tabla S3**. Varianzas normalizadas de cada carácter en aprendices y no aprendices vocales, así como la razón entre ellos. Un valor > 1 indica que el carácter evoluciona más rápido en aprendices, mientras que un valor < 1 indica que evoluciona más rápido en no aprendices.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Varianza Normalizada Aprendices** | **Varianza Normalizada No Aprendices** | **Razón Normalizada de Varianzas entre Aprendices y No Aprendices** |
| **Centroide Espectral** | 0.033 | 0.037 | 0.884 |
| **Entropía Espectral** | 0.037 | 0.048 | 0.774 |
| **Coeficiente de Tonalidad** | 0.040 | 0.049 | 0.819 |
| **RMS** | 0.103 | 0.069 | 1.490 |
| **Flujo Espectral** | 0.130 | 0.032 | 3.956 |

**Paso 6:** Evaluar la correlación de cambios evolutivos

Calculamos los cambios por rama para cada carácter, que son las diferencias absolutas entre los valores reconstruidos en los nodos adyacentes. Luego calculamos las correlaciones mediante el *coeficiente de correlación de Pearson*, entre estos vectores de cambio para cada par de caracteres.

**Tabla S4**. Correlaciones de cambio entre caracteres en un clado hipotético de aprendices vocales.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Centroide Espectral** | **Entropía Espectral** | **Coeficiente de Tonalidad** | **RMS** | **Flujo Espectral** |
| **Centroide Espectral** | 1 | 0.279 | 0.220 | 0.444 | 0.261 |
| **Entropía Espectral** | 0.279 | 1 | 0.212 | 0.388 | 0.207 |
| **Coeficiente de Tonalidad** | 0.220 | 0.212 | 1 | 0.296 | 0.133 |
| **RMS** | 0.444 | 0.388 | 0.296 | 1 | 0.426 |
| **Flujo Espectral** | 0.261 | 0.207 | 0.133 | 0.426 | 1 |

**Tabla S4**. Correlaciones de cambio entre caracteres en un clado hipotético de no aprendices vocales.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Centroide Espectral** | **Entropía Espectral** | **Coeficiente de Tonalidad** | **RMS** | **Flujo Espectral** |
| **Centroide Espectral** | 1 | **0.670** | -0.084 | 0.281 | 0.338 |
| **Entropía Espectral** | **0.670** | 1 | 0.021 | 0.341 | 0.046 |
| **Coeficiente de Tonalidad** | -0.084 | 0.021 | 1 | **0.715** | -0.016 |
| **RMS** | 0.281 | 0.341 | **0.715** | 1 | 0.0007 |
| **Flujo Espectral** | 0.338 | 0.046 | -0.016 | 0.0007 | 1 |