

CARTILLA DE PRÁCTICOS

Curso de Evolución 2022

Laboratorio de Evolución. Facultad de Ciencias.
Universidad de la República.

Práctico 1

Filogenias I

Introducción

Los chimpancé, *Pan troglodytes*, de la región centro-oeste del continente Africano son reconocidos como un reservorio de virus de inmunodeficiencia en Simios (SIVcpzPtt), los cuales han cruzado la barrera específica en al menos dos oportunidades, resultando en la pandemia provocada por el Síndrome de Inmuno Deficiencia Adquirida (HIV-1, grupo M) y por otro lado en la infección aislada de unos pocos individuos en Camerún (HIV-1, grupo N). Un tercer linaje de virus HIV-1 (grupo O), también de la región centro-oeste de África, cae igualmente dentro de la radiación de los virus de tipo «SIVcpzPtt». Más de 30 especies de primates son portadores de virus que provocan inmunodeficiencia en Simios, pero los chimpancés son los principales portadores de los tipos cercanamente emparentados al HIV-1. Con la finalidad de establecer el origen de la cepa HIV-1 (grupo O) se secuenciaron algunos genes de varias muestras de chimpancés (SIVcpz) y gorilas (*Gorilla gorilla*, SIVgor) de Camerún. El objetivo de esta actividad consiste, mediante un análisis filogenético de secuencias de ADN, investigar el posible origen y relacionamiento de las diferentes cepas de HIV-1 presentes en chimpancés, gorilas y humanos.

Datos y programa de análisis

La base de datos a analizar consiste en secuencias de los genes *env* –que codifica proteínas de la envoltura– y *pol* –que codifica la transcriptasa inversa– del virus HIV-1. Las secuencias ya se encuentran alineadas (las homologías posicionales entre las distintas secuencias ya están establecidas), por lo que ya están listas para ser analizadas.

El programa que se usará para generar las hipótesis filogenéticas es el MEGA11. Éste es un programa que se baja gratis de la red en <http://www.megasoftware.net> y que tiene varias prestaciones, incluyendo el estudio descriptivo de las secuencias y reconstrucciones filogenéticas mediante métodos basados en distancias genéticas, máxima parsimonia y máxima verosimilitud.

Actividades a realizar durante el práctico

1) Abrir el archivo “HIV.meg” en MEGA11 (“File/Open A File/Session”). Visualizar la matriz de datos (“Data/Explore Active Data”) e identificar los primeros 5 sitios variables (botón “V”) y los primeros 3 informativos (botón “Pi”). ¿Por qué algunos de los sitios variables no son informativos?

2) Ir a la ventana principal de MEGA11 y realizar un análisis utilizando el criterio de Máxima Parsimonia [“Phylogeny/Construct/Test Maximum Parsimony Tree(s)”]. Analice el apoyo de los clados obtenidos utilizando «bootstrap», con 100 pseudoréplicas (opción “Test of Phylogeny-> Bootstrap Method; No. of Bootstrap Replication -> 100”). Mantenga los demás

parámetros del análisis en sus condiciones por defecto. Luego de obtenido el árbol filogenético, moverse de la pestaña “Original Tree” a la opción “Bootstrap consensus tree”. ¿Qué representan las reconstrucciones filogenéticas que se muestran en cada una de esas dos opciones? Definir el grupo externo en el “Bootstrap consensus tree”, utilizando las secuencias virales obtenidas del Mono verde Africano, *Cercopithecus aethiops* (SIVagm). Para lograr esto, seleccione primero la rama que representa a ese taxón y luego la opción: “Subtree/Root Tree” (también se puede hacer desde un botón ubicado en la barra lateral). ¿Qué función cumple el grupo externo?

3) Registre la longitud e índice de consistencia del árbol consenso obtenido. Esta información se encuentra disponible en la leyenda de la figura, la cuál se activa en el botón de “Display caption” de la barra lateral. Registre el valor de apoyo de los clados obtenidos utilizando bootstrap. Indique cuáles son los 3 clados que reciben menor apoyo estadístico.

4) Compare los valores de bootstrap con los obtenidos por otros compañeros. ¿Por qué difieren?

5) Opcional: Repita la reconstrucción filogenética –incluyendo el análisis de bootstrap– utilizando el algoritmo de unión de vecinos (“Phylogeny/Construct/Test Neighbor-Joining Tree...”).

6) De acuerdo a los resultados obtenidos con la reconstrucción filogenética discutir el posible origen y vías de contagio interespecíficas de los diferentes grupos de HIV-1. ¿Cuál fue el origen del contagio del grupo “HIV-1 (grupo 0)” en humanos? Teniendo en cuenta las vías de transmisión del virus, ¿cuáles serían las posibles formas de que se traspasaran las barreras específicas?

Basado en:

Hillis, D. 2010. Phylogenetic Progress and Applications of the Tree of Life, 421-449 p. En: Evolution since Darwin: The first 150 years, Editado por: Bell, M.; Futuyma, D.; Eanes, W.; & Levinton, J.; 688pp.

Van Heuverswyn et al. 2006. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature 444:164, doi:10.1038/444164a.

Práctico 2

Métodos Comparativos Filogenéticos

Resumen del problema

Es bien sabido que la presión parcial de oxígeno disminuye con la altura. Las personas no habituadas “se apunan” o sienten malestar al trasladarse a zonas de montaña, especialmente al momento de realizar actividades físicas exigentes. Un individuo que se traslada desde una zona baja hasta una de altura, experimenta a lo largo del tiempo una serie de ajustes fisiológicos para paliar al menos parte de esos desajustes.

Por lo tanto, cabe preguntarse si hay cambios genéticos que han sido favorecidos por la selección natural y que permiten a especies y poblaciones que viven a altas elevaciones adaptarse mejor a dichas condiciones. Tanto en humanos como en especies animales (y muchas otras), hay estudios orientados a identificar estos cambios.

Puesto que la afinidad de la sangre por el oxígeno es un factor clave para la vida en altura, y dicha afinidad depende fuertemente de las características de la hemoglobina (recordemos que la estructura cuaternaria de la hemoglobina combina dos cadenas de tipo alfa y dos de tipo beta en un tetrámero, en torno a un núcleo de hierro), esta proteína ha sido el blanco de muchos estudios.

En un trabajo reciente, Natarajan et al. (2016) se plantean identificar algunas de las adaptaciones de la hemoglobina para la vida en la altura. La hipótesis de trabajo es que la selección natural pudo haber favorecido cambios en las características de la hemoglobina de las especies asociados a la elevación en la que vive cada una.

Algunas de las ideas del artículo son: - Realizar un estudio comparando múltiples especies de aves, procurando elegir pares de especies cercanamente relacionadas, de modo que una de las especies de cada par viva en tierras altas y otra en las tierras bajas cercanas. - Para cada una de estas especies, aislar la hemoglobina y estudiar su afinidad con el oxígeno en el laboratorio.

El estudio incluye un análisis de los cambios en las secuencias de las hemoglobinas, que usaremos más adelante en el curso. Por el momento, extraemos del artículo los siguientes datos para cada una de las 56 especies estudiadas:

1. La elevación en la que viven (específicamente la altura de la localidad en la que fueron estudiadas).
2. La afinidad de su alfa globina por el oxígeno, tomada en condiciones controladas de laboratorio (en un medio que aproxima las condiciones en sangre), resumida por el valor conocido como P50.
3. El árbol filogenético que se utiliza en el artículo. Este árbol se toma como la mejor hipótesis disponible de las relaciones entre las especies.

Preguntas 1: Obtener de internet una curva de afinidad de la hemoglobina con el oxígeno. ¿Qué es el valor P50? ¿Cómo debería cambiar el P50 para que la hemoglobina tenga mayor afinidad con el oxígeno, por ejemplo en una especie de altura?

Activando los paquetes de R

Para esta actividad práctica usaremos, **R** (<https://cran.r-project.org/>) el cual es un entorno y lenguaje de programación enfocado al análisis estadístico. Se recomienda el uso **RStudio** (<https://rstudio.com/>), que es un entorno de desarrollo integrado (IDE), en conjunción con **R**, lo cual facilita el manejo de datos y la realización de los análisis.

Utilizaremos los siguientes paquetes de **R**, los cuales deben ser previamente instalados:

R Markdown: para formatear el archivo de salida de los análisis.

knitr: para convertir los archivos de R markdown en varios formatos.

phytools: varias funciones para análisis filogenéticos, principalmente orientado a la biología comparada.

Trabajando con los datos originales

Mediante el comando “setwd(dir)” podemos indicarle a R en donde están ubicados los archivos con los cuales vamos a estar trabajando, siendo “dir” la ruta hacia nuestro directorio (e.g. C:\User\Desktop\R). También podemos ubicar el directorio de trabajo de R mediante “getwd()” y colocar ahí nuestros archivos de interés.

El bloque de código siguiente lee solamente los datos de elevación y los de P50 de la HbA. Se hace una exploración de esos datos.

```
# Leemos las dos tablas de datos y combinarlas en un solo "data frame".
# Notar que en esta parte son archivos nuevos con los nombres de las
# variables cambiados
P50<-read.table("HbA_KCLHIP_P50.txt")
elevacion<-read.table("Elevación_2.txt", header = TRUE)
Datos = cbind(elevacion, P50) # acá se combinan los datos de las dos
# tablas
Datos$Elev.cat <- c("Alta", "Baja") # agrego variable categórica
# Elevación a cada par de spp.
table(Datos$Elev.cat) # chequeo que quedan 28 pares

##
## Alta Baja
## 28 28

print(Datos)

##
##          Familia          Especie
Elevacion
## Metriopelia_melanoptera  Columbidae  Metriopelia_melanoptera
4178
## Columbina_cruziana      Columbidae  Columbina_cruziana
372
## Hydropsalis_longirostris  Caprimulgidae  Hydropsalis_longirostris
4401
## Hydropsalis_decussata    Caprimulgidae  Hydropsalis_decussata
309
## Colibri_coruscans        Trochilidae    Colibri_coruscans
4030
## Schistes_geoffroyi       Trochilidae    Schistes_geoffroyi
```

| | | |
|------------------------------------|---------------|-------------------------|
| 1395 | | |
| ## Selasphorus_platycercus 2470 | Trochilidae | Selasphorus_platycercus |
| ## Archilochus_alexandri 2050 | Trochilidae | Archilochus_alexandri |
| ## Amazilia_viridicauda 3005 | Trochilidae | Amazilia_viridicauda |
| ## Amazilia_amazilia 366 | Trochilidae | Amazilia_amazilia |
| ## Chalcostigma_stanleyi 4300 | Trochilidae | Chalcostigma_stanleyi |
| ## Chalcostigma_ruficeps 2858 | Trochilidae | Chalcostigma_ruficeps |
| ## Oreotrochilus_estella 4391 | Trochilidae | Oreotrochilus_estella |
| ## Adelomyia_melanogenys 2147 | Trochilidae | Adelomyia_melanogenys |
| ## Eriocnemis_luciani 3680 | Trochilidae | Eriocnemis_luciani |
| ## Haplophaedia_aurelieae 1740 | Trochilidae | Haplophaedia_aurelieae |
| ## Aglaeactis_castelnaudii 4578 | Trochilidae | Aglaeactis_castelnaudii |
| ## Heliodoxa_leadbeateri 1890 | Trochilidae | Heliodoxa_leadbeateri |
| ## Pterophanes_cyanopterus 4200 | Trochilidae | Pterophanes_cyanopterus |
| ## Boissonneaua_matthewsii 2850 | Trochilidae | Boissonneaua_matthewsii |
| ## Coeligena_violifer 3779 | Trochilidae | Coeligena_violifer |
| ## Coeligena_coeligena 2132 | Trochilidae | Coeligena_coeligena |
| ## Cinclodes_albiventris 4401 | Furnariidae | Cinclodes_albiventris |
| ## Furnarius_leucopus 350 | Furnariidae | Furnarius_leucopus |
| ## Notiochelidon_murina 4470 | Hirundinidae | Notiochelidon_murina |
| ## Pygochelidon_cyanoleuca 2550 | Hirundinidae | Pygochelidon_cyanoleuca |
| ## Troglodytes_aedonH 4375 | Troglodytidae | Troglodytes_aedonH |
| ## Troglodytes_aedonL 143 | Troglodytidae | Troglodytes_aedonL |
| ## Spinus_magellanicaH 3945 | Fringillidae | Spinus_magellanicaH |
| ## Spinus_magellanicaL 39 | Fringillidae | Spinus_magellanicaL |
| ## Zonotrichia_capensisH 4079 | Emberizidae | Zonotrichia_capensisH |
| ## Zonotrichia_capensisL 39 | Emberizidae | Zonotrichia_capensisL |
| ## Tangara_vassorii | Thraupidae | Tangara_vassorii |

| | | | |
|--------------------------------|--------------|------|-----------------------------|
| 3421 | | | |
| ## Tangara_nigroviridis | Thraupidae | | Tangara_nigroviridis |
| 2698 | | | |
| ## Conirostrum_cinereumH | Thraupidae | | Conirostrum_cinereumH |
| 4056 | | | |
| ## Conirostrum_cinereumL | Thraupidae | | Conirostrum_cinereumL |
| 352 | | | |
| ## Diglossa_brunneiventris | Thraupidae | | Diglossa_brunneiventris |
| 4385 | | | |
| ## Diglossa_glauca | Thraupidae | | Diglossa_glauca |
| 1395 | | | |
| ## Catamenia_analisH | Thraupidae | | Catamenia_analisH |
| 3573 | | | |
| ## Catamenia_analisL | Thraupidae | | Catamenia_analisL |
| 39 | | | |
| ## Oxyura_jamaicensisH | Anatidae | | Oxyura_jamaicensisH |
| 1250 | | | |
| ## Oxyura_jamaicensisL | Anatidae | | Oxyura_jamaicensisL |
| 350 | | | |
| ## Merganetta_armataH | Anatidae | | Merganetta_armataH |
| 4500 | | | |
| ## Merganetta_armataL | Anatidae | | Merganetta_armataL |
| 1500 | | | |
| ## Chloephaga_melanoptera | Anatidae | | Chloephaga_melanoptera |
| 4700 | | | |
| ## Neochen_jubata | Anatidae | | Neochen_jubata |
| 300 | | | |
| ## Lophonetta_s_alticola | Anatidae | | Lophonetta_s_alticola |
| 4800 | | | |
| ## Lophonetta_s_specularioides | Anatidae | | Lophonetta_s_specularioides |
| 1500 | | | |
| ## Anas_georgicaH | Anatidae | | Anas_georgicaH |
| 4600 | | | |
| ## Anas_georgicaL | Anatidae | | Anas_georgicaL |
| 39 | | | |
| ## Anas_f_oxypetra | Anatidae | | Anas_f_oxypetra |
| 4400 | | | |
| ## Anas_f_flavirostris | Anatidae | | Anas_f_flavirostris |
| 39 | | | |
| ## Anas_c_orinoma | Anatidae | | Anas_c_orinoma |
| 4600 | | | |
| ## Anas_c_cyanoptera | Anatidae | | Anas_c_cyanoptera |
| 39 | | | |
| ## Anas_puna | Anatidae | | Anas_puna |
| 4600 | | | |
| ## Anas_versicolor | Anatidae | | Anas_versicolor |
| 39 | | | |
| ## | P50 Elev.cat | | |
| ## Metriopelia_melanoptera | 26.86 | Alta | |
| ## Columbina_cruziana | 28.38 | Baja | |
| ## Hydropsalis_longirostris | 30.21 | Alta | |
| ## Hydropsalis_decussata | 36.12 | Baja | |
| ## Colibri_coruscans | 31.08 | Alta | |
| ## Schistes_geoffroyi | 36.81 | Baja | |

| | | |
|---------------------------------------|-------|------|
| ## <i>Selasphorus_platycercus</i> | 38.25 | Alta |
| ## <i>Archilochus_alexandri</i> | 39.12 | Baja |
| ## <i>Amazilia_viridicauda</i> | 24.24 | Alta |
| ## <i>Amazilia_amazilia</i> | 29.84 | Baja |
| ## <i>Chalcostigma_stanleyi</i> | 32.34 | Alta |
| ## <i>Chalcostigma_ruficeps</i> | 33.02 | Baja |
| ## <i>Oreotrochilus_estella</i> | 20.20 | Alta |
| ## <i>Adelomyia_melanogenys</i> | 32.02 | Baja |
| ## <i>Eriocnemis_luciani</i> | 31.81 | Alta |
| ## <i>Haplophaedia_aurelieae</i> | 35.67 | Baja |
| ## <i>Aglaeactis_castelnaudii</i> | 17.23 | Alta |
| ## <i>Heliodoxa_leadbeateri</i> | 17.23 | Baja |
| ## <i>Pterophanes_cyanopterus</i> | 30.43 | Alta |
| ## <i>Boissonneaua_matthewsii</i> | 36.51 | Baja |
| ## <i>Coeligena_violifer</i> | 19.12 | Alta |
| ## <i>Coeligena_coeligena</i> | 22.90 | Baja |
| ## <i>Cinclodes_albiventris</i> | 25.12 | Alta |
| ## <i>Furnarius_leucopus</i> | 44.69 | Baja |
| ## <i>Notiochelidon_murina</i> | 30.89 | Alta |
| ## <i>Pygochelidon_cyanoleuca</i> | 39.59 | Baja |
| ## <i>Troglodytes_aedonH</i> | 17.07 | Alta |
| ## <i>Troglodytes_aedonL</i> | 25.88 | Baja |
| ## <i>Spinus_magellanicaH</i> | 27.82 | Alta |
| ## <i>Spinus_magellanicaL</i> | 31.53 | Baja |
| ## <i>Zonotrichia_capensisH</i> | 39.98 | Alta |
| ## <i>Zonotrichia_capensisL</i> | 36.28 | Baja |
| ## <i>Tangara_vassorii</i> | 43.09 | Alta |
| ## <i>Tangara_nigroviridis</i> | 43.22 | Baja |
| ## <i>Conirostrum_cinereumH</i> | 39.47 | Alta |
| ## <i>Conirostrum_cinereumL</i> | 40.50 | Baja |
| ## <i>Diglossa_brunneiventris</i> | 32.06 | Alta |
| ## <i>Diglossa_glauca</i> | 40.68 | Baja |
| ## <i>Catamenia_analisH</i> | 37.46 | Alta |
| ## <i>Catamenia_analisL</i> | 37.95 | Baja |
| ## <i>Oxyura_jamaicensisH</i> | 30.05 | Alta |
| ## <i>Oxyura_jamaicensisL</i> | 28.52 | Baja |
| ## <i>Merganetta_armataH</i> | 26.60 | Alta |
| ## <i>Merganetta_armataL</i> | 27.97 | Baja |
| ## <i>Chloephaga_melanoptera</i> | 27.64 | Alta |
| ## <i>Neochen_jubata</i> | 35.31 | Baja |
| ## <i>Lophonetta_s_alticola</i> | 25.14 | Alta |
| ## <i>Lophonetta_s_specularioides</i> | 37.98 | Baja |
| ## <i>Anas_georgicaH</i> | 35.62 | Alta |
| ## <i>Anas_georgicaL</i> | 42.20 | Baja |
| ## <i>Anas_f_oxypetra</i> | 30.08 | Alta |
| ## <i>Anas_f_flavirostris</i> | 30.96 | Baja |
| ## <i>Anas_c_orinoma</i> | 29.36 | Alta |
| ## <i>Anas_c_cyanoptera</i> | 37.43 | Baja |
| ## <i>Anas_puna</i> | 27.32 | Alta |
| ## <i>Anas_versicolor</i> | 39.66 | Baja |

rm(P50, elevacion)

Preguntas 2: La tabla de datos incluye varios pares de especies de un mismo género. Elegir algunos de esos pares para discutir:

1. ¿De qué especies se trata? Averiguar algo de los nombres comunes, familias a las que pertenecen, distribución geográfica. 2. ¿Qué tendencias se observan al examinar en varios de esos pares la relación entre altura y P50?

Examinando la relación entre elevación y P50

En el siguiente bloque se obtiene el coeficiente de correlación entre las dos variables de interés, se aplica un modelo de regresión lineal entre dichas variables, y se grafican los valores junto con la línea de tendencia obtenida en la regresión.

```
attach(Datos) # Fijamos Las Variables de Datos (ya no es necesario
Llamarlas a través del comando Datos$x)

# Análisis exploratorios
str(Datos) # Resumen de número y tipo de variables y nro de observaciones

## 'data.frame': 56 obs. of 5 variables:
## $ Familia : chr "Columbidae" "Columbidae" "Caprimulgidae"
"Caprimulgidae" ...
## $ Especie : chr "Metriopelia_melanoptera" "Columbina_cruziana"
"Hydropsalis_longirostris" "Hydropsalis_decussata" ...
## $ Elevacion: int 4178 372 4401 309 4030 1395 2470 2050 3005 366 ...
## $ P50 : num 26.9 28.4 30.2 36.1 31.1 ...
## $ Elev.cat : chr "Alta" "Baja" "Alta" "Baja" ...

Datos$Elevacion <- as.numeric(as.character(Datos$Elevacion))

summary(Familia) # cuantas observaciones de cada Familia

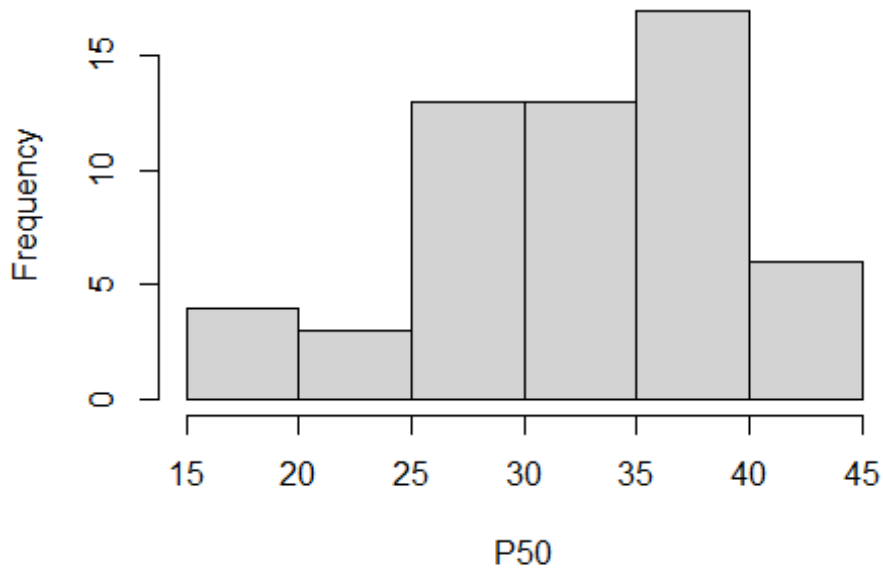
## Length Class Mode
## 56 character character

summary(P50)

## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 17.07 27.77 31.67 32.04 37.58 44.69

hist(P50)
```

Histogram of P50

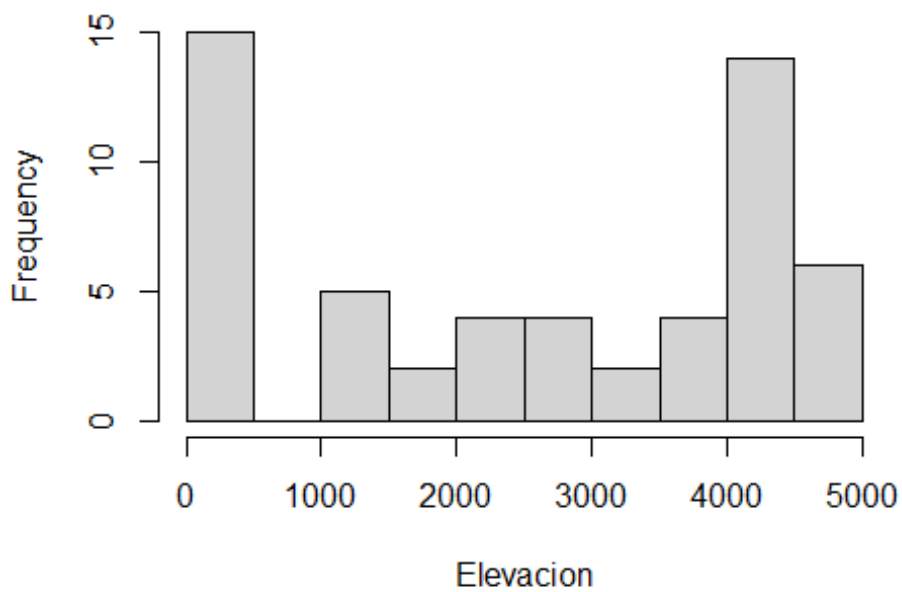


```
summary(Elevacion)
```

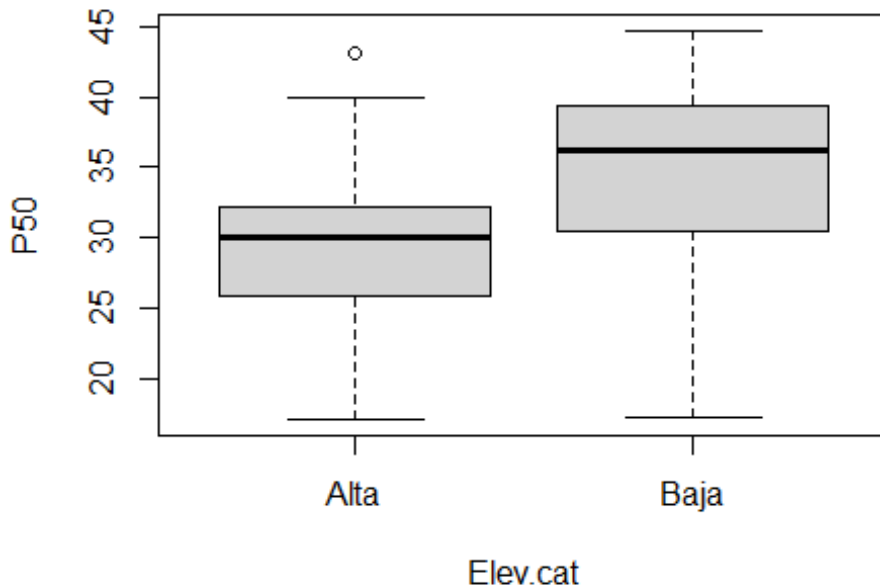
```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   39.0  370.5  2774.0  2548.0  4318.8  4800.0
```

```
hist(Elevacion)
```

Histogram of Elevacion



```
# Exploro relación P50 vs. Elevación (categórica)
boxplot(P50~Elev.cat)
```

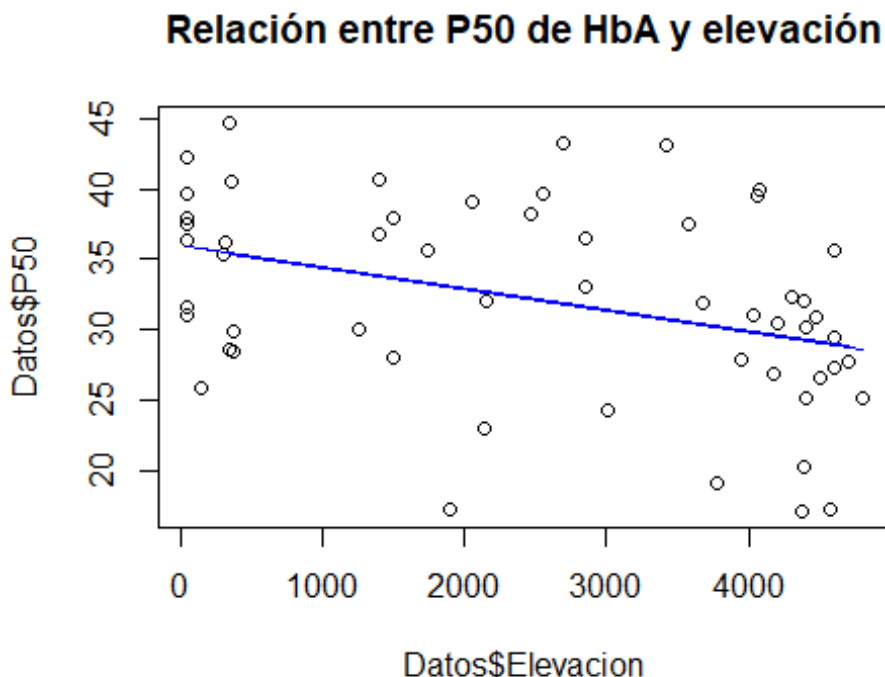


```
# ...y la relación P50 vs. Elevación (numérica)
# Calculamos el coeficiente de correlación entre las dos variables
R = cor(Datos$Elevacion,Datos$P50)

# Aplicar un modelo lineal (lm) de regresión. Se define a P50 como la
variable dependiente y a Elevacion como la variable independiente.
Regresion <- lm(Datos$P50~Datos$Elevacion)
summary.lm(Regresion)

##
## Call:
## lm(formula = Datos$P50 ~ Datos$Elevacion)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.8125  -4.1662   0.8666   4.4904  12.3688
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  35.9080738  1.5388246  23.335 < 2e-16 ***
## Datos$Elevacion -0.0015162  0.0004991  -3.038  0.00367 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.483 on 54 degrees of freedom
## Multiple R-squared:  0.1459, Adjusted R-squared:  0.1301
## F-statistic: 9.228 on 1 and 54 DF, p-value: 0.003667
```

```
# Graficamos Los valores de P50 y elevación, agregando La línea de ajuste
obtenida de La regresión Lineal.
plot(Datos$P50~Datos$Elevacion, main = "Relación entre P50 de HbA y
elevación")
lines(Datos$Elevacion, fitted(Regresion), col="blue")
```



Preguntas 3. ¿Cuál es el valor del coeficiente de correlación? ¿Qué sugiere respecto a la relación entre las variables (signo, intensidad)? 2. Examinar la tabla de regresión. ¿Qué fracción de la varianza está explicada por el modelo? 3. Identificar y explicar los parámetros relevantes del modelo y el p-valor asociado a cada uno. ¿Qué sugieren estos resultados sobre la hipótesis de trabajo? 4. Examinar la gráfica, incluyendo la predicción obtenida por regresión lineal. ¿Qué sugiere la gráfica y cómo se relaciona con los puntos discutidos más arriba?

Las especies, y por lo tanto sus rasgos, no pueden ser tratadas como variables independientes desde el punto de vista estadístico dado que comparten una historia evolutiva en común. Este problema se hace aún más evidente en taxa cercanamente emparentados. La respuesta a tal problema fue propuesta por Felsenstein (1985) mediante el cálculo de Contrastes Filogenéticos Independientes (CFI), mediante el cual se transforma a los rasgos analizados en variables independientes, empleando la filogenia de las especies como marco de análisis. En la siguiente sección, vamos a abordar este problema y analizaremos nuevamente la correlación entre la P50 y la Altura, pero esta vez corrigiendo la falta de independencia de los datos, utilizando los CFI como nuevas variables, independientes de la historia evolutiva compartida entre las diferentes especies de aves.

```
#Leemos el árbol filogenético
tree<-read.tree("tree.tre")

#enraizamos el árbol en su punto medio
```

```
treeR<-midpoint.root(tree)

#visualizamos el árbol enraizado
plotTree(treeR, edge.width=1, ftype="i", fsize=0.7)
```



```
#para simplificar los comandos, leemos otro archivo que contiene solo las
especies y las máximas alturas registradas, así como otro archivo con los
valores de P50 (se puede corroborar y si no se modificó para la etapa
anterior, se puede usar el mismo en ambas partes)
elevacion_2<-read.table("Elevacion.txt")
P50_2<-read.table("HbA_KCLHIP_P50.txt")

#Creamos un vector numérico de las variables para calcular los contrastes
VP50 <- P50_2$P50
Velevacion <- elevacion_2$Elevacion

#Para que los datos y los taxa terminales se asocien correctamente
names(VP50) <- row.names(P50_2)
names(Velevacion) <- row.names(elevacion_2)

#calculamos los contrastes filogenéticamente independientes
ContrasteVP50 <- pic(VP50, treeR)
ContrasteVelevacion <- pic(Velevacion, treeR)

#si queremos extraer además de los contrastes, su varianza asociada
ContrasteVP50.var <- pic(VP50, treeR, var.contrasts=TRUE)
ContrasteVelevacion.var <- pic(Velevacion, treeR, var.contrasts=TRUE)

#ahora visualizamos los contrastes filogenéticos calculados anteriormente
ContrasteVP50
```

| | | | | | |
|-------------|--------------|-------------|-------------|-------------|-------------|
| ## | 57 | 58 | 59 | 60 | 61 |
| 62 | | | | | |
| ## | 1.21599821 | 1.56277493 | -1.81378821 | -4.42508324 | 6.92515888 |
| 5.37509655 | | | | | |
| ## | 63 | 64 | 65 | 66 | 67 |
| 68 | | | | | |
| ## | 1.63034354 | -2.49099309 | -1.63497883 | 0.77076261 | 0.34648232 |
| -6.09526045 | | | | | |
| ## | 69 | 70 | 71 | 72 | 73 |
| 74 | | | | | |
| ## | -0.72831998 | -0.09192388 | 2.61629509 | -2.62336616 | -6.22961074 |
| -6.15182900 | | | | | |
| ## | 75 | 76 | 77 | 78 | 79 |
| 80 | | | | | |
| ## | -13.83807971 | 0.92603541 | 1.74299931 | 1.37103384 | -0.68017986 |
| 0.05484828 | | | | | |
| ## | 81 | 82 | 83 | 84 | 85 |
| 86 | | | | | |
| ## | -8.72569768 | -5.70635172 | -4.84396876 | -0.62225397 | -4.65276262 |
| -9.07925107 | | | | | |
| ## | 87 | 88 | 89 | 90 | 91 |
| 92 | | | | | |
| ## | 2.41909763 | -5.42350901 | -0.96873629 | 1.08187338 | 2.90637083 |
| -0.29722727 | | | | | |
| ## | 93 | 94 | 95 | 96 | 97 |
| 98 | | | | | |
| ## | -1.63227832 | -2.13005429 | -0.82035448 | -7.12925805 | 3.18727496 |
| -3.00177253 | | | | | |
| ## | 99 | 100 | 101 | 102 | 103 |
| 104 | | | | | |
| ## | -2.67286363 | 8.68064514 | -4.29920923 | -2.72943218 | -2.38730372 |
| 6.67084060 | | | | | |
| ## | 105 | 106 | 107 | 108 | 109 |
| 110 | | | | | |
| ## | -0.48083261 | -6.72324388 | -3.95979797 | -0.61518290 | -4.05172186 |
| -4.17900108 | | | | | |
| ## | 111 | | | | |
| ## | -1.07480231 | | | | |

ContrasteVelevacion

| | | | | | | |
|----|------------|------------|------------|------------|-------------|------------|
| ## | 57 | 58 | 59 | 60 | 61 | 62 |
| ## | -68.16633 | 443.42685 | 300.75919 | -712.44242 | -57.95270 | 196.07592 |
| ## | 63 | 64 | 65 | 66 | 67 | 68 |
| ## | 352.72895 | -433.84602 | 79.87683 | -625.84769 | -2498.91536 | 2114.24928 |
| ## | 69 | 70 | 71 | 72 | 73 | 74 |
| ## | 2619.12352 | 511.23820 | 2856.71140 | 2761.95909 | 2992.47590 | 1357.64502 |
| ## | 75 | 76 | 77 | 78 | 79 | 80 |
| ## | 2864.48957 | 1063.50194 | 10.44802 | -465.67522 | 26.72612 | 0.00000 |
| ## | 81 | 82 | 83 | 84 | 85 | 86 |
| ## | 3225.11403 | 3225.11403 | -57.73503 | 3083.69267 | 3225.11403 | 2333.45238 |
| ## | 87 | 88 | 89 | 90 | 91 | 92 |
| ## | -288.67513 | 3111.26984 | 2121.32034 | 636.39610 | 109.80928 | 146.61554 |
| ## | 93 | 94 | 95 | 96 | 97 | 98 |

```

## -110.31928 588.93661 35.75795 532.61130 -859.03997 71.61372
## 99 100 101 102 103 104
## 1164.60487 873.94426 954.59415 1371.78716 906.33164 -434.03226
## 105 106 107 108 109 110
## 1019.64798 -331.68773 1866.05480 296.98485 1863.22637 2893.48095
## 111
## 2691.24841

```

vemos los contrastes filogenéticos calculados y también su varianza asociada

ContrasteVP50.var

```

##      contrasts variance
## 57  1.21599821 2.933534
## 58  1.56277493 4.660203
## 59 -1.81378821 3.322870
## 60 -4.42508324 3.322849
## 61  6.92515888 3.322746
## 62  5.37509655 3.322242
## 63  1.63034354 3.319767
## 64 -2.49099309 3.307692
## 65 -1.63497883 3.250000
## 66  0.77076261 3.000000
## 67  0.34648232 2.000000
## 68 -6.09526045 2.000000
## 69 -0.72831998 2.000000
## 70 -0.09192388 2.000000
## 71  2.61629509 2.000000
## 72 -2.62336616 2.000000
## 73 -6.22961074 2.000000
## 74 -6.15182900 2.000000
## 75 -13.83807971 2.000000
## 76  0.92603541 3.395349
## 77  1.74299931 3.583333
## 78  1.37103384 3.375000
## 79 -0.68017986 3.500000
## 80  0.05484828 3.000000
## 81 -8.72569768 2.000000
## 82 -5.70635172 2.000000
## 83 -4.84396876 3.000000
## 84 -0.62225397 2.000000
## 85 -4.65276262 2.000000
## 86 -9.07925107 2.000000
## 87  2.41909763 3.000000
## 88 -5.42350901 2.000000
## 89 -0.96873629 2.000000
## 90  1.08187338 2.000000
## 91  2.90637083 3.326879
## 92 -0.29722727 3.342637
## 93 -1.63227832 3.422768
## 94 -2.13005429 3.702106
## 95 -0.82035448 3.811838
## 96 -7.12925805 3.455556
## 97  3.18727496 3.829787

```

```
## 98 -3.00177253 3.357143
## 99 -2.67286363 2.000000
## 100 8.68064514 3.500000
## 101 -4.29920923 2.000000
## 102 -2.72943218 2.000000
## 103 -2.38730372 3.857143
## 104 6.67084060 3.500000
## 105 -0.48083261 2.000000
## 106 -6.72324388 3.000000
## 107 -3.95979797 2.000000
## 108 -0.61518290 2.000000
## 109 -4.05172186 2.000000
## 110 -4.17900108 2.000000
## 111 -1.07480231 2.000000
```

ContrasteVelevacion.var

```
## contrasts variance
## 57 -68.16633 2.933534
## 58 443.42685 4.660203
## 59 300.75919 3.322870
## 60 -712.44242 3.322849
## 61 -57.95270 3.322746
## 62 196.07592 3.322242
## 63 352.72895 3.319767
## 64 -433.84602 3.307692
## 65 79.87683 3.250000
## 66 -625.84769 3.000000
## 67 -2498.91536 2.000000
## 68 2114.24928 2.000000
## 69 2619.12352 2.000000
## 70 511.23820 2.000000
## 71 2856.71140 2.000000
## 72 2761.95909 2.000000
## 73 2992.47590 2.000000
## 74 1357.64502 2.000000
## 75 2864.48957 2.000000
## 76 1063.50194 3.395349
## 77 10.44802 3.583333
## 78 -465.67522 3.375000
## 79 26.72612 3.500000
## 80 0.00000 3.000000
## 81 3225.11403 2.000000
## 82 3225.11403 2.000000
## 83 -57.73503 3.000000
## 84 3083.69267 2.000000
## 85 3225.11403 2.000000
## 86 2333.45238 2.000000
## 87 -288.67513 3.000000
## 88 3111.26984 2.000000
## 89 2121.32034 2.000000
## 90 636.39610 2.000000
## 91 109.80928 3.326879
## 92 146.61554 3.342637
```



```

## 93 -110.31928 3.422768
## 94 588.93661 3.702106
## 95 35.75795 3.811838
## 96 532.61130 3.455556
## 97 -859.03997 3.829787
## 98 71.61372 3.357143
## 99 1164.60487 2.000000
## 100 873.94426 3.500000
## 101 954.59415 2.000000
## 102 1371.78716 2.000000
## 103 906.33164 3.857143
## 104 -434.03226 3.500000
## 105 1019.64798 2.000000
## 106 -331.68773 3.000000
## 107 1866.05480 2.000000
## 108 296.98485 2.000000
## 109 1863.22637 2.000000
## 110 2893.48095 2.000000
## 111 2691.24841 2.000000

```

#vemos Los contrastes asociados a cada nodo, a partir del archivo que también contiene el cálculo de Las varianzas. Los contrastes asociados a P50 aparecerán en azul y Los asociados a La Elevación aparecerán en verde.

```

plot(treeR, no.margin=TRUE,edge.width=1 ,cex=c(0.6,0.6))
nodelabels(round(ContrasteVelevacion.var [,1], 3), adj = c(1, -0.4),
frame="n", col = "blue", cex=0.6)

```



```
#volvemos a plotear la filogenia solo para que nos se superpongan ambos contrastes
```

```
plot(treeR, no.margin=TRUE, edge.width=1 ,cex=c(0.6,0.6))  
nodelabels(round (ContrasteVP50.var [,1], 3), adj = c(1, 1.1), frame="n",  
cex=0.6, col = "darkgreen")
```



```
#Analizamos la evolución correlacionada de ambos rasgos a partir de Los  
archivos sin el cálculo de varianzas
```

```
RegresionP50_elevacion <- lm(ContrasteVP50~ContrasteVelevacion)
```

```
#Visualizamos Los parámetros de la regresión
```

```
summary.lm(RegresionP50_elevacion)
```

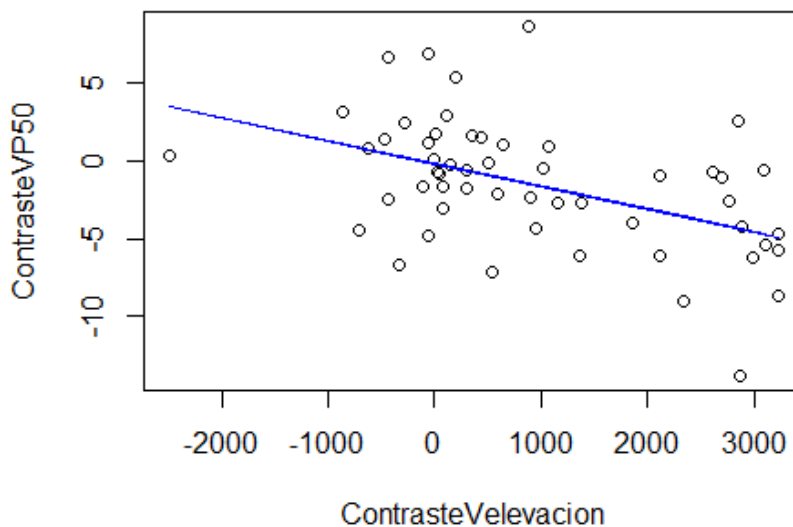
```
##  
## Call:  
## lm(formula = ContrasteVP50 ~ ContrasteVelevacion)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max    
## -9.4792 -1.6509 -0.0061  2.1798 10.1324   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   -0.17539    0.60601  -0.289  0.773392   
## ContrasteVelevacion -0.00146    0.00037  -3.947  0.000235 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##
```

```
## Residual standard error: 3.671 on 53 degrees of freedom
## Multiple R-squared: 0.2271, Adjusted R-squared: 0.2126
## F-statistic: 15.58 on 1 and 53 DF, p-value: 0.0002347
```

```
#visualizamos la relación entre los contrastes y agregamos una línea de
tendencia a la regresión
```

```
plot(ContrasteVP50~ContrasteVelevacion, main = "Relación entre contrastes
de P50 de HbA y de elevación")
lines(ContrasteVelevacion, fitted(RegresionP50_elevacion), col="blue")
```

Relación entre contrastes de P50 de HbA y de elevación



```
#Veamos ahora cómo se distribuyen los caracteres sobre la filogenia
```

```
#primero definimos el nombre de los datos
```

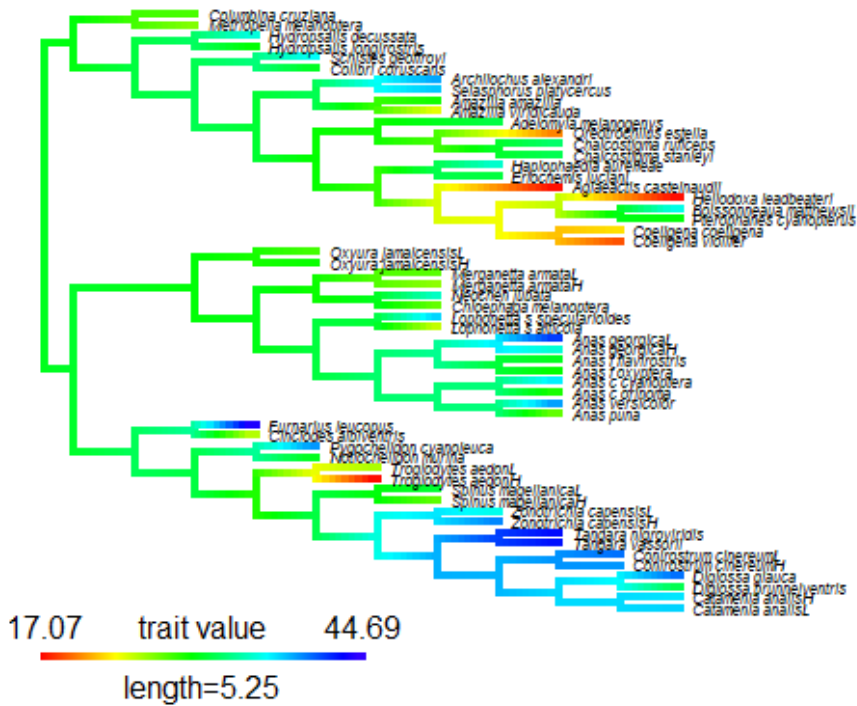
```
P50_2<-setNames(P50_2[,1],rownames(P50_2))
```

```
elevacion_2<-setNames(elevacion_2[,1],rownames(elevacion_2))
```

```
#y luego obtenemos su distribución sobre la filogenia propuesta
```

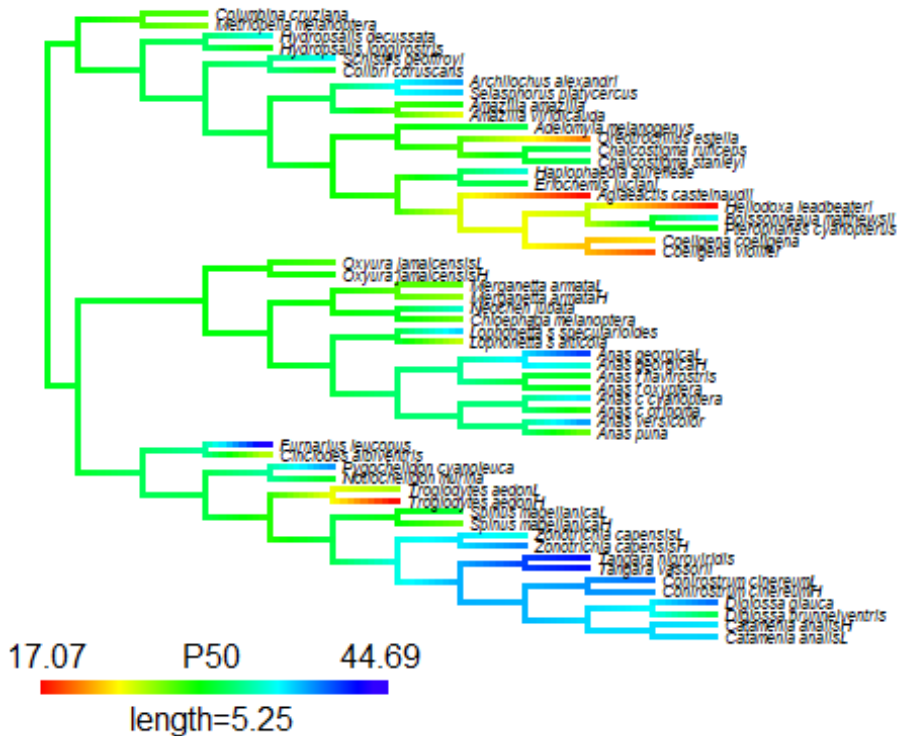
```
#para el valor de P50
```

```
Dist_P50<-contMap(treeR,P50_2, fsize=c(0.5,1),outline=FALSE)
```



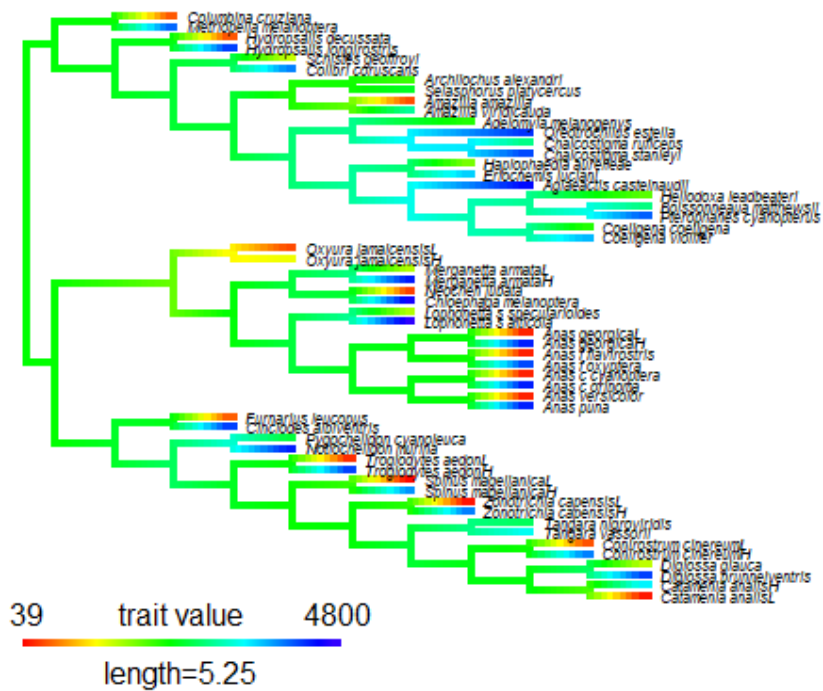
#modificamos la Leyenda del gráfico

```
plot(Dist_P50, fsize=c(0.5,1), outline=FALSE, lwd=c(3,7), leg.txt="P50")
```

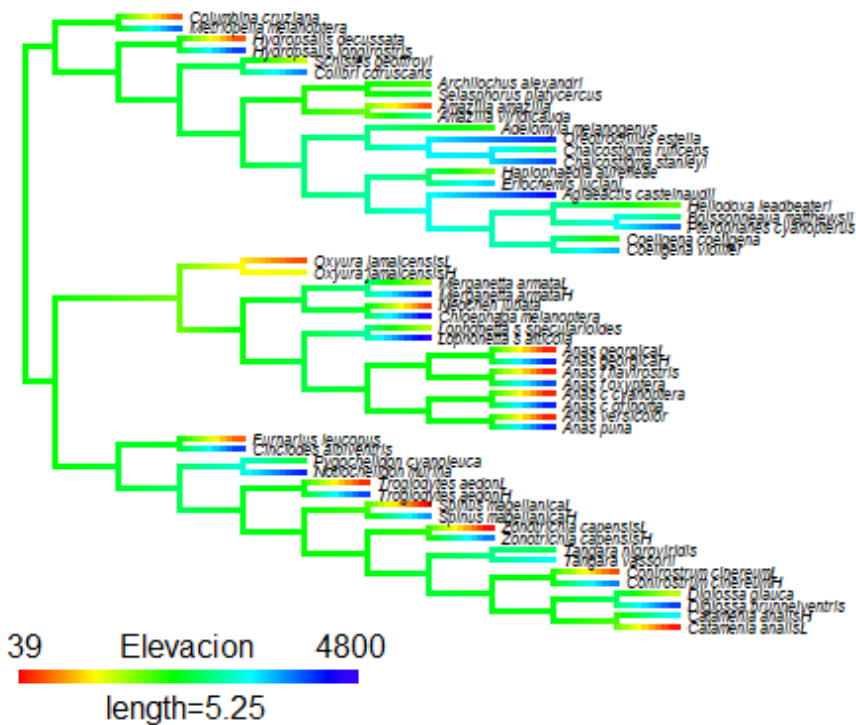


#para las altitudes máximas registradas para cada taxón

```
Dist_elevacion<-contMap(treeR, elevacion_2, fsize=c(0.5,1), outline=FALSE)
```



```
#nuevamente, modificamos La Leyenda del gráfico
plot(Dist_elevacion, fsize=c(0.5,1), outline=FALSE, lwd=c(3,7), leg.txt="Elevacion")
```



Pregunta 4

¿Según los resultados obtenidos, la historia filogenética compartida de las especies parece haber influido en la adaptación de la hemoglobina a la hipoxia?

Práctico 3

Deriva genética

Parte I. Ley de Hardy-Weinberg

El modelo de Hardy-Weinberg describe el comportamiento de las frecuencias génicas de una generación a la otra en poblaciones ideales. Se dice que una población está en equilibrio de Hardy-Weinberg cuando sus frecuencias alélicas y genotípicas se ajustan a las predicciones del modelo.

Bajo los supuestos del modelo, siendo p_i la frecuencia del alelo A_i , las predicciones de frecuencias genotípicas serían las siguientes:

$$E(A_i A_i) = p_i^2$$

$$E(A_i A_j) = 2p_i p_j$$

Las predicciones de Heterocigosidad, para múltiples alelos, estarán dadas por la siguiente ecuación:

$$E(H) = 1 - \sum p_i^2$$

Siendo " $\sum p_i^2$ " la suma de las frecuencias esperada de homocigotos, o sea la Homocigosidad.

La Heterocigosidad (medida de la variabilidad en la población) aumenta a medida que aumenta el número de alelos en la población.

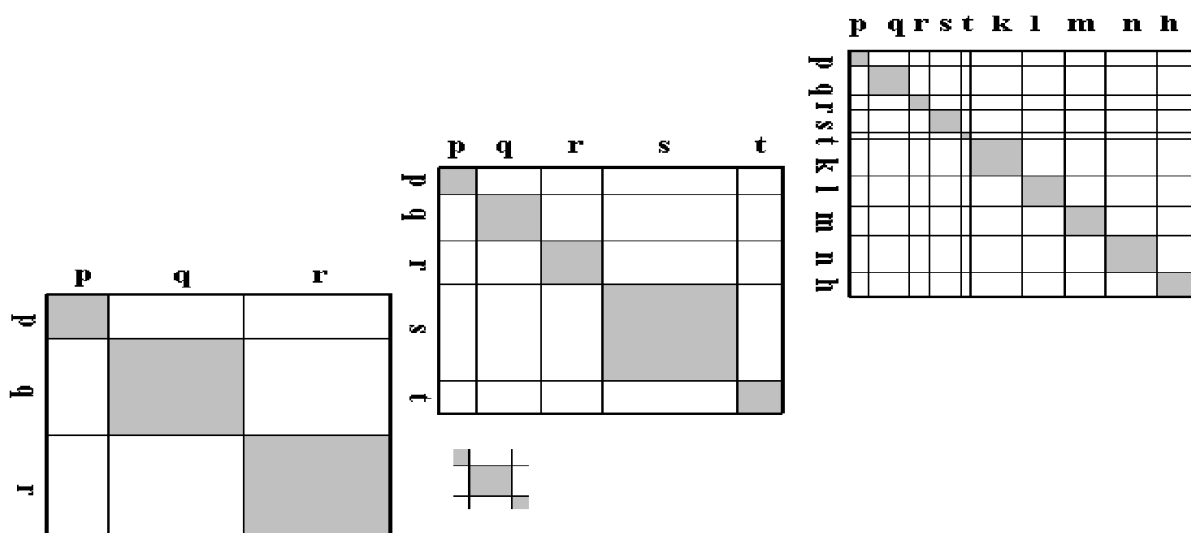


Figura- El área sombreada representa la fracción de homocigotos esperados bajo equilibrio Hardy-Weinberg.

Caso de estudio- Se estudia la variación genética de una población de *Ctenomys rionegrensis*. A partir del mismo, calcule las frecuencias alélicas y genotípicas observadas globales. Calcule la heterocigosis observada y la esperada a partir del modelo de Hardy-Weinberg considerando a toda la muestra como parte de una única población. Interprete los resultados.

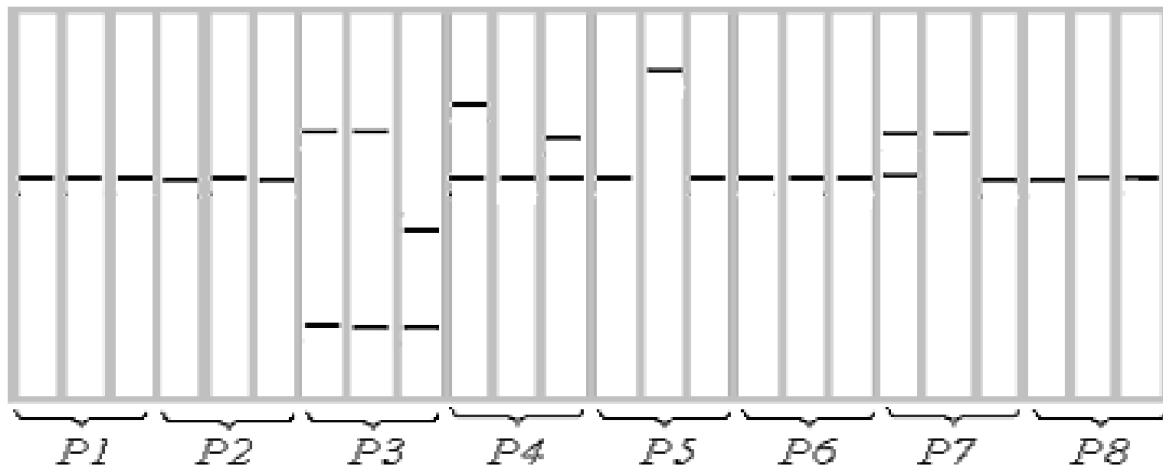
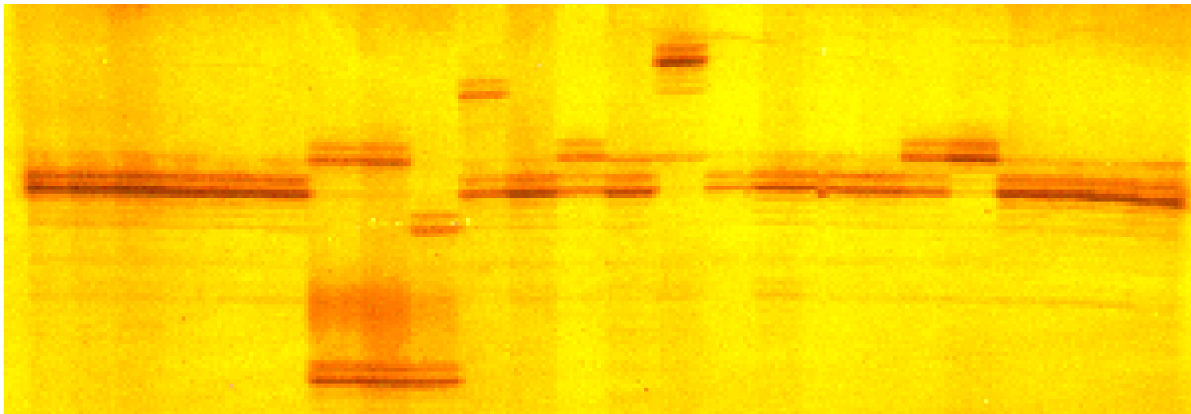


Figura- Gel de la poliacrilamida mostrando los resultados de la amplificación de microsatélites para tres individuos de cada una de ocho subpoblaciones de *Ctenomys rionegrensis* (arriba), y su representación esquemática (abajo).

La resolución de este problema se encuentra disponible en los materiales del práctico.

Parte II. Deriva genética

Cuando de los supuestos del modelo de Hardy-Weinberg se abandona la condición de que las poblaciones tienen tamaño infinito, se puede observar que de una generación a otra se producen cambios aleatorios en las frecuencias génicas. Esto es efecto directo de que los alelos que forman una nueva generación son un muestreo al azar de los alelos presentes en la generación parental. En este caso, el sentido de la evolución no se puede predecir, ya que está causada por los efectos de muestreo de generación en generación, un proceso llamado deriva genética.

El modelo poblacional que sólo difiere del de Hardy-Weinberg en poseer un tamaño poblacional finito es el modelo Fisher-Wright, que modela la probabilidad de un alelo de tener determinada frecuencia en la generación siguiente, únicamente a partir de la frecuencia inicial del alelo y el tamaño poblacional, realizando muestreos al azar con reposición.

Para los siguientes ejercicios se utilizará el programa AlleleA1 que permite simular variaciones en la frecuencia génica en función del tiempo, cambiando parámetros como la frecuencia alélica inicial y el tamaño poblacional, entre otros.

Las simulaciones se pueden hacer para uno o varios loci no ligados (Number of Loci) cada uno con dos alelos, lo que también se puede interpretar como varios ensayos sucesivos independientes para un solo locus con dos alelos. En cualquiera de los casos, el programa grafica la frecuencia de uno de los dos alelos del locus, y el otro alelo será el complemento de esa. El botón "Run" en la ventana gráfica permite realizar nuevas simulaciones manteniendo los mismos parámetros. El botón "Multiple" permite realizar varias simulaciones en una misma ventana, donde se podrá elegir el color a cada corrida. Si se selecciona "Auto" el color cambiará automáticamente. Con el botón "Clear" se borrarán las simulaciones realizadas.

Ejercicio 1. Efecto del tamaño poblacional

¿Cómo crees que es el efecto de la deriva genética según el tamaño poblacional?

Para comparar el efecto de diferentes tamaños poblacionales, realizaremos simulaciones manteniendo una frecuencia inicial de 0.5, para 5 loci, durante 500 generaciones, y dos tamaños poblacionales diferentes:

Simulación A: $N = 500$ (gran tamaño poblacional)

Simulación B: $N = 50$ (pequeño tamaño poblacional)

Anotar por lo menos para una simulación de cada tipo las frecuencias finales de los alelos ¿Cuántos alelos se fijan o eliminan? ¿en cuántas generaciones? Comparar y discutir tus resultados con los de tus compañeros. ¿Son estos resultados coherentes con tus predicciones?

Ejercicio 2. Efecto de la frecuencia alélica inicial

Reflexione acerca de con cuál frecuencia inicial se encontrará una mutación que recién surge en un población. ¿Cuál será su probabilidad de fijarse?

Corre algunas veces la siguiente simulación durante 500 generaciones:

Simulación C: $N = 500$, y una baja frecuencia alélica inicial: 0,1

Nuevamente registre para al menos una corrida, cuántos alelos se fijan o eliminan y en cuántas generaciones. Comparar los resultados obtenidos en esta última simulación con los de la simulación A. ¿Son estos resultados coherentes con sus predicciones?

Ejercicio 3. Efecto a largo plazo sobre la heterocigosis

Realizar una nueva simulación con $N=500$, por 1000 generaciones, y frecuencia inicial de 0,5, y calcula la heterocigosis promedio mediante la ecuación básica $H = 2pq$ en $t = 0, 500$ y 1000. ¿Existe alguna tendencia en los resultados?

Parte III. Deriva genética y mutación

Ahora introduciremos una fuente de variación abandonando otro de los supuestos de los modelos de Hardy-Weinberg y de Fisher-Wright. Permitiremos el surgimiento de mutaciones neutrales (sin selección) a una tasa μ por alelo por generación. Consideramos que cada variante nueva puede surgir por mutación una única vez.

Ejercicio 4

Reflexione sobre: ¿cuántas mutaciones neutrales, en promedio, se fijarán por deriva en cada generación? ¿Cómo afecta el tamaño de la población a esta tasa de sustitución?

Realizar dos simulaciones con una tasa de mutación $m=0,00005$ y 1000 generaciones, para los dos tamaños poblacionales: $N=50$ y $N=500$. Si estamos simulando el comportamiento de A1, probar definiendo solamente la tasa de mutación de A2 a A1. En 5 iteraciones, registre la cantidad de sustituciones para ambos casos. ¿Son estos resultados acordes a lo esperado?

Opcional

Para visualizar una manera de cómo pueden realizarse algunas de estas simulaciones a partir de la distribución binomial, usaremos nuevamente el entorno de programación R con el archivo "Deriva genética.rmd", en donde también podremos cambiar algunos parámetros, pero sabiendo cuál es la base del cálculo.

Práctico 4

Selección Natural y Selección Sexual

Objetivo: Familiarizarse con el concepto de selección natural y algunos métodos para detectar su acción en distintos tipos de datos.

1. Estimación de eficacia y coeficiente de selección a partir de datos de campo

El tucu-tucu de Río Negro presenta fenotipos melánicos, agutís y dorso oscuro, con frecuencia coexistiendo en suelos arenosos. El fenotipo melánico debería ser más fácilmente detectable, y por tanto sufrir mayor mortalidad por depredación. Un estudio de la población de Estancia “La Tabaré”, Depto. de Río Negro, –en donde solo coexisten las variedades melánica y agoutí– obtuvo 74 juveniles en una estación reproductiva: algunos en trampas colocadas dentro de las cuevas y otros en bolos regurgitados en torno a un nido de “Lechuza de las vizcacheras” (*Athene cunicularia*) localizado en el mismo campo¹. En la siguiente tabla, se estima la supervivencia de cada fenotipo descontando los observados en bolos de lechuza del total de juveniles de cada fenotipo²:

| Fenotipos | Total de juveniles | Sobrevivientes | Eficacia absoluta | Eficacia relativa (w) | Coefficiente de selección (s) | S (%) |
|-----------|--------------------|----------------|-------------------|-----------------------|-------------------------------|-------|
| Agutís | 46 | 38 | | | | |
| Melánicos | 28 | 19 | | | | |

Ejercicio

- Calcular la eficacia absoluta (fracción de sobrevivientes), la eficacia relativa w (cociente entre eficacia absoluta de un fenotipo y la correspondiente al fenotipo más apto).
- Calcular el coeficiente de selección (s) para cada fenotipo. ¿Qué significan en términos biológicos los coeficientes de selección calculados?

¿Qué utilidad tienen estos resultados para entender la posible mortalidad diferencial de los dos fenotipos por causa de la depredación? ¿Cuáles serían sus limitaciones?

¿Qué utilidad tienen estos resultados para entender la eficacia darwiniana general de estos fenotipos? Razone sobre las posibles limitaciones de estos resultados.

Si la mortalidad diferencial de juveniles fuese un fenómeno real, ¿cómo podría explicarse la persistencia del fenotipo melánico? ¿Cuáles estudios podrían hacerse para avanzar en la comprensión del problema?

¹ Vasquez Herrera, A. 2003. Posible depredación diferencial sobre individuos agutís y melánicos de *Ctenomys rionegrensis*, reflejada en bolos de *Athene cunicularia*. Informe de Pasantía, Licenciatura en Ciencias Biológicas, Facultad de Ciencias, Universidad de la República, 35 pp.

² Datos adicionales: De los 50 bolos estudiados, 26 contenían restos de tucu-tucus. De estos, fue posible determinar el color del pelaje en 17. El estudio mencionado incluye una prueba de X^2 con un p-valor de 0,14, que no muestra diferencias significativas en la distribución de agutís y melánicos al comparar las colectas en las cuevas con los observados en los bolos.

2. Selección Sexual

La selección sexual resulta en variación en el éxito reproductivo entre individuos del mismo sexo y típicamente actúa más fuertemente sobre los machos. Puede ser dividida en intrasexual e intersexual y, aunque la evolución de ciertos rasgos de los machos puede ser promovida exclusivamente por uno de los dos componentes, a menudo actúan en forma simultánea. La evidencia empírica sugiere que podrían actuar en direcciones opuestas y el resultado neto reflejaría el equilibrio entre esos dos procesos. Sin embargo, en muchos casos la selección intra e intersexual tienen efectos complementarios, promoviendo la expresión de los mismos rasgos en machos.

El caso de los peces anuales

Austrolebias charrua es una especie de pez anual endémica del sistema de lagunas Patos-Merín. Habitan charcos temporales en cuyos sustratos depositan huevos resistentes a la desecación. La especie presenta un dimorfismo sexual muy marcado: los machos son más grandes, tienen distinta coloración y un patrón de bandas verticales oscuras en los flancos del cuerpo. Un estudio evaluó la selección sexual sobre el tamaño corporal de los machos por medio de la elección de la hembra y la competencia entre machos (Passos et al. 2013³).

Actividad i) Preferencias de las hembras.

Se llevaron a cabo observaciones comportamentales usando el diseño del experimento que se muestra en la figura 1a (ver video). Se registró el tiempo que una hembra interactuaba con cada macho, la frecuencia de actividades de cortejo realizada por cada macho y se calculó un índice de aparencia del macho que resume la intensidad del color y el despliegue de la aleta dorsal. Los resultados se ejemplifican en la Tabla 1.

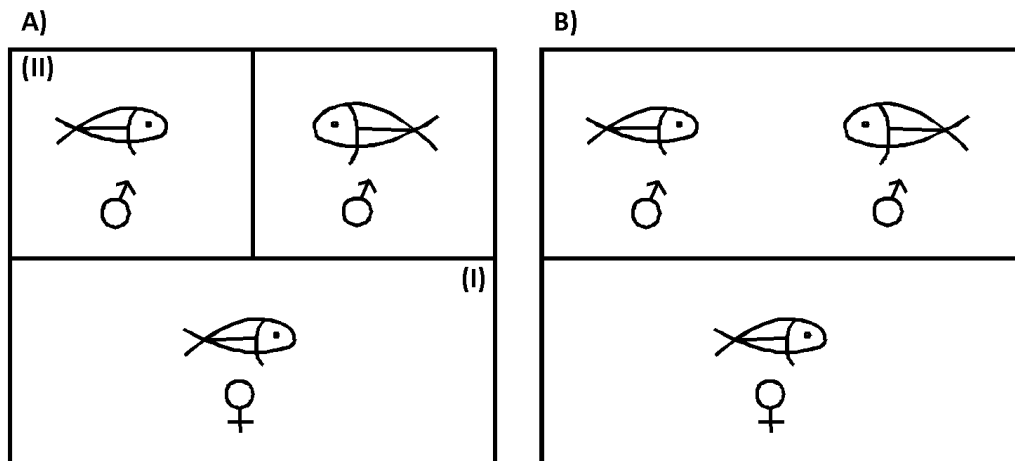


Figura 1- a) Diseño del experimento para evaluar la preferencia de la hembra. La hembra se encuentra sola en un compartimento (i) y puede ver los dos compartimentos de los machos (ii), separados por un tabique que impide que estos se vean entre ellos. b) Diseño del experimento para evaluar competencia entre machos. La hembra se encuentra sola en un compartimento y puede ver los dos machos en el compartimento contiguo.

³ Passos C, B Tassino, M Loureiro y GG Rosenthal. 2013. Intra- and intersexual selection on male body size in the annual killifish *Austrolebias charrua*. Behavioural Processes 96, 20–26.

Tabla 1. Tiempo que la hembra interactuó con cada macho (en segundos), la frecuencia de actividades de cortejo realizada por cada macho y el índice de apariencia del macho (2-6). Se muestran 3 casos y se presenta un promedio (última fila) para 30 casos.

| | Tiempo interacción | | Actividad Cortejo | | Apariencia Macho | |
|-----------------|--------------------|-------------|-------------------|--------------|------------------|--------------|
| | Grande | Chico | Preferido | No Preferido | Preferido | No Preferido |
| Hembra 1 | 388 | 192 | 0.571 | 0.909 | 4 | 4 |
| Hembra 2 | 245 | 145 | 0.833 | 0.708 | 5 | 5 |
| Hembra 3 | 284 | 435 | 0.613 | 0.751 | 5 | 4 |
| | | | | | | |
| Promedio | 813* | 425* | 0,683 | 0,641 | 5 | 4 |

*denota diferencias significativas ($p < 0,05$) entre los valores

- a) De las características estudiadas, ¿cuál/es condiciona/n la elección de la hembra? Justifique su respuesta.
 b) De ser así, ¿cuáles podrían ser las causas de dicha preferencia?

Actividad ii) Competencia entre machos.

Se llevaron a cabo observaciones comportamentales usando el diseño del experimento que se muestra en la Figura 1b (ver video). Se registraron datos de frecuencia de distintos comportamientos agonísticos, el tiempo de resolución del conflicto, la diferencia de tamaño entre los machos (dominante – subordinado) y a partir de ellos se confeccionaron las gráficas de la Figura 2.

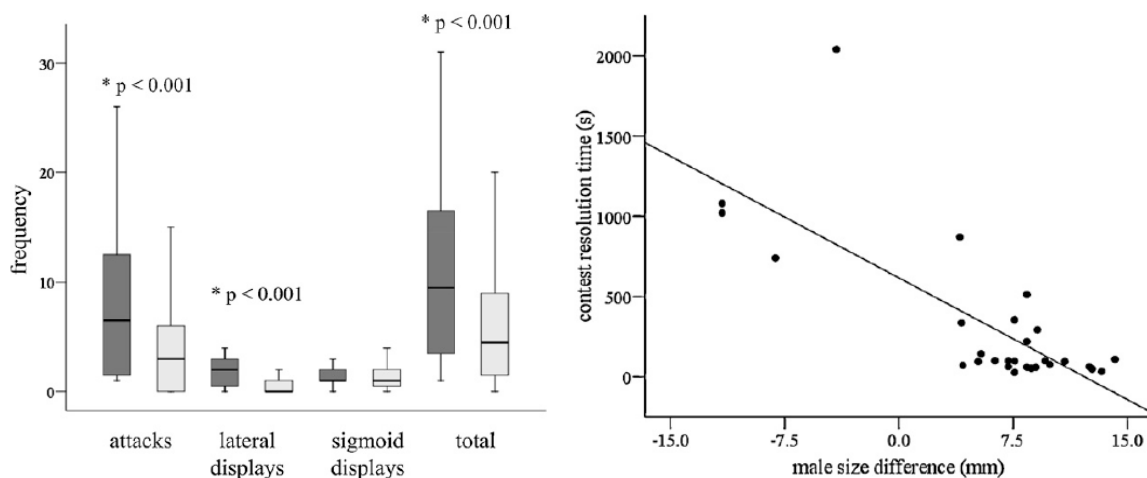


Figura 2. a) Frecuencia de los distintos comportamientos de interacción agonística. En gris oscuro se muestran los datos para los machos más grandes, y en gris claro para los más chicos de cada día. b) Tiempo de resolución del conflicto (s) en función de la diferencia de tamaño de los machos (mm).

- c) ¿Qué diferencias comportamentales se observan entre los machos dominantes y los subordinados?
 d) ¿Qué relación hay entre el tiempo de resolución del conflicto y la diferencia de tamaño entre machos?
 e) ¿Qué consecuencias fenotípicas tendrían las preferencias de las hembras a nivel poblacional?; y ¿qué otros procesos podrían modificar dichas consecuencias fenotípicas?

Práctico 5

Estructura poblacional y flujo génico

Introducción

Ctenomys rionegrensis es una de las tres especies de este género de roedores subterráneos reconocidas en Uruguay. Su distribución geográfica para nuestro país está restringida a un área de aproximadamente 60 x 50 km al suroeste del departamento de Río Negro (fig. 1).

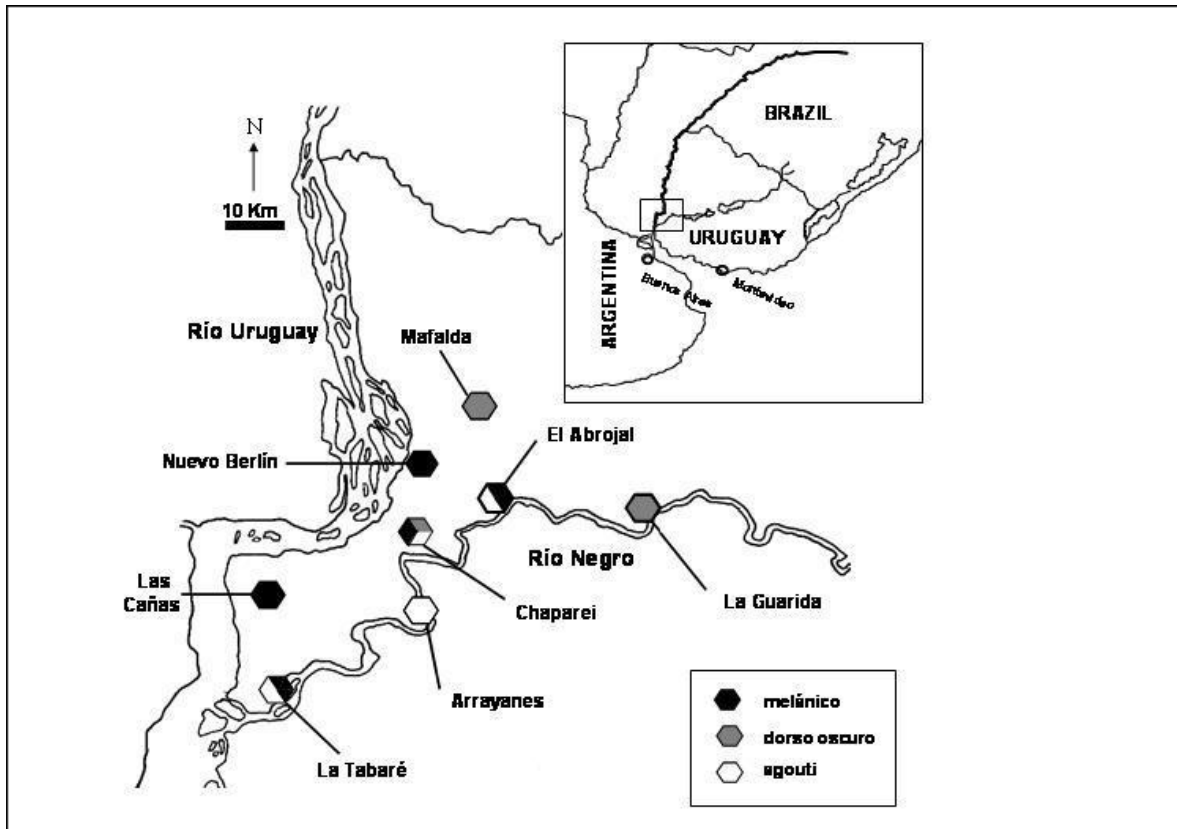


Figura 1. Distribución geográfica de *Ctenomys rionegrensis* en Uruguay. Los sitios que se detallan corresponden a las localidades de muestreo.

Pese a su pequeña distribución, existen tres coloraciones de pelaje marcadamente diferentes: melánico, agouti y dorso oscuro. Mientras que algunas de las poblaciones están enteramente constituidas por individuos que presentan la misma coloración, otras son polimórficas, presentando distintas combinaciones de pelajes.

Esta marcada diferenciación cromática de *C. rionegrensis* es llamativa por varias razones. En primer lugar, por ocurrir en un área geográfica tan reducida. Segundo, no se cumple con la regla general (que incluye a roedores subterráneos) de correspondencia entre color de pelaje y color del sustrato en el que se habita (debida probablemente a presiones selectivas impuestas por la depredación). Todas las poblaciones de esta especie se limitan a la ocupación de suelos arenosos claros, sin diferencias obvias en la vegetación, y de esta forma los individuos melánicos contrastan marcadamente con el sustrato donde viven. A diferencia del comportamiento de otros roedores subterráneos, los tucos emplean más tiempo en la superficie, dedicándose a tareas asociadas al forrajeo y dispersión, incrementando así el riesgo de depredación.

La estructura genética de las poblaciones de roedores subterráneos, como *Ctenomys rionegrensis*, estaría principalmente determinada por:

- a) su baja vagilidad, y por lo tanto bajos niveles de flujo génico, y
- b) el pequeño tamaño de sus poblaciones, que las hace más susceptibles a los efectos de la deriva genética (operando en mayor grado, fijando o eliminando alelos al azar).

Bajo estas premisas, y dado que hasta el momento no se ha encontrado ninguna posible explicación seleccionista que explique la fijación del pelaje melánico en algunas poblaciones, se ha planteado la hipótesis de una posible fijación del melanismo por deriva genética.

Esta hipótesis prevé:

- 1) Una reducción de la variación genética en las poblaciones melánicas.
- 2) Bajos niveles de flujo génico entre las poblaciones, posibilitando que la deriva se sobreponga a los efectos homogeneizadores del primero.

Métodos y Resultados

Para poner a prueba esta hipótesis, se analizaron 11 loci de microsatélites en 150 individuos de *C. rionegrensis* pertenecientes a 8 poblaciones donde se encuentran representados los tres tipos de pelaje, tanto en alopatria como en simpatria (ver mapa).

- 1) Se calculó la heterocigosis promedio observada H_o y la esperada H_e para cada población y los estadísticos F de Wright.

| Valores promedio | Chapareí | Abrojal | Guarida | Tabaré | Las Cañas | Arrayanes | Mafalda | Nuevo Berlín |
|------------------|----------|---------|---------|--------|-----------|-----------|---------|--------------|
| # alelos | 4 | 3.18 | 2.09 | 3.36 | 2.27 | 4.73 | 3.91 | 1 |
| H_o | 0.54 | 0.42 | 0.19 | 0.43 | 0.39 | 0.65 | 0.54 | 0 |
| H_e | 0.63 | 0.46 | 0.25 | 0.48 | 0.39 | 0.64 | 0.54 | 0 |
| F_{IS} | | | | | | | | |

Actividad 1) Calcular los F_{IS} para cada población.

Los niveles de flujo génico, globales y entre pares de poblaciones fueron estimados de dos formas:

a) *Niveles globales:* a partir del estadístico F_{ST} , mediante la fórmula $Nm = \frac{\left(\frac{1}{F_{ST}} - 1\right)}{4}$, obteniéndose un $Nm = 0.45$.

b) *Estimaciones pareadas:* a partir de F_{ST} calculados para cada par de poblaciones.

Para evidenciar si existe un patrón de “aislamiento por distancia”, se construyó el siguiente gráfico con los valores de flujo génico entre pares de poblaciones (obtenidos a partir de las estimaciones pareadas de F_{ST}) en función de la distancia geográfica que las separa (fig. 2). Los valores de F_{ST} fueron estimados para dos marcadores genéticos: microsatélites y el gen del citocromo b del ADN mitocondrial.

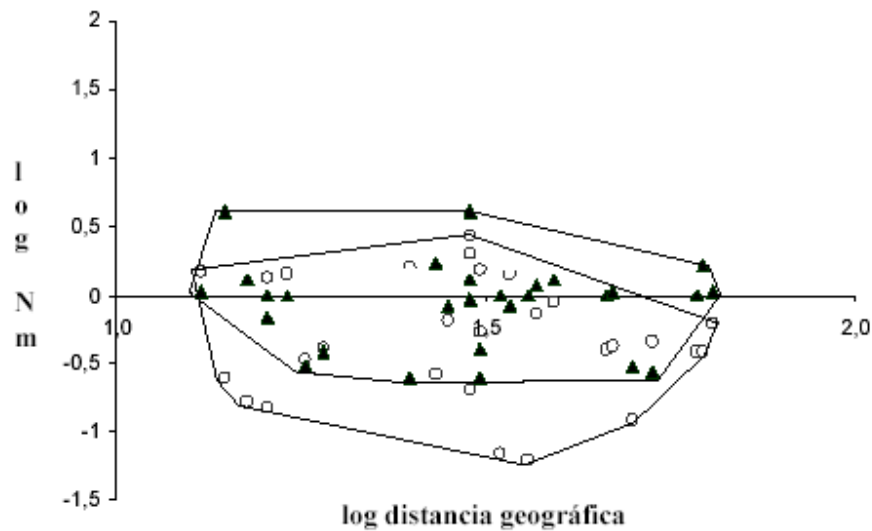


Figura 2. Gráfica del logaritmo del número de migrantes, expresado como Nm, vs. el logaritmo de las distancias geográficas. Los círculos y los triángulos denotan estimaciones basadas en microsatélites y citocromo b, respectivamente.

Además, para el gen del citocromo b se obtuvieron las distintas variantes de las secuencias (haplotipos) y se estudió su frecuencia, su relación con la procedencia de la muestra y la relación entre ellas (fig 3).

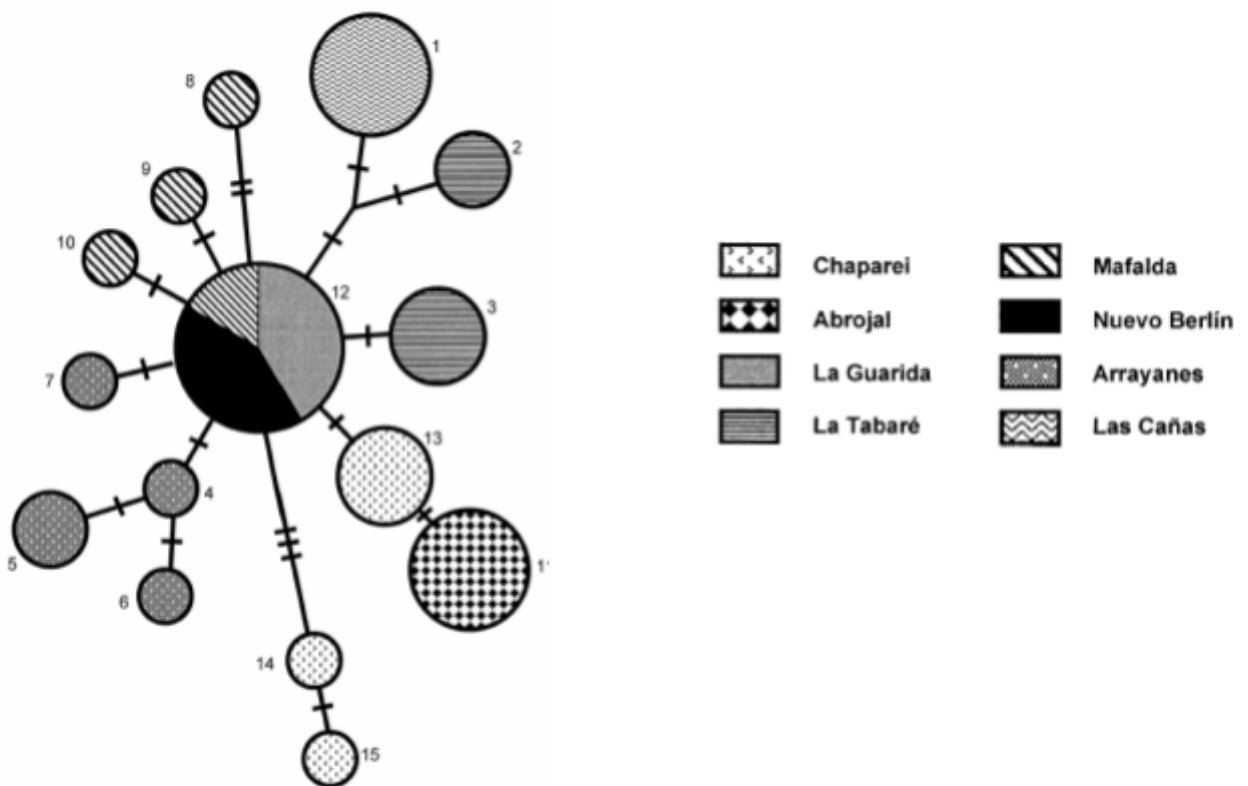


Figura 3. Red de distancias mínimas para los 15 haplotipos de citocromo b encontrados. Cada haplotipo se representa con un círculo, cuya área es proporcional a la frecuencia. Sobre las líneas que conectan haplotipos, los cambios están marcados como rayas transversales. Cada trama representa una población (ver referencia).

Actividad 2) ¿Qué sugiere el gráfico de la Fig. 2 de flujo génico en función de distancias geográficas para pares de poblaciones? (vea también la Fig. 3)

Actividad 3) ¿Las estimaciones de flujo génico obtenidas, corresponden a métodos directos o indirectos? ¿Tiene esto alguna implicancia para nuestros resultados?

Un estudio anterior que considera algunas de las mismas poblaciones y que emplea marcadores alozímicos, propone una estimación de $Nm \approx 6$ a 10 y los siguientes valores de F_{IS} :

| | Localidad | | | | |
|----------|-----------|---------|---------|-----------|--------------|
| | Abrojal | Guarida | Mafalda | Las Cañas | Nuevo Berlín |
| F_{IS} | 0,365 | 0,577 | 0,312 | 0,349 | 0,242 |

Actividad 4) Discutir a que pueden atribuirse las discrepancias entre ambas aproximaciones al mismo problema.

Actividad 5) Evaluando la totalidad de los resultados presentados, ¿acepta o rechaza la hipótesis de que el melanismo se fijó por deriva en algunas poblaciones? ¿Qué interpretación plantearías para explicar el patrón de estructura poblacional encontrado?

Bibliografía

Wlasiuk, G., Garza, J. C. y Lessa, E. P. 2003. Genetic and geographic differentiation in the Río Negro Tuco-tuco (*Ctenomys rionegrensis*): inferring the roles of migration and drift from multiple genetic markers. *Evolution*, 57 (4), pp. 913-926.

Práctico 6

Selección natural: análisis a nivel molecular

Objetivo: Familiarizarse con el concepto de selección natural y algunos métodos para detectar su acción usando datos moleculares.

Introducción

Cuando la selección natural actúa sobre las poblaciones deja huellas que pueden ser reconocidas en el ADN. Para identificar esas huellas se han desarrollado diferentes pruebas aplicables a secuencias nucleotídicas codificantes de proteínas. Una aproximación robusta y sencilla desarrollada por McDonald y Kreitman es considerar que, bajo neutralidad, la relación entre la tasa de cambio nucleotídico sinónimo (dS) y no sinónimo (dN) o de reemplazo aminoacídico será la misma dentro y entre poblaciones. Cualquier desviación sugiere un apartamiento de la neutralidad, incluyendo algún tipo de selección positiva. Si no se cuenta con información poblacional, otra aproximación muy utilizada aunque exigente es considerar que, bajo neutralidad estricta, ambas tasas deberían ser iguales, por lo que dN/dS , también conocido como ω será 1. Si dN supera ampliamente dS, es decir si $\omega > 1$, se asume que actuó selección positiva (el caso inverso, $\omega < 1$ indicaría selección purificadora). En este práctico aplicaremos ambas aproximaciones.

Ejercicio 1 - Test de McDonald y Kreitman

Cuando McDonald y Kreitman en 1991 propusieron su test de neutralidad, lo aplicaron a un conjunto de secuencias de la enzima Alcohol deshidrogenasa (Adh), para tres especies diferentes.

Actividad

La Tabla 1 muestra el resumen de los sitios variables de las secuencias de Adh incluidas en la base de datos.

- Interprete la Tabla 1: ¿qué hay en las filas y las columnas?
- Completar los siguientes cuadros a partir de los datos de la tabla.
- Utilizando los cuadros completados, realizar el test de McDonald y Kreitman (MK) y sacar conclusiones.

| | <i>D. simulans vs. D. yakuba</i> | |
|-----------|----------------------------------|---------------|
| | Polimorfismos | Sustituciones |
| Reemplazo | | |
| Sinónimos | | |

| | <i>D. melanogaster vs. D. yakuba</i> | |
|-----------|--------------------------------------|---------------|
| | Polimorfismos | Sustituciones |
| Reemplazo | | |
| Sinónimos | | |

| | <i>D. melanogaster vs. D. yakuba</i> | |
|-----------|--------------------------------------|---------------|
| | Polimorfismos | Sustituciones |
| Reemplazo | | |
| Sinónimos | | |

Ejercicio 2 - Variación en el ω entre sitios y entre linajes

En este ejercicio usaremos las secuencias de la Hemaglutinina (HA) del virus de la gripe. Esta proteína es una glucoproteína antigénica que se encuentra en la superficie del virus y es la mayor responsable de la unión del virus a la célula infectada. Esta proteína es muy estudiada en el diseño de vacunas, porque presenta una evolución asimétrica que sugiere una fuerte selección de aquellas variantes que son las que mejor escapan al sistema inmune del hospedero. Además, el análisis de los cambios nucleotídicos sinónimos y no sinónimos muestra que muchos residuos aminoacídicos en la HA concentrados en el extremo distal y externo de la proteína (que son aquellos sitios que interactúan con el sistema inmune del hospedero) están siendo seleccionados positivamente para cambiar (Fig. 1).

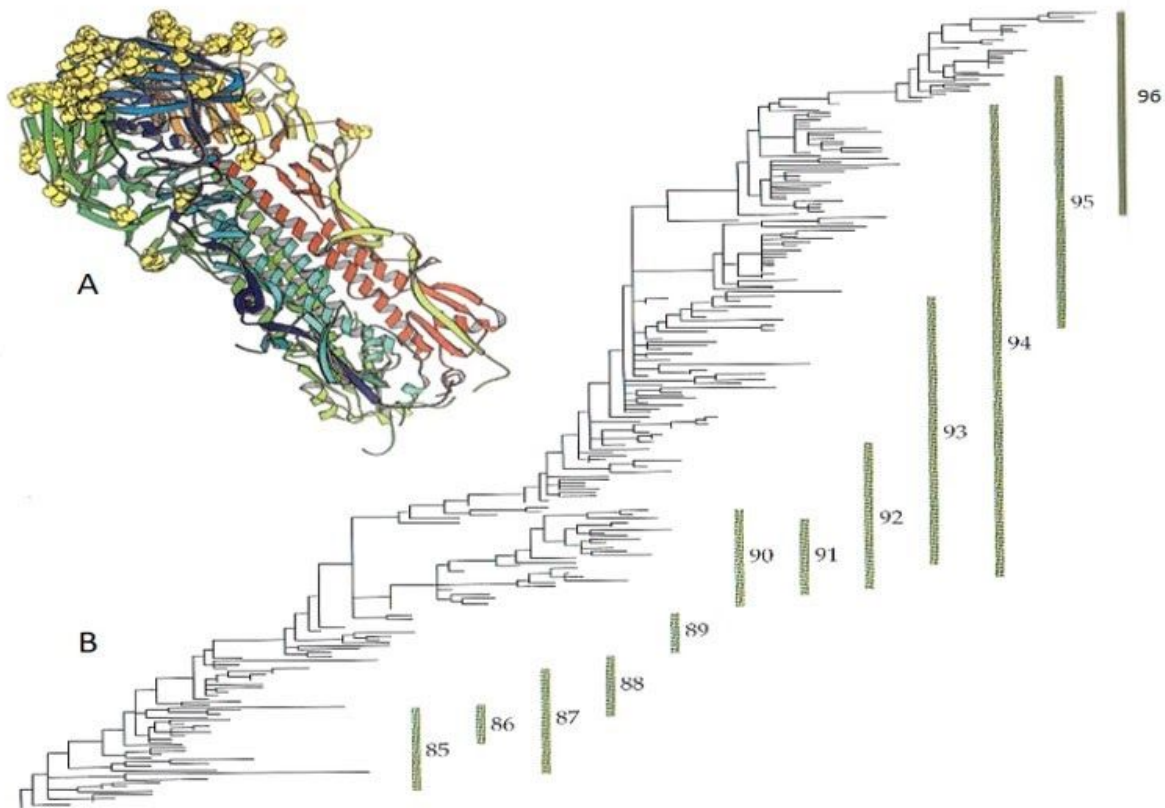


Fig 1. A) Modelo tridimensional de la Hemaglutinina del virus de la gripe, mostrando los sitios aminoacídicos seleccionados positivamente para cambiar. B) Filogenia de cepas del virus aisladas desde 1985 a 1996, basada en el análisis de las secuencias nucleotídicas de ese gen (Tomado de Hillis, 2009).

A continuación, estimaremos la tasa de cambio nucleotídico sinónimo (dS) y no sinónimo (dN) y su relación mediante máxima verosimilitud (ML). Estas estimaciones pueden ser realizadas para cada codón y/o linaje, considerando una filogenia que permita estimar las secuencias de los nodos ancestrales. Utilizaremos una aproximación asociada al MEGA llamada SLAC (Single-Likelihood Ancestor Counting) implementada en el servidor Datamonkey (<https://www.datamonkey.org/>) donde se pueden implementar otros algoritmos.

SLAC es una de las tantas formas de evaluar sitios bajo selección. Utiliza una combinación de enfoques de ML y conteos para inferir dN y dS por sitio para un conjunto de secuencias nucleotídicas codificantes de proteínas alineadas y la filogenia que las vincula. En este caso, cargaremos al programa la información de las secuencias, y el árbol será estimado por el propio programa. SLAC comienza optimizando las longitudes de las ramas y los parámetros de sustitución de nucleótidos bajo un modelo complejo (denominado MG94xREV). Sin embargo, en lugar de usar ML para ajustar los parámetros dN y dS específicos del sitio, SLAC usa ML para inferir la secuencia ancestral más probable en cada nodo de la filogenia. Así, compara secuencias nucleotídicas entre nodos adyacentes, y luego cuenta directamente el número total de cambios no sinónimos y sinónimos que se han producido en cada sitio (para eso emplea una versión modificada del método de recuento Suzuki-Gojobori). La significancia estadística se determina en cada sitio utilizando una distribución binomial extendida. Es importante destacar que, debido a su enfoque basado en el recuento, es posible que SLAC no sea preciso para conjuntos de datos con altos niveles de divergencia. A su vez, esta aproximación asume que la presión de selección para cada sitio es constante a lo largo de toda la filogenia, algo poco realista. Existen otros métodos más flexibles y sensibles que pueden ser explorados en el servidor Datamonkey.

Actividad

- Abrir el programa MEGAX, y dentro del programa abrir el archivo SecuenciasHA.meg.
- Ir a la opción "SELECTION" en el menú principal, donde se pueden implementar algunos tests para identificar sitios bajo selección.
- Marcar la opción DATAMONKEY (donde se aplicará la opción "codon by codon selection by SLAC"). Una vez en esta opción no hay que especificar el archivo (porque ya fue cargado) pero sí algunas opciones, como el código genético (Universal) y los sitios a usar (all sites). El resultado demorará unos minutos. En caso que haya dificultades puede ver los resultados tal y como se verían en MEGAX, en el siguiente enlace: <http://datamonkey.org/slac/5f3d5c21503f2f234815f7ff>.
- Interprete los resultados. ¿Existe algún sitio y/o linaje seleccionado? ¿Qué sitio presenta una fuerte evidencia de selección positiva? ¿Qué valores de dN y dS tiene ese sitio? ¿Qué cambios aminoacídicos se registran en ese sitio? ¿Dónde cree que se ubicará ese sitio en la proteína dados los antecedentes planteados?
- ¿Qué similitudes y diferencias encuentra entre esta aproximación y el test de MK?
- Explore otros programas implementados en Datamonkey.org. En este enlace verá el análisis de el mismo conjunto de secuencias pero habiendo aplicado otro método para inferir selección a nivel molecular denominado MEME (Mixed Effects Model of Evolution), que permite estimar sitios y ramas bajo selección simultáneamente (<https://www.datamonkey.org/meme/62f3ebed955d162e04ea09bd>).

Referencias

- McDonald, J., Kreitman, M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351, 652–654 (1991). <https://doi.org/10.1038/351652a0>
- Hillis, DM. (2009). Phylogenetic Progress and Applications of the Tree of Life. En: *Evolution since Darwin: The First 150 Years*, pp. 421-449. Eds: MA Bell, DJ Futuyma, WF Eanes, JS Levinton. Sinauer Associates, Inc. • Publishers Sunderland, Massachusetts U.S.A.
- Kosakovsky Pond, SL and Frost, SDW. "Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection." *Mol. Biol. Evol.* 22, 1208--1222 (2005).
- Murrell, B et al. "Detecting individual sites subject to episodic diversifying selection." *PLoS Genetics* 8, e1002764 (2012).

Práctico 7

Patrones de evolución molecular

Objetivos: A partir del análisis del patrón de sustituciones nucleotídicas de una secuencia codificante en un grupo taxonómico particular: 1) visualizar patrones generales de evolución molecular y 2) discutir la validez y el alcance de la idea de “reloj molecular”, identificando factores que pueden producir desviaciones aparentes del mismo.

Datos.

El archivo Primates_datos.meg contiene los 1000 primeros sitios del gen del citocromo b del ADN mitocondrial de 13 especies de primates.

Clasificación de los primates considerados en este práctico

| Suborden | Familia | Género | Nombre común | Distribución |
|---------------|-----------------|---|--|---|
| Strepsirrhini | Lemuridae | <i>Lemur</i> <i>Microcebus</i> | Lemur | Madagascar |
| Platyrrhini | Cebidae | <i>Cebus</i> <i>Saimiri</i> | Mono capuchino Mono ardilla | Neotrópico |
| Catarrhini | Cercopithecidae | <i>Papio</i> <i>Macaca</i> | Babuino Macaco | África |
| | Hylobatidae | <i>Hylobates</i> | Gibón | Asia, Indonesia |
| | Hominidae | <i>Pongo</i> <i>Gorilla</i> <i>Pan</i> <i>Homo</i> | Orangután Gorila Chimpancé Humano | Borneo África África Cosmopolita |

Tiempos de divergencia

Datos paleontológicos sugieren los siguientes tiempos de divergencia⁴ desde el ancestro común (dados en millones de años desde el presente):

58 Lemúridos vs. los restantes primates

40 Platyrrinos vs. Catarrinos

15 Orangután vs restantes homínidos

6 Gorila vs. Chimpancé y Humanos

Actividades

1) Usar el programa Mega X

Una vez abierta la base de datos “primates_datos.meg” en el programa, realizar las siguientes actividades.

- ¿Por qué será útil indicarle al programa el carácter codificante de la secuencia? ¿y que el origen de la secuencia sea ADN mitocondrial de mamíferos?

⁴ En la discusión sobre el reloj molecular y temas relacionados se habla de “divergencia” para referirse al cambio total que ha ocurrido en la evolución de dos especies desde su ancestro común. Este cambio se cuenta, por tanto, a lo largo de dos líneas evolutivas; bajo la hipótesis del reloj molecular la “tasa de divergencia” de un gen o región cualquiera es el doble que la “tasa de evolución”.

- ¿Las secuencias aminoacídicas son más o menos informativas que las secuencias nucleotídicas?
- Obtener una filogenia usando el criterio de Máxima Parsimonia (utilizando las opciones que vienen por defecto). Definir como grupo externo a Lemuridae, reportar el índice de consistencia (en $i >$ general). ¿Qué información aporta este índice acerca de la filogenia?
- Representar el árbol anterior como filograma (por defecto aparece un cladograma). Reportar si existen diferencias entre grupos en la tasa de evolución y reflexionar las posibles causas que pueden producirlas.
- Obtener una tabla de distancias absolutas pareadas. Escoger en el menú la opción Distances, Compute Pairwise y elegir la opción Model / Nucleotide / No. of Differences. Visualizar las otras opciones.
- Observar la copia de la matriz obtenida anteriormente que se encuentra a continuación. Luego: a) Completar la información ausente, b) en la matriz reconocer los recuadros para las 2 comparaciones con que se cuenta con información paleontológica.

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|----|-------------------------------|-----|-----|-----|-----|-----|-----|-----|---|---|-----|-----|----|----|
| 1 | <i>Lemur_catta</i> | | | | | | | | | | | | | |
| 2 | <i>Microcebus_griseorufus</i> | 170 | | | | | | | | | | | | |
| 3 | <i>Macaca_mulatta</i> | 263 | 281 | | | | | | | | | | | |
| 4 | <i>Papio_hamadryas</i> | 262 | 278 | 135 | | | | | | | | | | |
| 5 | <i>Cebus_albifrons</i> | 245 | 265 | 255 | 266 | | | | | | | | | |
| 6 | <i>Saimiri_sciureus</i> | | | | | | | | | | | | | |
| 7 | <i>Hylobates_agilis</i> | 254 | 263 | 203 | 209 | 249 | 246 | | | | | | | |
| 8 | <i>Hylobates_lar</i> | 251 | 262 | 206 | 215 | 258 | 243 | 55 | | | | | | |
| 9 | <i>Pongo_pygmaeus</i> | 238 | 276 | 198 | 207 | 248 | 255 | 184 | | | | | | |
| 10 | <i>Gorilla_gorilla</i> | 239 | 262 | 190 | 208 | 245 | 240 | 163 | | | | | | |
| 11 | <i>Homo_sapiens</i> | 241 | 270 | 199 | 213 | 253 | 253 | 164 | | | 127 | | | |
| 12 | <i>Pan_paniscus</i> | 248 | 275 | 188 | 197 | 224 | 243 | 159 | | | 117 | 112 | | |
| 13 | <i>Pan_troglodytes</i> | 239 | 276 | 189 | 201 | 225 | 236 | 164 | | | 120 | 115 | 50 | |

2) Usar el programa Excel

- Abrir el archivo "distancias primates.xls". Encontrarás el número de diferencias discriminadas entre las posiciones del codón, así como entre transiciones y transversiones, obtenidas de la forma anterior. Estas distancias pareadas se han graficado para cada uno de los tiempos de divergencia. Observe los rangos de valores para cada una de estas medias, y saque conclusiones de las dos gráficas.
- Pensar, discutir y responder: para el caso de la cantidad de cambios según las posiciones del codón, ¿qué gráfico esperarías obtener según el reloj molecular? ¿Se ajustan en apariencia las gráficas a la idea del reloj molecular? ¿Qué factores pueden dar cuenta de las variaciones observadas? ¿Cómo pueden explicarse estas tasas en términos neutralistas?
- De ser posible, estimar el tiempo de divergencia de los gibones y los homínidos.

3) Mencione los problemas asociados a la estimación del tiempo de divergencia entre gibones y homínidos, realizada anteriormente.

De encontrarse un pseudogen del citocromo b para estas especies, discuta cómo espera que sea el patrón de sustituciones nucleotídicas en esta secuencia.

Bajo neutralidad, ¿es esperable obtener un gen, o región de un gen, con mayor cantidad de cambios no sinónimos que sinónimos?, ¿qué interpretación podría darle a este fenómeno?

Práctico 8

Familias multigénicas - “Globinas”

Objetivo: Mediante el estudio de los genes de Globinas en primates, familiarizarse con la identificación de familias multigénicas, su problemática y limitaciones.

Una familia génica o familia multigénica es un grupo de loci cromosómicos cuya secuencia de nucleótidos es similar y derivan de una secuencia común ancestral. Puede incluir copias de genes ligeramente diferentes y/o pseudogenes más variables, en uno o varios cromosomas.

Los genes codificantes para las globinas representan un ejemplo clásico de familia multigénica. En general, estas proteínas portan un grupo “heme” y se caracterizan por unirse y transportar oxígeno. Presentan dominios homólogos en varios taxa: a) hemoglobinas tetraméricas de vertebrados (componente proteico principal de los eritrocitos), b) flavohemoglobinas en microorganismos, c) hemoglobinas homodiméricas bacterianas, d) leghemoglobinas en plantas (asociadas al metabolismo del nitrógeno en plantas con rhizomas), e) hemoglobinas de invertebrados, f) mioglobinas monoméricas usualmente encontradas en el músculo animal entre otras.

En primates existen varias clases de globinas dentro de la familia, generalmente designadas con letras griegas (por ejemplo, α , β , γ , δ , ϵ , ξ y Mioglobina MB). Recientemente se ha identificado una nueva clase, la citoglobina (CYGB). En humanos, el gen que la codifica parece expresarse en todos los tejidos y se localiza en el brazo largo del par cromosómico 17. Se sugiere que su tamaño está conservado en varios mamíferos y es de 190 aminoácidos.

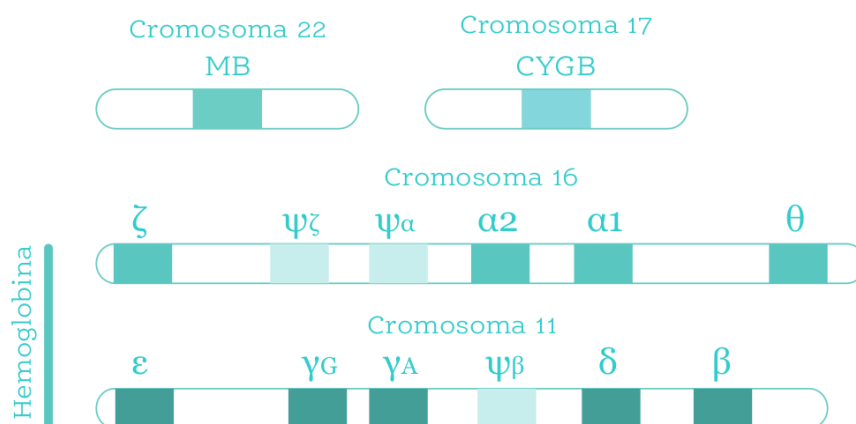


Figura 1 - Localización de los genes codificantes para las globinas en humanos.

Conceptos generales

- ¿Qué procesos podrían estar involucrados en la aparición de las familias multigénicas?
- ¿Qué ventajas conferirían estas familias génicas?

Análisis del alineamiento de secuencias proteicas de globinas, identificación de regiones conservadas

Cargue el alineamiento en el programa MEGA X: Ir a **Align** → **Edit/Build Alignment** → **Retrieve a sequence from a file** → cargar el archivo “**globinas_aa.fas**” y alinear: Ir a **Alignment** y elegir cualquiera de las opciones (ClustalW o MUSCLE).

- Identifique los sitios conservados en todas las secuencias y sugiera su posición en la proteína. Justifique.
- ¿Visualiza algún patrón particular en la secuencia de estas proteínas? Interprete.
- Selecciones y copie (ctr + c) la región conservada (al menos 50 aa.) en una de las secuencias y haga un blastp (blast de sec. aminoacídicas) contra la base de datos del genbank (<http://blast.ncbi.nlm.nih.gov/>). ¿A qué corresponde dicha región?
- Opcional : Una forma de identificar a que corresponden dichas regiones es mediante la base del genbank, “Conserved Domain Database (CDD)” (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). La búsqueda de dominios similares en tu secuencia está automáticamente incorporada al blastp.

Reconstrucción filogenética de las globinas

Cargue los datos en el programa MEGA (**File / Open Data / globinas_nt.meg**). Esta base de datos tiene las secuencias codificadas (ADNc) reportadas de genes miembros de la familia de las globinas en los siguientes primates: *Homo sapiens*, *Gorilla gorilla*, *Pongo abelii*, *Macaca mulatta*, *Callithrix jacchus*, *Papio anubis*, *Pan troglodytes*, *Microcebus murinus* y *Otolemur garnettii* (ver Figura 2).

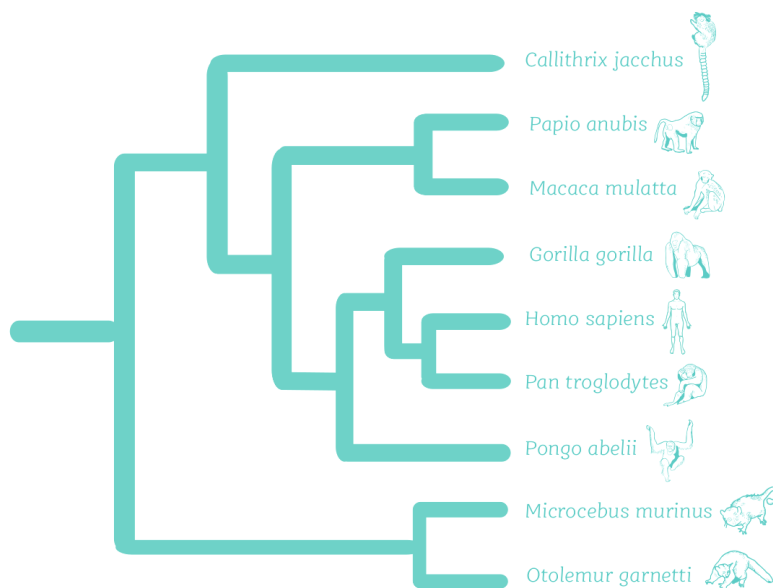


Figura 2 - Cladograma de las especies incluidas los datos

- En base a los datos disponibles (ver Fig 1 y 2), ¿cómo piensa usted que ocurrió la evolución de los diferentes tipos de globinas dentro de los primates? Genere un esquema.

Realice la reconstrucción filogenética en base a los datos que se le brindan (Método: *Neighbor-Joining*, Gaps: *Pairwise deletion*, Bootstrap: *1000 pseudoréplicas*).

- ¿Son los resultados coincidentes con el esquema que usted generó?
- Identifique aquellos nodos que se corresponden con eventos de especiación y/o duplicación.
- Analice e interprete algunos casos particulares:
 - ◆ clase delta en humanos
 - ◆ ausencia de γ -globina en *M. murinus* y *O. garnettii*
 - ◆ ausencia de μ -globina en *M. murinus*
 - ◆ relación de las citoglobinas con las otras globinas
- Si le interesara, ¿en cuál cromosoma buscaría la μ -globina en humanos?
- ¿Qué información adicional sería útil para establecer el origen de un nuevo gen de globinas?

Estimación del parámetro ω (dN/dS) entre clases y dentro de las clases

Los genes ya están agrupados en las clases correspondientes en la base de datos. Calcular las distancias media dentro (ortólogos) y entre clases (parálogos).

En MEGA... Ir a *Distances* → *Compute within group mean* y luego *Compute between group mean*, incluyendo sustituciones sinónimas (dS) y no-sinónimas (dN), estimadas con el modelo de Kumar (*LogDet (Tamura-Kumar)*) y Gaps: *Pairwise deletion*.

- ¿Cuál espera sea el resultado dentro y entre clases? ¿Por qué? ¿Qué relación tiene esto con el árbol reconstruido? ¿Cómo piensa que será la estimación de la tasa sinónima y por qué?
- A partir de los valores de distancia media sinónima y no-sinónima calcule la relación entre ellas (\approx dN/dS) dentro de las clases y entre clases, y tome nota.
- Interprete ambos resultados, relacione los mismos con procesos evolutivos dados en clase.

Práctico 9

Genómica comparada

Objetivo: Mediante la comparación de dos genomas completos de bacterias cercanamente emparentadas, una infecciosa y otra no, se pretende destacar las potencialidades de la genómica comparada como herramienta para poner a prueba hipótesis en biología evolutiva.

Listeria monocytogenes es un patógeno intracelular causante de listeriosis, tanto epidémica como esporádica. Es una enfermedad transmitida por el alimento, particularmente peligrosa en recién nacidos, ancianos, embarazadas y pacientes inmunodeprimidos. Una vez ingerida, se dirige desde el lumen intestinal hacia el sistema nervioso y a la placenta, causando meningitis, meningo-encefalitis, septicemia, aborto espontáneo, infecciones perinatales y gastroenteritis. Se ha encontrado en varios animales domésticos y en un amplio rango de condiciones (temperatura, salinidad, etc.), incluso extremas. Poco se conocía sobre las bases moleculares de su patogenicidad, hasta que en 2003 fue secuenciado completamente su genoma y el de una especie cercanamente emparentada no patógena, *L. innocua*.

Buchrieser y cols. (2003), plantean que la patogenicidad de *L. monocytogenes* se debe a la adquisición (por transferencia horizontal), de genes de proteínas de superficie, en particular internalinas, InlB (necesarias para entrar a las células eucariotas) o ActA, que juega un papel clave en la movilidad basada en actina. Así mismo, ambas especies presentan un conjunto de genes homólogos a los de *Salmonella* que les permitirían tener metabolismo anaeróbico.

Familiarización con el programa

Descargar el programa Artemis Comparison Tool (ACT) del sitio web del Instituto Sanger (<http://sanger-pathogens.github.io/Artemis/ACT/>) y elegir la opción de acuerdo a su sistema operativo (Windows, Linux, Mac). Este programa requiere de la plataforma Java para poder ejecutarse. Descargar la versión actualizada de Java del siguiente sitio: <https://www.oracle.com/java/technologies/javase-jdk15-downloads.html>.

Abrir el programa Artemis Comparison Tool (ACT) haciendo click derecho sobre el archivo [act.jar](#), “abrir con” e indicar la aplicación Java. Cargar los siguientes tres archivos en el siguiente orden:

Ir a File / Open...

Sequence file 1: [L_innocua.embl](#) - genoma de *L. innocua*,

Comparison file 1: [comp_innocua_vs_monocytogenes.txt](#) - comparación entre ambos genomas

Sequence file 2: [L_monocytogenes.embl](#) - genoma de *L. monocytogenes*

Estas secuencias con su descripción se pueden obtener de internet en forma gratuita.

Una vez abierto el programa ACT reconocer las diferentes regiones de la interfase: la secuencia de *L. innocua* aparece arriba, la comparación en el medio y la secuencia de *L. monocytogenes* abajo. Utilizando los cursores verticales y horizontales es posible acercar y desplazar. Probar la opción "Locked" para una mejor visualización (botón derecho sobre la zona media).

Trabajar con una de las secuencias, por ejemplo, la de *L. innocua*. Se sugiere realizar la siguiente secuencia de pasos:

- Poner el máximo de aumento y visualizar los tres marcos de lectura. Con las opciones del botón derecho del mouse (BD), active y desactive diferentes funciones (Ej.: codones stop).
- Desactivar la anotación automática (BD, [Entries](#), desmarcar el archivo *Nombre de la especie.embl*), y moverse sobre la secuencia.
- Identificar los marcos abiertos de lectura (Open Reading Frames: ORFs).
- Verifique que los ORF por Ud. identificados se corresponden con genes ya anotados, volviendo a activar el archivo .embl. Identificar la orientación de los mismos. Cuente el número de genes en cada organismo ([View / L_innocua / Overview](#)).

¿Qué mecanismos podrían explicar la diferencia en el número de genes entre ambas especies?

- Seleccione uno de esos genes y vea sus propiedades (nombre, función, largo, etc.) (BD, [View](#), [View Selected Features](#)).
- Usar la opción [Graph](#) del menú principal para graficar el contenido GC ([GC content %](#)) en una de las secuencias y luego el [GCskew](#) o [GCdeviation](#)

¿Qué factores evolutivos pueden dar cuenta de los patrones observados? ¿Cómo pueden afectar estos sesgos mutacionales a las reconstrucciones filogenéticas o los estudios de evolución molecular?

- Buscar algún gen particular, por ejemplo *Imo0459* en *L. monocytogenes* (BD, [Goto](#), [Navigator, feature with gene name](#)).

Trabajar con secuencias la comparación entre las secuencias

En esta etapa, puede ser útil eliminar los marcos de lectura (Menú principal, Display). Se sugiere comenzar por el principio de las secuencias, a un nivel medio de aumento.

- Visualizar códigos de colores rojo (sintenia: se mantiene el orden y orientación relativa de genes ortólogos), azul (regiones invertidas), blanco (sin homología), amarillo (selección del usuario).
- Identificar genes especie específicos e invertidos.

¿Cómo diría que es el grado de sintenia entre ambas especies? ¿Tiene esto algún significado biológico?

→ Establezca algunos genes especie específicos y agregue la mayor cantidad posible de información que los caractericen.

¿Tienen estos genes algún significado biológico en relación con la diferente patogenicidad de las especies?

¿Qué tipo de información puede ser útil para saber si ha ocurrido transferencia horizontal de esos genes? ¿Qué otras formas de adquisición de genes serían posibles?

El programa ACT acepta la inclusión de otros genomas y comparaciones. ¿Qué información puede aportar al estudio el hecho de incorporar un grupo externo al análisis?

Con las herramientas y datos disponibles, ¿aceptaría o rechazaría la hipótesis planteada por Buchrieser y cols. (2003)?

Material suplementario

MedlinePlus Enciclopedia Médica Listeriosis (información de la enfermedad):
<http://www.nlm.nih.gov/medlineplus/spanish/ency/article/001380.htm>

Trabajo original: Buchrieser C, Rusniok C, The Listeria Consortium, Kunst F, Cossart P and Glaser P. 2003. Comparison of the genome sequences of *Listeria monocytogenes* and *Listeria innocua*: clues for evolution and pathogenicity. FEMS Immunology and Medical Microbiology, 35: 207-213.