

Sistemática filogenética  
Introducción a la práctica.

Segunda Edición

Daniel Rafael Miranda-Esquivel,  
Joan Salvador Arias,  
Ivonne Janeth Garzón-Orduña

Versión: 2016.01.01 26 de abril de 2016

# Índice general

Introducción general	I
1. Selección de caracteres	1
2. Matrices de datos	14
A. Programas de cómputo	26
A.1. Editores y manejadores de búsqueda . . . . .	28
A.1.1. WinClada . . . . .	28
A.1.2. MacClade & Mesquite . . . . .	31
A.2. Búsqueda de árboles . . . . .	36
A.2.1. TNT . . . . .	36
A.2.2. NONA . . . . .	39
A.2.3. POY . . . . .	40

A.2.4. PAUP* . . . . .	43
A.2.5. MrBayes . . . . .	45
B. Formatos de archivos . . . . .	47
B.1. Formatos de matrices . . . . .	47
B.1.1. NONA . . . . .	47
B.1.2. PAUP* . . . . .	49
B.1.3. POY . . . . .	51
B.2. Formatos de árboles . . . . .	51
B.2.1. NONA . . . . .	51
B.2.2. PAUP* . . . . .	52
C. Algunos comandos básicos para POY . . . . .	54
C.1. Análisis de sensibilidad con costos diferenciales . . . . .	54



# Introducción general

Este libro ha sido pensado como uno de los tantos soportes posibles para las clases de sistemática de nivel básico y medio, además de servir de repaso a conceptos teóricos generales pero sobre la base empírica, y no como reemplazo de los libros básicos o avanzados sobre análisis filogenético.

## ¿Cómo está estructurado este manual?

El libro consta de 12 prácticas que van desde manejo de caracteres, pasando por editores de matrices y árboles a búsquedas tanto para análisis de parsimonia y máxima verosimilitud (ML), como para análisis bayesiano. En todos los casos se presentan

las técnicas, la metodología a seguir, los programas de cómputo a usar (y sus comandos), además de una serie de preguntas sobre la práctica o en general sobre la técnica. La literatura recomendada es una sugerencia de lecturas, desde el punto de vista de los autores, críticas, pero obviamente no cubre todos los artículos posibles; exploraciones constantes de revistas como *Cladistics*, *Systematic Biology*, *Zoologica Scripta* o *Molecular Phylogenetics and Evolution* y similares, actualizarían las perspectivas aquí presentadas. Al final se presenta un capítulo que trata sobre los programas usados, que se espera funcione como una guía rápida para el uso de los mismos pero que no reemplaza al manual.

A lo largo del libro se utiliza una tipografía consistente para indicar el nombre de los programas (v.g., `TNT`, `Component`), las instrucciones que deben ser escritas en los programas de línea de comando, por ejemplo:

```
> mult=replic 10;
```

y si se accede las instrucciones mediante un menú o un cuadro de diálogo en los programas de interfaz gráfica (v.g., *Anali-*

ze/heuristics).

El orden de las prácticas obedece a la estrategia de enseñanza seguido en la UIS pero puede usarse en otras secuencias, por ejemplo, las prácticas caracteres, excluyendo alineamiento, seguido de búsquedas incluida la búsqueda del modelo, consensos, por último soporte y al final discutir la práctica de alineamiento.

Por la misma naturaleza del análisis filogenético, la práctica más larga es la de búsquedas, es factible que todos los ejercicios no puedan completarse en una única sesión de laboratorio.

Usted podrá encontrar material adicional, los datos, algunos macros para los distintos programas y demás chismes en el sitio web del laboratorio de Sistemática & Biogeografía (LSB) de la UIS, en la dirección: [ciencias.uis.edu.co/labsist](http://ciencias.uis.edu.co/labsist).

## Público objeto

Se espera que el usuario de este manual tenga conocimientos básicos, o que esté tomando un curso formal de sistemática filo-

genética a nivel de pre o posgrado. No se esperan características especiales en cuanto al dominio de computadores, pero el manejo básico de un editor de texto que permita grabar archivos sin formato es muy recomendable, además de conocer el entorno de línea de comandos, el cual es usado en una gran cantidad de programas. Los usuarios pueden trabajar en cualquier plataforma desde Linux a MacOSX, pasando por Windows; en general se citan los programas apropiados para cada plataforma.

El formato de la bibliografía ha sido modificado para que en la medida de lo posible, muestre los DOI (Digital Object Identifier) para que la búsqueda de la bibliografía básica sea más fácil.

## Programas eliminados

Algunos programas han sido eliminados, por ejemplo Piwe, Modeltest o component, y algunos se presentan al mínimo, en particular NONA, dado el poco uso actual. Modeltest y MrModeltest han sido superados por JModeltest y PAUP\* en modo nativo, por lo que no es muy factible que se mantengan



en uso, solo *Modeltest* es interesante para aprender sobre los comandos de *PAUP\**.

## Programas adicionados

El área más activa en términos de programas ha sido análisis bajo Verosimilitud y Bayesiano, se han incluido los programas: *RaxML/ExaML*, *PhyML*, *JModeltest* y *R*; en especial para este último se dan ejemplos detallados para familiarizar al estudiante con la plataforma.

## Agradecimientos

A la Universidad Industrial de Santander, en particular a la Escuela de Biología. A los estudiantes del programa de pregrado de la escuela de Biología de la UIS (Universidad Industrial de Santander) estudiantes profundización, a Viviana Romero quien reescribió casi por completo el capítulo de búsquedas.



# Práctica 1

## Selección de caracteres

### Introducción

En un análisis filogenético cladístico es necesario identificar y usar caracteres homólogos, de hecho, la importancia de los caracteres es tal que Hennig (1968) se refería a los especímenes estudiados como semaforontes (o los que llevan caracteres). Aún a pesar del papel central del concepto de carácter, la definición no es tan fácil.

Richards (2003) señala que el término carácter no está bien definido y que su reconocimiento depende en gran parte del entrenamiento del taxónomo (Richards, 2002). Es por eso que la mayoría de las ocasiones se enfatiza en la importancia de un análisis de caracteres profundo y concienzudo (ver Hennig (1968); Neff (1986); Rieppel & Kearney (2001)).<sup>1</sup>

Las ideas de carácter y homología usadas aquí son independientes del tipo de caracteres usados, pero es mucho más sencillo presentar la discusión en términos de caracteres morfológicos. Más adelante se tratará directamente el tema de los caracteres moleculares.

Aunque no existe una fórmula mágica para reconocer caracteres homólogos, la principal idea es la similitud "especial", basada en la definición de homología como similar dada por ancestría común ((Rieppel & Kearney, 2001); contra Kluge (2003); Grant & Kluge (2004)). Por similitud no debe entenderse solo el parecido general, sino que existe una correlación estructural y de posición (la similitud y conjunción en el sentido de Patterson

---

<sup>1</sup>Este análisis de caracteres también implica revisar las afirmaciones cuantitativas o cualitativas hechas, ver Wiens (2001).

(1981)) del carácter en los taxa comparados.

Por ejemplo, los brazos humanos, las alas de los murciélagos y las patas de los caballos presentan una estructura ósea similar, con diversos huesos colocados en las mismas posiciones, y en general el miembro completo está en la misma posición con respecto al cuerpo en los tres taxa, aunque sus formas son muy diferentes.

Otro factor importante a tener en cuenta son los "grados" o niveles de generalidad de homología. Es decir, que una estructura es homóloga con otra en un cierto grado, pero no comparable con ella en otro. Por ejemplo, artrópodos y vertebrados comparten la cefalización y el tener miembros pareados. Comparar directamente esas estructuras no es apropiado, aunque es posible que sean homólogas a nivel molecular, donde los mismos genes controlan el desarrollo de estas estructuras, pero la larga historia independiente de cada linaje hace imposible una comparación de las estructuras como homologías, aunque no prohíbe una comparación funcional.

Es importante recalcar que cuando alguien compara los carac-

terres de diferentes taxa, hace una inferencia fuerte sobre los caracteres: se trata del mismo carácter en los taxa; es decir, da una identidad histórica al carácter al hacerlo el mismo carácter por homología, aunque estén modificados los distintos estados (Hennig, 1968). Esto es lo que le da soporte al uso de la congruencia de caracteres para descubrir las sinapomorfías que le dan identidad histórica a los grupos supraespecíficos.

La selección y observación de caracteres, no solo morfológicos, es un trabajo que sólo se aprende mediante el trabajo continuo con los datos derivados de los especímenes en el laboratorio y con una lectura crítica de la literatura sobre el grupo. Es importante que usted pueda reconocer cuándo los caracteres usados en una descripción, una clave o un análisis filogenético cumplen o no con el requisito de la identidad histórica.

Algunos libros de texto sistemática, presentan distintos criterios de reconocimiento de caracteres (por ejemplo Hennig (1968); Wiley & Lieberman (2011)), aunque no es posible plantear reglas estrictas o situaciones de casos perfectos, eso no es un motivo para suponer que los caracteres morfológicos carecen de base conceptual y por ello deben ser excluidos del análisis (contra

Scotland et al. (2003)).

Existen algunos criterios básicos aplicables en general a los caracteres: para cada carácter se reconocen diversos estados alternativos de ese carácter, los cuales son llamados estados de carácter o simplemente estados. Es posible que un carácter no sea aplicable a todos los organismos de estudio, pero dentro de los organismos donde es aplicable, los estados de un carácter deben dividir el universo en al menos dos grupos mutuamente excluyentes, cada uno de ellos reconocible por un estado de carácter.

Dado que se tratan los caracteres como entidades históricas, se debe ser cuidadoso con NO crear estados de carácter que sirvan como "cajas de basura". Por ejemplo, en el carácter:

Color:

- 0. rojo
- 1. otro color

el estado "otro color" es ambiguo y puede contener individuos de colores tales como verde o amarillo, cuya única (falsa) similitud es que no poseen el color rojo, pero no existe una identidad

histórica defendible, es decir el NO color rojo se ha generado desde múltiples ancestros. Todos los estados del carácter deben representar una unidad histórica, independientemente de si el estado es el apomórfico o el plesiomórfico. Siempre que encuentre un carácter donde uno de los estados está definido como negación (incluidas las ausencias), debe revisar cuidadosamente que el estado "negativo" indique claramente una unidad.

Cuando los caracteres solo tienen dos estados, no es necesaria ninguna suposición sobre la "dirección del cambio"; sin embargo, cuando hay más de dos estados, la dirección es una pregunta importante. Algunos caracteres simplemente no dan posibilidad de escoger una dirección particular (por ejemplo, las bases nucleotídicas), por lo cual son caracteres no ordenados. Bajo una posición, uno puede asignar un posible orden a los caracteres que se leen como diversos grados de homología, sin asumir nada sobre el proceso evolutivo, por ejemplo, el carácter:

Pilosidad en la pata:

0. solo en el extremo anterior
1. hasta la mitad



2. toda la superficie
3. toda la pata y con presencia de pelos más largos en el extremo posterior

podría ordenarse desde 0 hasta 3 (hacerlo aditivo) para reflejar que existe un mayor grado de homología entre, por ejemplo, el estado 2 y 3, que 1 y 3. La otra posición no privilegiará ningún cambio con respecto a otro y la dirección se asignaría por la congruencia con otros caracteres (el carácter es no aditivo).

En otros casos, la identidad histórica es difícil de apreciar; por ejemplo:

transparentable con KOH:

0. si
1. no

el carácter como un todo parece ser más un artefacto de laboratorio que una propiedad heredable del organismo. En esa misma línea están caracteres como

número de componentes en PCA:

0. uno
1. dos.

Al tratarse los estados de carácter como entradas alternativas, un mismo organismo no puede poseer dos estados o más estados del mismo carácter: para nuestra visión de homología de las alas y los miembros anteriores, no pueden existir ángeles o centauros, con dos versiones diferentes del miembro anterior en el mismo organismo. Este es el criterio de conjunción de Patterson; sin embargo, hay que tener en cuenta que existen homologías seriadas que contradicen esta idea, no solo en organismos segmentados como anélidos y artrópodos, sino en otros donde la segmentación no es clara (e.g., hojas de las plantas, pelos de los mamíferos, etc.). El criterio de conjunción de Patterson es un llamado al reconocimiento de las estructuras que se usan como caracteres en el análisis filogenético.

## Técnicas

Use el dibujo al final de la práctica que corresponde a la figura 1 en Sokal (1983). Los "organismos" son caminálculos, seres hipotéticos creados por J. Camin para estudios sobre clasificaciones

en fenética<sup>2</sup>.

A partir de los dibujos:

1. Elabore una pequeña guía morfológica de los organismos, de modo que pueda diferenciar las diferentes estructuras en los diferentes organismos (en buen romance significa poner nombres a sus organismos y ligar tales nombres a estructuras fácilmente reconocibles; el ejercicio más cercano en un análisis para separar por morfos sus organismos).
2. Elabore un listado de los diferentes caracteres de los organismos, reconozca al menos 5 caracteres únicos y 5 compartidos. Anote en una hoja separada los organismos y sus estados.
3. Intercambie su listado (acompañado de la guía morfológica y los nombres de los organismos) con un compañero:
  - a) Trate de reconocer los caracteres que su compañero describió, y anote cuales organismos poseen esos caracteres.
  - b) En caso de ser necesario, re-escriba los caracteres de

---

<sup>2</sup>Copyright, Society of Systematic Biologists, se reproduce con permiso.

su compañero e indique la razón de sus cambios.

4. Compare la lista de caracteres por organismos que usted realizó con la que su compañero re-escribió.

Pregunta(s) general(es).

1. Haga un ejercicio crítico de los caracteres de sus compañeros:
  - a) ¿están bien definidos?
  - b) ¿hay implícita una transformación?
  - c) ¿usaron de la misma forma la similaridad?
2. Trate de localizar los motivos de sus aciertos y fallas cuando trabajó con los caracteres de su compañero. ¿Son esos motivos consistentes con la crítica a los caracteres?
3. Evalúe la crítica de su compañero a sus caracteres: ¿es posible reescribir los caracteres para mejorarlos? Si así lo cree, reescribalos; en caso contrario, dé argumentos para desechar el carácter o para mantener su definición actual.
4. Use la literatura de la que disponga sobre un grupo en particular, trate de evaluar los caracteres presentes y hágase preguntas similares a las que realizó cuando examinó los caracteres de sus compañeros.
5. Si es posible, trate de usar y contrastar un análisis filogenético

con una descripción tradicional del mismo grupo.

6. Pleijel (1995) argumentó que los únicos caracteres válidos son las presencias, por lo que todos los caracteres son una enumeración de presencia/ausencia (no hay estados de carácter), trate de identificar las posibles ventajas y falencias de esa aproximación.
7. Lea el artículo de Goloboff et al. (2006) e identifique la forma como manejan los caracteres cuantitativos. ¿usan un esquema similar al usado para los caracteres discretos?

## Literatura recomendada

de Pinna (1991) [Una discusión sobre la formulación de los caracteres].

Goloboff et al. (2006) [Una visión a los caracteres cuantitativos].

Kluge (2003) [Una crítica al uso de la similitud como criterio de selección de caracteres].

Neff (1986) [Una normalización del análisis de caracteres].

Patterson (1981) [La verdadera prueba de homología].

Platnick (1979) [Un artículo clásico sobre la jerarquía -cladismo de patrón-, tanto de caracteres como de taxa].

Pleijel (1995) [Una visión al manejo de caracteres, diferente a la ofrecida aquí].

Rieppel & Kearney (2002) [Una visión crítica y extensa de la selección de caracteres, con ejemplos empíricos].

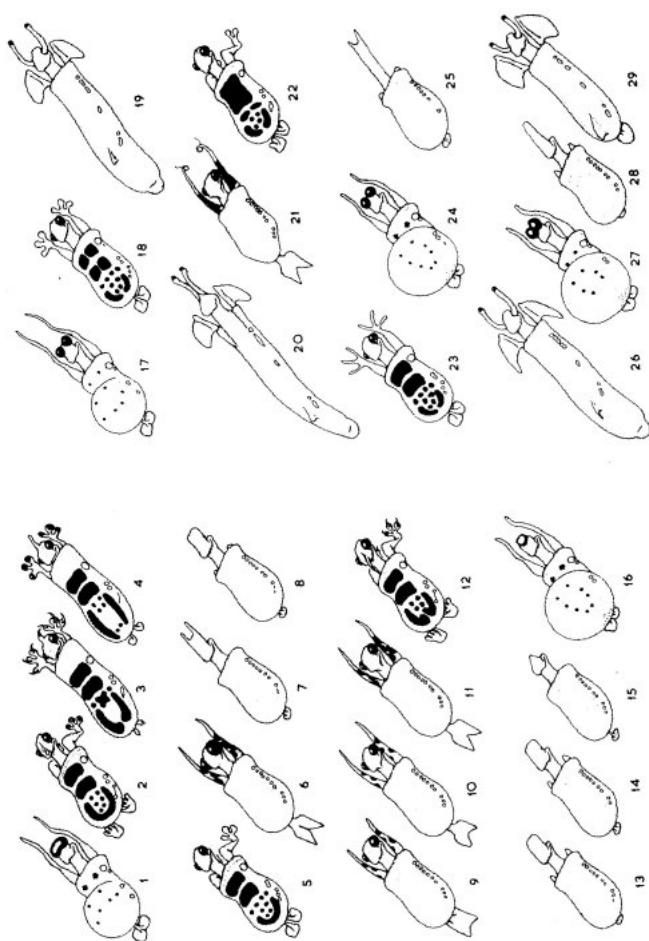


Figura 1.1: Los Caminálculos, tomado de Sokal (1983)  
 Copyright Society of Systematic Biologists. Se reproduce con  
 permiso.

# Práctica 2

## Matrices de datos

### Introducción

Una de las diferencias más importantes entre los trabajos taxonómicos con un enfoque clásico y los análisis filogenéticos es que estos últimos incluyen explícitamente una matriz de datos donde se pueden evidenciar los caracteres examinados, y cómo fueron interpretados.

Básicamente, la matriz de datos es una lista tabulada de las observaciones de los caracteres en los distintos taxa. Para facilitar



su lectura y su uso en programas de análisis filogenético, los caracteres y sus estados son codificados como números o como letras.

Una vez construida, la matriz es el punto de partida para la búsqueda de árboles y la manera más sencilla con la que otros investigadores pueden recuperar la información recopilada durante el análisis de caracteres.<sup>1</sup>

Tal y como se hizo en la práctica sobre selección de caracteres, lo que se busca es que el investigador sea consistente en la codificación de los caracteres, lo cual es importante en el manejo de caracteres inaplicables y no observados. Algunos autores los codifican de diferentes formas en la matriz (usualmente, con - y ?) para facilitar la recuperación de información (puede encontrar una discusión más completa en Strong & Lipscomb (1999)), otros autores, y en general los programas, no reconocen diferencia entre unos y otros, mientras que programas como TNT, POY o PAUP\* pueden reconocer los eventos de inserción-pérdida como un quinto carácter o como gaps (Giribet

---

<sup>1</sup>si desea puede ver nuestra visión ampliada en Miranda-Esquivel et al. (2004).

& Wheeler, 1999).

Cuando se usan caracteres con múltiples estados, es necesario clarificar cuáles fueron usados como aditivos y cuáles como no aditivos, y en el caso de haber sido recodificados, cómo se hizo esa codificación. Si se han codificado como aditivos, se debe indicar el ¿por qué? de tal codificación e incluir los argumentos que muestren que los estados de carácter se hallan anidados entre sí, la aditividad puede generar la estructura en la topología final, sin que haya realmente transformaciones que soporten los nodos.

## Técnicas

Existen diferentes programas para manipular matrices de datos y árboles. Algunos de ellos permiten la interacción matriz-árbol

(WinClada<sup>2</sup>, Mesquite<sup>3</sup>, MacClade<sup>4</sup>, R<sup>5</sup> o Seaview<sup>6</sup>).

En su mayoría estos programas sirven de plataforma para manejar otros programas que realizan el análisis filogenético como tal. Muchos de los programas que manejan matrices están diseñados básicamente para algún tipo particular de datos (por ejemplo, ADN) y aquellos que tienen un marco amplio se quedan cortos para manejar cierta clase de información (por ejemplo, no interpretan la codificación IUPAC para polimorfismos de ADN).

Los programas pueden leer uno o varios formatos de archivos, pero solo algunos programas como Mesquite leen y escriben todos los formatos de datos; es importante revisar la compatibilidad de los programas para el manejo de archivos. En la mayoría de ellos es posible exportar entre los diferentes tipos

---

<sup>2</sup>[www.cladistics.com](http://www.cladistics.com)

<sup>3</sup>[www.mesquiteproject.org](http://www.mesquiteproject.org)

<sup>4</sup>el programa no corre en versiones de MacOS X superiores a 10.6 (desde Lion en adelante)

<sup>5</sup>R es uno de los programas para análisis estadístico más versátiles, además de ser gratuito permite la implementación de múltiples procesos de cálculo desde estadística básica hasta múltiples análisis en evolución; sirve de plataforma para ape Paradis et al. (2004), un módulo que permite distintas acciones con matrices y árboles, lo puede obtener desde <http://cran.r-project.org>.

<sup>6</sup><http://doua.prabi.fr/software/seaview>

de datos, o por lo menos entre los más usados. En general, la mayor parte de los programas trabaja bien con el formato NEXUS (Maddison et al., 1997), tanto para exportar como para importar, aunque el formato es en ocasiones muy diferente y algunos programas pueden no identificarlo correctamente.

Otro formato importante es el de Hennig86 o NONA, pero en muchos programas, especialmente moleculares, su uso no está implementado.

Se han hecho algunos ejercicios de extender XML, un lenguaje de marcas, para que sea posible usarlo en análisis filogenéticos<sup>7</sup>, se busca en última instancia que los análisis filogenéticos cumplan con los estándares de reproductibilidad (ver por ejemplo Cranstoun et al. (2014)).

Revise siempre la documentación del programa que desea usar y así podrá estar seguro si el programa que va a utilizar cumple con los requisitos que usted necesita y cual es el formato de las matrices.

Existen listas de programas para análisis filogenético que pue-

---

<sup>7</sup>[www.phyloxml.org](http://www.phyloxml.org)

den ser consultadas en internet, por ejemplo:

<http://evolution.genetics.washington.edu/phylip/software.html>

o en

<http://taxonomy.zoology.gla.ac.uk/software/>

Abra el archivo de datos morfológicos para vertebrados:

**”datosVertebrados.xls”**

con un programa para hojas de cálculo y manipule las matrices con WinClada, Mesquite o TNT para Windows:

1. Anote cuáles caracteres son multiestado y cuáles son binarios.
2. Identifique si hay o no caracteres aditivos.
3. A partir de los datos, construya una matriz nueva (en TNT para Windows use el menú Data - edit data).
4. Dado que los programas no cuentan con opciones de salvado automático, periódicamente salve la matriz.
5. Nomine los terminales y los caracteres y sus estados. Explore diferentes formas de llevar esa tarea a cabo.

6. Suponga que algunos autores consideran que las plumas son escamas modificadas. Aceptando esa información, recodifique la matriz y determine si el carácter es aditivo o no.
7. Seleccione el último carácter y colóquelo al principio de la matriz.
8. Introduzca un carácter nuevo en la posición 4.
9. Exporte la matriz en otros formatos, **NEXUS** si está usando **WinClada**, **Nona** si está usando **Mesquite**.
10. Verifique la compatibilidad de los datos, abriendo la matriz en el programa correspondiente.
11. Revise con un editor de texto los archivos que usted creó, trate de identificar cuáles son las partes claves del formato<sup>8</sup>.
12. Abra una de las matrices moleculares en cada programa y con un editor de texto y trate de identificar en qué se diferencia de la matriz morfológica.
13. en R:

---

<sup>8</sup>Aunque en general no se usan los editores de texto, este paso es crítico para después ser capaz de rastrear los problemas que pueden tener las matrices de datos.

- a) Instale en su ordenador la versión más reciente de R ( $\geq 3.2.0$ ).
- b) Cargue las bibliotecas `ape` y `phangorn`<sup>9</sup>:

```
> library(phangorn)
```

- c) En caso de no tener las bibliotecas disponibles primero bájelas con la instrucción<sup>10</sup> :

```
> install.packages(c("ape", "phangorn"),
dependencies=T)
```

- d) Abra la matriz de datos en formato de texto simple y asígnela a un objeto de R:

---

<sup>9</sup>La biblioteca `phangorn` está diseñada para análisis filogenéticos que tiene como objetivo estimar árboles y redes, utilizando diferentes métodos como máxima verosimilitud, parsimonia, distancia y conjugación de Handamard, requiere las bibliotecas `ape` y `rgl`. Puede ser descargado desde <http://cran.r-project.org/web/packages/phangorn/index.html>

<sup>10</sup>La función `install.packages("Nombre_paquete")` permite hacer la descarga de los paquetes directamente del repositorio desde el entorno de R. También es posible hacer la descarga directamente de la página web e instalarlo desde cualquier directorio en el ordenador dando directamente la ruta a dicho sitio: `install.packages("../usuario/R/Packages/Paquete.tar.gz")`, pero deberá descargar e instalar, independiente, todas las dependencias requeridas por el paquete. La instrucción `library()` le permitirá cargar el paquete en el entorno de R para poder empezar a trabajar con todas sus utilidades.

```
> datos <- read.table("matriz.txt")
```

- e) Revise los nombres de las variables en la matriz con  
`names(datos)`
- f) Revise los nombres de las terminales en la matriz con  
`row.names(datos)`
- g) Abra la matriz de datos en formato **phylip/newick** y asígnela a un objeto de R:

```
> datos.Phylip <- read.phyDat("DNA1.phy",  
format="phylip", type="DNA")
```

- h) Escriba en formato **Nexus** la matriz leída datos:

```
> write.nexus.data(datos.Phylip,  
file="DNA1.nex")
```

- i) Abra la matriz escrita con **Winclada** o **Mesquite** y revise la conversión.
14. En el directorio de datos existen dos matrices con distintos problemas "**problema1.txt**" y "**problema2.txt**", intente abrir las matrices, busque y corrija el error.<sup>11</sup>

---

<sup>11</sup>tradicionalmente, las terminaciones de los archivos son .ss en **Win-**



Dependiendo de la plataforma que trabaje, usted tiene disponible distintos programas. *Mesquite* y *R* son programas gratuitos y válidos para todas las plataformas pero en general las búsquedas no son eficientes, aunque le permiten trabajar con varios tipos de datos e interactuar directa o indirectamente con programas como *PAUP\**, *TNT* o *PhyML*. Sobre Windows usted cuenta con *WinClada* que funciona tanto en modo de manejo de edición de matrices (windada) como en modo de edición y manipulación de árboles. Además, le permite hacer búsquedas con *NONA*, también puede usar la versión de Windows de *TNT* que tiene interfaz gráfica. Si usa Mac una opción es *McClade*, el cual funciona como *Mesquite* pero es más veloz y eficiente, aunque es muy factible que no lo pueda usar con la versión actual de Mac OS X.

Revise la sección de Programas de cómputo para ver las instrucciones que utilizaría con un programa distinto a *Winclada/NONA*. En todos los casos familiarícese con los menús e instrucciones para abrir/cerrar y editar tanto matrices como ár-

---

*Clada*, .nex con *Mesquite* o *MacClade*, pero debe recordar que la terminación del archivo no es necesariamente el formato.

boles, tenga en cuenta que los manuales de los programas traen información adicional, por lo tanto su lectura es una muy buena opción.

Pregunta(s) general(es).

1. Si desea transformar de un formato de matriz a otro usando exclusivamente editor de texto, ¿cuáles son los pasos a seguir?  
Ensaye la lista construida con un ejemplo.
2. Haga el listado de los aspectos comunes a todos los formatos de matrices.
3. En el laboratorio anterior se insistió en la claridad de los caracteres y sus estados. A la luz de los resultados obtenidos usando árboles y la matriz:
  - a) ¿puede usted conectar la importancia del análisis de caracteres con la forma como se interpretan los cladogramas?
  - b) Revise su bibliografía, y de ser posible compare trabajos con y sin matriz explícita. ¿Puede notar alguna diferencia?
  - c) ¿Cree que es una ventaja incluir y publicar la matriz, o por el contrario es una desventaja?

## Literatura recomendada

Maddison et al. (1997) [Una introducción al formato NEXUS].

Maddison & Maddison (2014) [El programa más versátil para manipulación de matrices, por lo que la documentación DEBE ser leída].

Maddison & Maddison (2005) [Aunque el programa no esté en uso, el manual presenta gran cantidad de información básica].

Neff (1986) [Una de las primeras discusiones sobre el manejo de pesos en caracteres].

# Apéndice A

## Programas de cómputo

Hoy en día los análisis filogenéticos usan decenas, cientos a miles de terminales y de caracteres, por lo que se requiere gran cantidad de cálculos. Aunque es posible hacerlos a mano, esto se hace imposible para más de 10 taxa y una veintena de caracteres. Esta sección busca familiarizar al estudiante con los programas más comunes, aparte de los usados durante las distintas prácticas. Es casi un truismo que la selección de programas depende de la(s) plataforma(s) utilizada(s), el problema a resolver, y en menor grado de los fondos disponibles, gratis o "por unos pocos dólares" se puede tener un repertorio apropiado de programas, la condición central a tener en cuenta es la

velocidad y la eficiencia. La facilidad de manejo en la gran mayoría de casos es secundaria; aunque la curva de aprendizaje puede ser lenta y tortuosa, las habilidades ganadas hacen que los análisis en el largo plazo sean más fáciles de ejecutar; a eso se suma la posibilidad de manejar "lenguajes" por casi todos los programas. Es posible que al final del proceso todo sea asunto de un par de instrucciones, una vez que ha perfeccionado sus habilidades y que posee archivos de instrucciones preajustadas a sus gustos y necesidades.

En la medida de lo posible, familiarícese con la línea de comandos, casi todos los programas interactúan de esa manera y lo aprendido le permitirá analizar los datos en un menor tiempo ya que podrá hacer análisis a medida, más allá de lo que permita la interfaz gráfica y podrá hacer los cálculos en paralelo, algo que no es fácil o posible desde la interfaz gráfica.

## A.1. Editores y manejadores de búsqueda

### A.1.1. WinClada

Autor: Nixon, 2002.

Plataforma: Windows 9x o superior [Funciona MUY bien en wine dentro de Linux].

Disponibilidad: Shareware (30 días de prueba), tiene un costo de US 50.00, <http://www.cladistics.com>.

El programa requiere de **NONA** para realizar las búsquedas (**PI-WE** o **Hennig86** son deseables, pero no obligatorios).

Aunque es un programa obsoleto, es uno de los programas más útiles, tanto para principiantes como para iniciados. El programa tiene dos entornos para manejar matrices y árboles. El manejo e impresión de árboles permite que se tengan imágenes listas para publicación con pocos pasos y sin necesidad de retoque posterior. Tiene múltiples "conflictos" con el manejo de imágenes que pueden ser incómodos; además, el usuario debe tener presente que el programa numera por defecto (taxa, caracteres y árboles) desde cero (la manera lógica para los programadores), o desde uno, por lo que se debe tener cuidado

al escribir el texto y referenciar la gráfica.

WinDada Matrix: Le permite cambiar algunos aspectos de la matriz (crearlas, cambiar tamaño, salvarlas, fusionarlas).

Edit: Lo más importante es que permite que los datos sean o no modificables. Además, permite adicionar polimórfismos.

Terms/Chars: Estos menú le permiten modificar cosas específicas de terminales y caracteres, tales como nombre y orden, seleccionarlos, borrarlos, adicionar.

View: Opciones de presentación. Además, puede ver estadísticos de los caracteres, aditivos, activos, pasos mínimos y máximos; y le permite intercambiar modo numérico o IUPAC (para DNA).

Output: Le permite exportar en formatos diferentes, o información variada sobre los caracteres.

Key: Si tiene los caracteres nominados, funciona como clave interactiva.

Analyze: Tiene diferentes entradas para búsquedas y cálculo de soporte. En heuristics, es posible configurar la cantidad máxima de árboles a retener, la cantidad de árboles por cada réplica (árbol de Wagner+TBR) (i.e. estrategias NONA/PAUP) y otros detalles de

la búsqueda, como especificar una semilla para los números aleatorios (0 es el tiempo interno de la computadora). En ratchet: se puede modificar la cantidad de iteraciones de ratchet, el número de árboles a retener por iteración y el número de búsquedas de ratchet, bien sean secuenciales o simultáneas. Mientras que con bootstrap/jackknife/CR with NONA permite manipular las diferentes opciones para hacer soporte con remuestreo. Recuerde borrar todos los árboles antes de ejecutar la búsqueda de remuestreo.

CPanel Para llenar la matriz. Mode: Cambia el modo de acceso de los caracteres.

Winclados Para manejar el árbol. En general, la mayor parte de estas acciones son accesibles desde la barra de herramientas con el ratón.

Las teclas F2-F12 tienen (con y sin SHIFT) diversas funcionalidades de visualización, como engrosar/adelgazar el ancho de las ramas, expandir/contraer el costo, moverse entre árboles.

Trees: opciones para visualizar y manipular el árbol sin modificar su topología (forma, consensos, ir a un árbol determinado).

Nodes: Para ver frecuencias de nodos, desafortunadamente su funcio-



nalidad es muy inestable, y a veces produce resultados inesperados.

HashMarks: Para ver las transformaciones en los nodos.

Edit/Mouse modes: Modifica las acciones posibles del ratón para modificar el árbol, como mover nodos, colapsarlos, cambiar la raíz.

Diagnoser: Para mapear caracteres.

### A.1.2. MacClade & Mesquite

Autores: Maddison & Maddison (2005)

Plataforma (McClade): Macintosh (MacOS X, PPC y Classic 68k). Esta última funciona bien con emuladores como BasiliskII).

Plataforma (Mesquite): Cualquiera, requiere Java virtual machine.

Disponibilidad (McClade): Gratuito pero sin actualizaciones, se puede acceder desde el sitio de los autores: [macclade.org](http://macclade.org).

Disponibilidad (Mesquite): Gratuito, se puede acceder desde el sitio [mesquitemproject.org](http://mesquitemproject.org).

MacClade es un programa cuyas cualidades gráficas son impresionantes, su interfaz es agradable, sencilla y no transmite miedo a

los novatos. Dadas sus magnificas propiedades gráficas, es muy recomendable para la edición de árboles y matrices, pero principalmente para la optimización de caracteres, la cual permite distintos tipos de cambio de los mismos (v.g., Dollo, ACTRAN, DELTRAN) y adicionalmente tiene una característica única de mapeo, el "equivocal cycling" [Maddison & Maddison, 1992].

En **MacClade** la edición de los datos y el manejo de los árboles se llevan a cabo en dos interfaces separadas para árboles y matrices, las cuales son accesibles en la ventana Windows. Dado que Macclade está diseñado para ser igualmente funcional con datos morfológicos y moleculares, existen múltiples herramientas de edición.

**Data Editor** En esta interfaz se construye y edita la matriz. Sus ventanas son:

- **Edit:** se pueden duplicar caracteres o taxa, editar bloques de comandos que corran directamente en PAUP\*.
- **Utilities:** se pueden buscar secuencias particulares dentro de la secuencia total, reemplazar caracteres por otros, reemplazar datos ausentes por gaps y viceversa, importar alineamientos desde el genbank y finalmente se puede lograr que la matriz

hable por sí sola ¡sí, que hable! Un lector automático lee en forma descendente los taxa de su matriz facilitándole un poco el trabajo (puede usar inglés británico, si esto lo hace más feliz).

- Characters: permite adicionar caracteres, incluir los estados, ver una lista de los caracteres que se han incluido, determinar el formato de los caracteres que están en la matriz (si son proteínas, nucleótidos, o estándar, que es el definido para simbología de números en las casillas); también permite determinar el tipo de cambio que se permitirá para los caracteres (cuando se está haciendo una optimización), el peso de los caracteres.
- Taxa: permite hacer cosas semejantes a characters, pero para el manejo de los taxa, incluir taxa nuevos, crear listas de taxa o reordenarlos.
- Display: finalmente esta ventana maneja toda la configuración gráfica de la matriz, el tipo de letra, su tamaño, el color de los caracteres, el ancho de columna, etc.
- Windows le permite moverse entre la interfaz de los datos y el árbol y llamar a una caja de herramientas que por omisión siempre está presente en la parte inferior izquierda de la ventana. Esta caja ofrece herramientas como llenar estados,

expandir columnas, cortar, entre otras. En esta ventana notes about trees y note file permiten incluir comentarios acerca de los árboles y sobre la matriz.

**Tree Window** Esta interfaz maneja y modifica los árboles y posee las ventanas básicas del data editor, no obstante, incluye otras nuevas, específicas para el mapeo de caracteres y manejo de los árboles. Solo se describen aquellas nuevas ventanas:

- **Trees:** Esta ventana permite cambiar, abrir archivos externos para incluir árboles, crear una lista de los árboles que esta incluidos a la matriz, guardar los árboles, exportarlos en formato de hennig86 o Phylip (ver anexo de formatos para más información) y manejar las politomías como blandas o duras (Coddington & Scharff, 1994).
- **Σ:** esta ventana incluye el manejo de todos los estadísticos del árbol, como el costo, el índice de consistencia, índice de retención, índice de consistencia escalonado, el número de cambios en el árbol y finalmente, puede generarse el archivo de comandos para correr el índice decay, ó soporte de Bremer en PAUP\* (vea la práctica ??).

- Trace: aquí se maneja todo lo relacionado con el mapeo de los caracteres; puede mapear todos los caracteres a la vez, o escoger un determinado carácter para ser mapeado. También se escoge el tipo de cambio de los caracteres (resolving options) el cual puede ser ACTRAN o DELTRAN, y escoger el tipo de mapeo para los caracteres continuos, *MacClade* y *Mesquite* son los únicos programas que permiten mapear tal tipo de caracteres.
- Chart: En esta ventana se pueden generar cuadros de estadísticas de los caracteres vs pasos y árboles (characters states/etc.) o de los cambios de caracteres, la ocurrencia de los estados a través de todos los caracteres. Finalmente puede comparar dos árboles resolviendo sus politomías o tal y como son.
- Display: finalmente en esta ventana se manejan los aspectos gráficos de la representación de los árboles tales como el tipo de letra, el tamaño y el estilo de la letra, el estilo y forma del árbol, numeración de las ramas, el tamaño del árbol, la configuración de los colores del mapeo y la simbología del los taxa (como números o nombres).

Aunque con menos capacidades que *McClade*, *Mesquite* es una herramienta poderosa, teniendo en cuenta que es gratuito. Su prin-

principal falla es que no permite ver las transformaciones en los nodos de todos los caracteres simultáneamente. Su interfaz es muy similar a la de *McClade* y la distribución de ventanas y comandos es más o menos equivalente.

## A.2. Búsqueda de árboles

### A.2.1. TNT

Autor: Goloboff et al. (2008), plataforma: Windows, MacOS, Unix.

Disponibilidad: Gratis, es posible descargarlo desde <http://www.lillo.org.ar/phylogeny/tnt/>.

Para parsimonia, es el programa más rápido que se ha desarrollado; incluye varios tipos de búsquedas especializadas, como la deriva y la fusión de árboles. Al igual que *NONA*, *proc* le permite abrir la matriz de datos o archivos de instrucciones. Para las búsquedas puede configurar los diferentes métodos usando *ratchet*, *drift* y *mult*; al usar *?* puede obtener los parámetros, y con *=* puede modificarlos. *Mult* puede usarse como la "central de búsqueda", definiendo un número de réplicas para una búsqueda de Wagner y las posteriores

mejoras con ratchet y drift (según como estén configurados). Para correr simplemente escriba `mult: replic X`; para ejecutar el número de réplicas que desee (X), con `tsave* nombre` se abre el archivo "nombre" para guardar árboles. Guárdelos con `save` y al final no olvide cerrar el archivo: `tsave/;`.

El programa cuenta con ayuda en línea que puede ser consultada usando `help` o escribiendo `help comando` para un comando específico. Este es único programa que tiene directamente implementado el remuestreo simétrico, además tiene implementado el soporte de Bremer, el soporte relativo de Bremer, soporte de Bremer dentro de los límites del Bremer absoluto y puede calcular la cantidad de grupos soportados-contradichos (FC), con `resample` usted puede configurar la permutación. Con `subop` (también en **NONA**) usted puede indicar el costo de los subóptimos, en diferencia de pasos con respecto al árbol óptimo. `Bsupport` hace el cálculo del índice de Bremer, con un `*` calcula el valor relativo, o puede usar `Bsupport ]` para calcular el soporte relativo dentro de los límites del absoluto. Los resultados se pueden almacenar usando algunas herramientas del lenguaje de macros<sup>1</sup>.

---

<sup>1</sup>[www.lillo.org.ar/phylogeny/tnt/scripts/General\\_Documentation.pdf](http://www.lillo.org.ar/phylogeny/tnt/scripts/General_Documentation.pdf)

Versión de menú En Windows es bastante similar a `winClada`, pero esta mucho mejor organizada. Algunas funciones son:

- **Settings:** Además de las diferentes opciones de macros, manejo de memoria, también contiene los parámetros usados en el colapsado de ramas, consensos, y pesado implícito.
- **Analyze:** Contiene los diferentes tipos de búsquedas y sus parámetros, el manejo de árboles subóptimos, y los remuestreos.
- **Optimize:** Permite ver y dibujar sinapomorfias, mapear caracteres, y revisar estadísticos de caracteres y árboles (por ejemplo `long` o `peso`).
- **Trees:** Allí se encuentran las entradas para el dibujo de árboles, el cálculo soporte de Bremer, realizar consensos y superárboles, así como árboles al azar, y el manejador de etiquetas de los nodos (`tags`).
- **Data:** Aparte de la configuración de caracteres y terminales, posee un editor básico de datos, que aunque muy simple, es extremadamente fácil de manejar, y más directo que los editores basados en mostrar la matriz (Como `winClada`, o `Mesquite`).



## A.2.2. NONA

Autor: Goloboff, 1998.

Plataforma: Linux, Mac o Windows (9x o superior).

Disponibilidad: Gratuito en conjunto con otros programas en [www.lillo.org.ar/phylogeny/Nona-PeeWee/](http://www.lillo.org.ar/phylogeny/Nona-PeeWee/).

Este programa es una buena opción que existe para búsquedas bajo parsimonia, dado que es gratuito, tiene lenguaje de macros y es veloz, pero dados los costos, velocidad y operatibilidad de TNT, este último es la mejor opción.

Básicamente los comandos de NONA son los mismos de TNT. Una diferencia importante (aparte de la velocidad y algoritmos implementados) es que en NONA los comandos son ejecutados (en vez de poder configurar sin ejecutar). A diferencia de PAUP\*, no puede desactivar terminales.

Al igual que en TNT, el comando simple de búsqueda es mult. Para permutar ramas se utiliza max. En ambos casos debe colocarse un asterisco mult\*; max\* para que utilice TBR como permutación, en caso contrario usará SPR.

Para salvar el árbol, el comando es sv. La primera vez que se lla-

ma, solo abre el archivo (Debe darse como parámetro o el programa solicita un nombre). Con `sv*` salva todos los árboles en memoria. Usando `ksv` se guardan los árboles colapsados, de lo contrario siempre se guardan dicotómicos. Para salvar el consenso es necesario usar `inters`.

El único método de soporte explícitamente implementado es el soporte de Bremer `bsupport`: con asterisco `bs*` muestra los soportes relativos, ó de lo contrario muestra el valor absoluto. Pero el programa viene acompañado de varios programas y macros que permiten calcular jackknife y bootstrap, así como hacer medidas basadas en frecuencias relativas (FC).

### A.2.3. POY

Autor: Varón et al., 2007.

Plataforma: Linux, MacOS X y Windows.

Disponibilidad: gratuito, [research.amnh.org/scicomp/projects/POY.php](http://research.amnh.org/scicomp/projects/POY.php)

Es el único programa disponible que realiza un análisis simultáneo sin utilizar alineamiento previo de las secuencias moleculares. Tam-

bién puede usarse para realizar alineamientos, o para hacer búsquedas convencionales de datos previamente alineados o morfológicos. La versión actual posee únicamente una implementación de parsimonia, pero se planean versiones que también realicen búsquedas bajo máxima verosimilitud.

El programa posee cuatro ventanas: una de salida, donde se imprimen salidas pedidas por el usuario (por ejemplo, un árbol, o la ayuda). Una ventana de comandos, donde se escriben las órdenes al programa, una donde muestra que clase de tarea se esta realizando y en la última donde se muestra el progreso de las búsquedas.

Las matrices y arboles son abiertos usando `read("archivo")`, donde `archivo` es el archivo a leer. Puede leer múltiples conjuntos de datos, bien sea abriendo uno por uno, o varios en la misma instrucción como `read("arch1", "arch2")`. Los datos se van agregando. Para limpiar la memoria se utiliza `wipe()`. Se buscan árboles con `build(x)`, que realiza `x` arboles de Wagner. Para mejorar dichos árboles hay que usar `swap()`. Note que las instrucciones están separadas. Es posible hacer búsquedas más sofisticadas usando nuevas tecnologías, como `ratchet` implementado en `perturb(iterations: x, ratchet())`, donde se harían `x` iteraciones de `ratchet`, con `swap(drift)` o `swap(annealing)` se hace deriva de árboles y cristalización simulada, y `fuse()` hace fu-

sión de árboles.

El comando para manejar costos es `transform(tcm())`. Se puede modificar los costos de transformacion para datos "estáticos" como morfología con

```
> transform(static, weight:2)
```

que daría un peso de 2 a cada transformación de datos morfológicos. Con la instrucción:

```
> transform(tcm(1,2))
```

se da un peso de 1 a las sustituciones y de 2 a los gaps y caracteres estáticos, ese es el valor que viene por omisión. Con la instrucción:

```
> transform(fixedstates)
```

se implementa el método de estados fijos de Wheeler.

Es posible calcular soportes con `calculate_support()`, que incluye bootstrap, jackknife y soporte de Bremer (por omisión). Las salidas se realizan con `report("archivo")`, pueden ser cladogramas en formato de paréntesis, dibujados en ascii, o en postscript. También es útil para conocer estadísticas de los árboles, las optimaciones de los

nodos y la integridad de los datos.

El programa funciona desde computadoras de escritorio hasta clusters. Puede ejecutarse directamente en la línea de comando o en procesos de lotes.

#### A.2.4. PAUP\*

Autor: Swofford, 2002.

Plataforma: MWindows, MacOSX, Unix.

Disponibilidad: el programa para la interfaz gráfica tiene un costo entre US 80.00 a US150.00 dependiendo del sistema operativo y es distribuido por Sinauer, mientras que para línea de comandos es gratuito, [www.sinauer.com/detail.php?id=8060](http://www.sinauer.com/detail.php?id=8060).

PAUP\* es uno de los programas más usados en cuanto a búsquedas mediante ML, o si el sistema operativo es Mac. Dados los costos es muy posible que la interfaz dominante llegue a ser por línea de comandos; la interfaz para Windows y MacOSX en modo gráfico tiene el mismo acercamiento; como NONA y TNT tiene la opción de manejar un gran número de parámetros que permiten hacer una búsqueda sobre pedido, adicionalmente realiza todos los

consensos directamente sin necesitar un segundo programa. Su gran desventaja es su velocidad y la falta de un lenguaje de macros.

`set`: esta función define la configuración básica de algunos parámetros. Son importantes `increase=no`, para que no le pregunte si desea aumentar el número de árboles, y `maxtrees=N` para dar como opción que guarde un número de árboles (N).

`exec`: para abrir la matriz de datos o archivos de instrucciones.

`hsearch`: es la función de búsqueda de cladogramas. Las opciones de este comando son `addseq=random`, lo que asegura que la entrada de datos para el árbol de Wagner es al azar.

`swap=tbr/spr`: es el tipo de permutación.

`nreps=X`: indica el número de réplicas que se hará en la búsqueda.

`nchuck=X ckuckscore=1`: le permite usar la estrategia NONA, indicando cuántos árboles desea por réplica. Si usted desea hacer solo permutación de ramas del árbol previamente construido, la serie de instrucciones sería `search start=current chuckscore=no`.

Para guardar los árboles utilice el comando `savetre`.

Con el comando `contree` se generan los árboles de consenso. Estos deben ser guardados cuando se ejecuta el comando (parámetro `treefile=nombre`, donde nombre es el archivo.). Los métodos de medición de soporte implementados son el bootstrap con el comando

bootspttrap y el jackknife con jackknife.

En caso de dudas, el programa posee una ayuda en línea, consúltela usando `help` o escribiendo `help comando` para un comando específico; para ver las opciones del comando use `comando ?`.

### A.2.5. MrBayes

Autor: Ronquist et al., 2005.

Plataforma: Windows, MacOS, Unix.

Disponibilidad: gratuito, <http://mrbayes.net>

Es el programa más comúnmente usado para análisis bayesiano. Es gratuito y su interfaz es muy similar a la de PAUP\*. Los principales comandos del programa son descritos en el capítulo de análisis bayesiano. El tutorial incluido es muy completo y viene dentro del programa. Una de las características más interesantes es la implementación de los "modelos" usados para caracteres morfológicos (Lewis, 2001), así como el modelo 'parsimonia' tal y como fue descrito por Tuffley & Steel (1997) que se activa usando

```
> lset parsmodel=yes;
```

# Apéndice B

## Formatos de archivos

### B.1. Formatos de matrices

#### B.1.1. NONA

Es válido para `NONA`, `TNT`, `Hennig86` (formato de morfología de `POY`) y `WinClada`; comienza con `xread`, luego el número de caracteres y el número de taxa, los polimórficos entre paréntesis angulares y los desconocidos con `-` o `?`. Al final, un punto y coma y si se desea la aditividad de los caracteres (comenzando desde 0). Termina con `p/` o `p-`, que los programas interpretan como fin del archivo.



```

xread 'Matriz ejemplo' 5 5
out 00000
alpha 10-20
beta 1102[01]
gamma 1?111
lamda 11111
;
cc -0.2 +3 -4;
p/;

```

Para ADN (en **NONA**) se usa `dread`, con la cláusula `gap` seguida de `?` si se quieren asumir los gaps como desconocidos, o con `;` si quiere que sean un quinto estado. Se usa codificación tipo IUPAC.

```

dread gap ; match . 'DNA' 5 5
out ACGTC
alpha AT-CG
beta RTAAC
gamma CGAY-
lamda TCNCC

```

```
;
```

```
cc -.;
```

```
p/;
```

### B.1.2. PAUP\*

Se inicia con la cláusula `#nexus`, y luego con el bloque `data`. se puede definir si los caracteres son morfológicos, ADN o proteínas. Los polimorfismos se colocan entre paréntesis redondos. ADN en formato IUPAC.

```
#nexus
```

```
begin data;
```

```
dimensions ntax=5 nchar=5;
```

```
format missing=? gap=- symbols="0 1 2";
```

```
matrix
```

```
out 00000
```

```
alpha 10-20
```

```
beta 1102(01)
```

```
gamma 1?111
```

```
lamda 11111
```

```
;
```

```
end;

begin assumptions;

typeset tipoUno=unord:1-3 5, ord:4;

end;

begin paup;

[Aqui puede colocar instrucciones específicas de paup, por
ejemplo búsquedas]

hsearch add=random;

end;


#nexus

begin data;

dimensions ntax=5 nchar=5;

format missing=? gap=- datatype=dna;

matrix

out ACGTC

alpha AT-CG

beta RTAAC

gamma CGAY-

lamda TCNCC

;
```

```
end;
```

### B.1.3. POY

POY y Clustal, utilizan el formato FASTA (la primera línea inicia con > seguido por el nombre y comentarios de la secuencia; en la siguiente línea comienza la secuencia como tal).

```
> taxonA Comentarios
aaacgt
aac
> taxonB
aaacgt
```

## B.2. Formatos de árboles

### B.2.1. NONA

NONA, Hennig86, WinClada y TNT usan este formato; con \* indican que hay más árboles y con ; que es el último árbol. Nótese el espacio para separar los terminales. El primer taxon es 0. Al igual

que en las matrices, p- o p/ indican el final de lectura del archivo. Winclada puede incluir al inicio la lista de nombres, pero no es compatible con otros programas.

```
tread 'tres arboles'
(0 (1 (2 (3 4 ))))*)
(0 (1 (2 3 4 )))*
(0 ((1 2 )(3 4 )));
p/;
```

```
tread 'solo un arbol'
(0 (1 (2 (3 4 ))));
p/;
```

### B.2.2. PAUP\*

En PAUP\* el árbol está embebido en el archivo de la matriz o en un formato aparte. Los grupos son separados por comas y el primer taxon es 1.

```
#nexus
```

```

begin trees;
translate
1 lamda,
2 alpha,
3 beta,
4 gamma,
5 out
;
tree *primero=(5,(1,(2,(3,4))));
tree politomico=(5,(1,(2,3,4)));
tree tercero=(5,((1,2),(3,4)));
end;

```

El orden no altera los árboles en los programas. Así:

$(\emptyset, (1, (2, (3, 4))))$

es igual a

$((1, ((3, 4), 2)), \emptyset)$

# Apéndice C

## Algunos comandos básicos para POY

### C.1. Análisis de sensibilidad con costos diferenciales

(\* tomado de [urlgroups.google.com/group/POY4/](http://urlgroups.google.com/group/POY4/) \*)

(\* datos dos archivos de datos \*)

(\* 1.fas y 3.fas \*)

```

> read("1.fas", "3.fas")
> transform((names:("1.fas"),          tcm:"412.txt"),
             (names:("3.fas"), tcm:"121.txt"))
> store("misdatos")
> build(5, trees:2)
> select(best:1)
> transform((all, static_approx))
> report("todo13.mtr", phastwinclad) (* Resultados
    en el archivo Sensitivity_results.txt *)
> echo          ("412+121          Resultados",
    output:"Sensitivity_results.txt")
> report ("Sensitivity_results.txt", treestats)
    (* Deseche los árboles *)
> select(best:0)
> use("misdatos")
> select(characters,names:("1.fas"))
> build(5)
> select()
> echo("1.fas          Resultados",

```



```
    output:"Sensitivity_results.txt")
> report("Sensitivity_results.txt", treestats) (*
    Deseche los árboles *)
> transform((all, static_approx))
> report("matriz1.mtr", phastwinclad)
> use("misdatos")
> select(best:0)
> select(characters,names:"3.fas"))
> build(5)
> select(best:1)
> echo          ("3.fas          Resultados",
    output:"Sensitivity_results.txt")
> report ("Sensitivity_results.txt", treestats)
> transform((all, static_approx))
> report ("matriz3.mtr", phastwinclad)
> quit()
```



# Bibliografía

Coddington, J. & N. Scharff. 1994. Problems with zero-length branches. *Cladistics* 10:415--423. [doi:10.1111/j.1096-0031.1994.tb00187.x].

Cranston, K., L. Harmon, M. O'Leary, & C. Lisle. 2014. Best practices for data sharing in phylogenetic research. *PLOS Currents Tree of Life* [doi:10.1371/currents.tol.bf01eff4a6b60ca4825c69293dc59645].

de Pinna, M. C. C. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7:367--394. [doi:10.1111/j.1096-0031.1991.tb00045.x].

Giribet, G. & W. C. Wheeler. 1999. On gaps.

- Molecular Phylogenetics and Evolution 13:132 -- 143.  
[doi:dx.doi.org/10.1006/mpev.1999.0643].
- Goloboff, P. A., J. S. Farris, & K. C. Nixon. 2008. Tnt, a free program for phylogenetic analysis. Cladistics 24:774--786. [doi:10.1111/j.1096-0031.2008.00217.x].
- Goloboff, P. A., C. I. Mattoni, & A. S. Quinteros. 2006. Continuous characters analyzed as such. Cladistics 22:589--601. [doi:10.1111/j.1096-0031.2006.00122.x].
- Grant, T. & A. Kluge. 2004. Transformation series as an ideographic character concept. Cladistics 20:23--31.
- Hennig, W. 1968. Elementos de una sistemática filogenética. Eudeba, Buenos Aires.
- Kluge, A. G. 2003. The repugnant and the mature in phylogenetic inference: atemporal similarity and historical identity. Cladistics 19:356--368. [doi:10.1111/j.1096-0031.2003.tb00379.x].
- Maddison, D. R. & W. P. Maddison. 2005. MacClade 4: Analysis of phylogeny and character evolution. Version 4.08a. Sinauer Associates Incorporated.

- Maddison, D. R., D. L. Swofford, & W. P. Maddison. 1997. NEXUS: an extensible file format for systematic information. *Systematic Biology* 46:590--621. [doi:10.1093/sysbio/46.4.590].
- Maddison, W. P. & D. R. Maddison. 2014. Mesquite: a modular system for evolutionary analysis. (mesquiteproject.org).
- Miranda-Esquivel, D., I. Garzón, & J. Arias. 2004. ?'taxonomía sin historia? *Entomólogo* 32:4--7.
- Neff, N. 1986. A rational basis for a priori character weighing. *Systematic Zoology* 35:110--123.
- Paradis, E., J. Claude, & K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289--290. [doi:10.1093/bioinformatics/btg412].
- Patterson, C. 1981. Significance of fossils in determining evolutionary relationships. *Annual Review of Ecology and Systematics* 12:195--223.
- Platnick, N. I. 1979. Philosophy and the transformation

- of cladistics. *Systematic Biology* 28:537--546.  
[doi:10.2307/sysbio/28.4.537].
- Pleijel, F. 1995. On character coding for phylogeny reconstruction. *Cladistics* 11:309--315.  
[doi:10.1111/j.1096-0031.1995.tb00092.x].
- Richards, R. 2002. Kuhnian values and cladistic parsimony. *Perspectives on Science* 10:1--27.
- Richards, R. 2003. Character individuation in phylogenetic inference. *Philosophy of Science* 70:264--279.
- Rieppel, O. & M. Kearney. 2001. The origin of snakes: limits of a scientific debate. *Biologist* 48:110--114.
- Rieppel, O. & M. Kearney. 2002. Similarity. *Biological Journal of the Linnean Society* 75:59--82.  
[doi:10.1046/j.1095-8312.2002.00006.x].
- Scotland, R. W., R. Olmstead, & J. Bennett. 2003. Phylogeny reconstruction: the role of morphology. *Systematic Biology* 52:539--548.
- Sokal, R. 1983. A phylogenetic analysis of the caminalcules. i. the data base. *Systematic Zoology* 32(2):159--184.

Strong, E. & D. Lipscomb. 1999. Character coding and inaplicable data. *Cladistics* 15:363--371.

Tuffley, C. & M. Steel. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bulletin of mathematical biology* 59:581--607.

Wiens, J. 2001. Character analysis in morphological phylogenetics: problems and solutions. *Systematic Biology* 50:689--699.

Wiley, E. & B. Lieberman. 2011. John Wiley & Sons.