

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 XXYYXX secciones 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 dialogo en los programas de interfaz gráfica (v.g., **Analyze/heuristics**).

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

Público objeto

Se espera que el usuario de este manual tenga conocimientos básicos, o que esté tomando un curso formal de sistemática filogenética a nivel de pre o posgrado. No se esperan características especiales en cuanto a dominio de computadores, pero el manejo de un editor de texto que permita grabar archivos sin formato es *muy* recomendable, además de manejar 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 el 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 terminos 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 la 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 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)).¹

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

¹Este 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 caracteres 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 identidades 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. total a 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 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 el cual corresponde a la figura 1 en Sokal (1983). Los "organismos" son camináculos, seres hipotéticos creados por J. Camin para estudios sobre clasificaciones en fenética². 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

²Copyright, Society of Systematic Biologists, se reproduce con permiso

(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 cuáles organismos poseen esos caracteres.
 - b) En caso de ser necesario, re-escriba los caracteres de 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.

Preguntas generales.

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 re-escribir 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) [Ofrece una visión a 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].

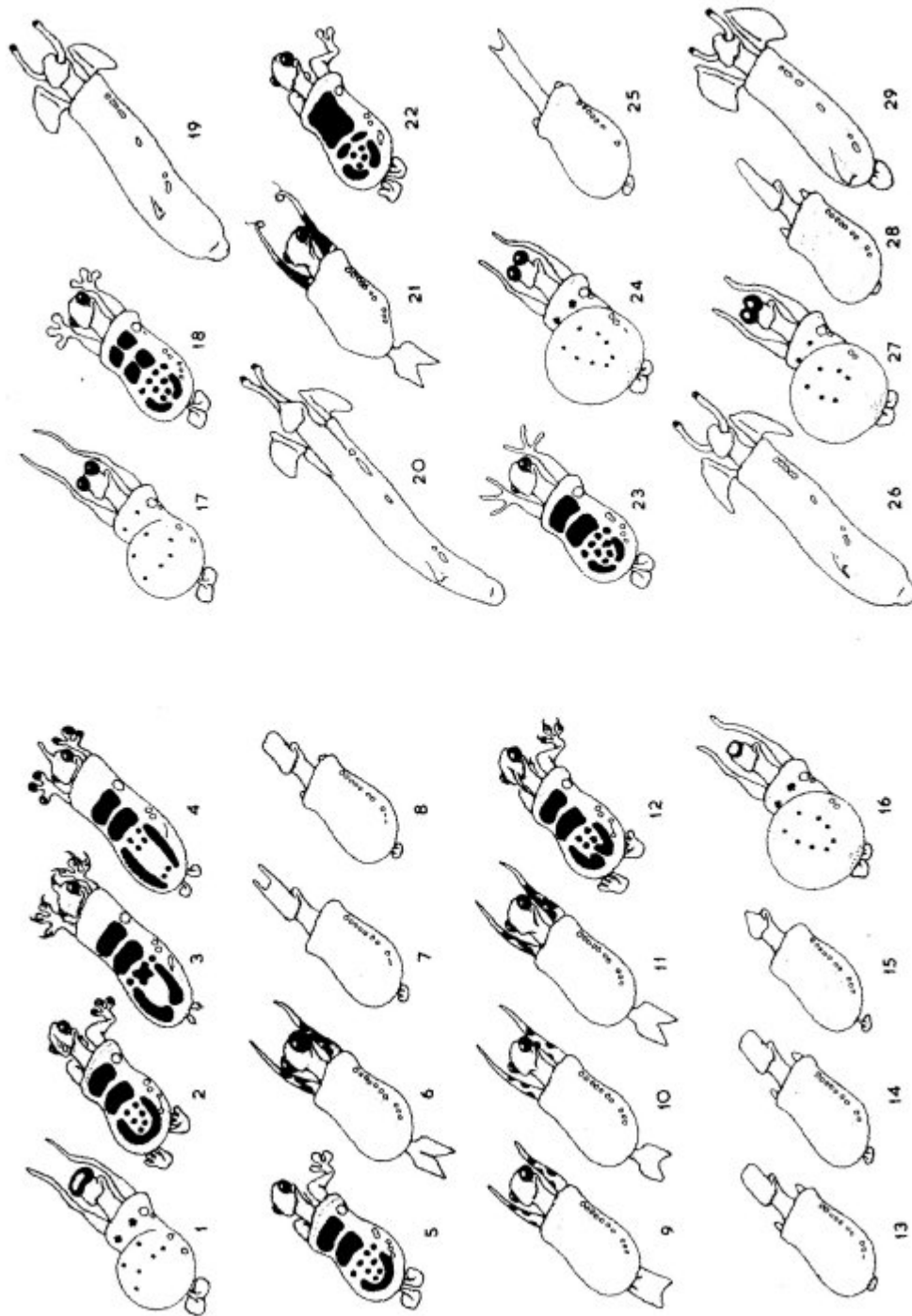


Figura 1.1: Los Camináculos, tomado de Sokal (1983)
 Copyright *Society of Systematic Biologists*, se reproduce con permiso.

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 puede 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**.¹

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 o Poy pueden reconocer como un quinto carácter los eventos de inserción-perdida o gaps (Giribet & Wheeler, 1999).

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

Cuando se usan caracteres con múltiples estados, es necesario clarificar cuáles fueron usados como aditivos y cuáles como no aditivos, o en el caso de haber sido recodificados, como 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 al 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](#)², [Mesquite](#)³, [MacClade](#)⁴, [R](#)⁵ o [Seaview](#)⁶). 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 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

²www.cladistics.com

³www.mesquite.org

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

⁵[R](#) es uno de los programas para análisis estadístico más versátiles del momento, además de ser gratuito permite la implementación de múltiples procesos de cálculo, de tal manera que es posible realizar distintos tipos de operaciones y cálculos desde estadística básica hasta análisis en evolución; sirve de plataforma para [ape](#) Paradis et al. (2004), un módulo que permite distintas acciones con matrices y árboles cran.r-project.org

⁶doua.prabi.fr/software/seaview

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 filogenéticos⁷, se busca en última instancia que los análisis filogenéticos cumplan con los estándares de reproducibilidad (ver por ejemplo Cranston 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 pueden ser consultadas en internet, por ejemplo:

evolution.genetics.washington.edu/phylip/software.html

o en

taxonomy.zoology.gla.ac.uk/software/

Técnicas

Abra el archivo de datos morfológicos para vertebrados "datos.vertebrados.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.

⁷www.phyloxml.org

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 formato⁸.
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:
 - a) Instale en su ordenador la versión más reciente de R (≥ 3.00 -11).
 - b) Cargue las bibliotecas ***ape*** y ***phangorn***⁹:

```
> library(ape); library(phangorn)
```

- c) En caso de no tenerlas disponibles primero bajelas con la instrucción:

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

⁸Aunque 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.

⁹La 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", y puede ser descargado desde <http://cran.r-project.org/web/packages/phangorn/index.html>

10

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

```
> 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** y asignela a un objeto de R:

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

- h) Escriba en formato **Nexus** la matriz leida anteriormente:

```
> write.nexus.data(datos.Phylip,file="NuevaDNA1.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.¹¹

¹⁰La 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 pagina web e instalarlo desde cualquier directorio en el ordenador dando directamente la ruta a dicho sitio:

```
> install.packages("../usuario/R/Packages/Paquete.tar.gz")
```

Tenga en cuenta que de este modo deberá descargar e instalar todas las dependencias requeridas por el paquete de manera independiente. La instrucción `library()` le permitirá cargar el paquete en el entorno de R para poder empezar a trabajar con todas sus utilidades.

¹¹tradicionalmente, las terminaciones de los archivos son `.ss` en **WinClada**, `.nex` con

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

Preguntas generales.

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 contruida 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?

Mesquite o **McClade**, pero **debe** recordar que la terminación del archivo no es necesariamente el formato.

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 este en uso, el manual presenta gran cantidad de información básica].

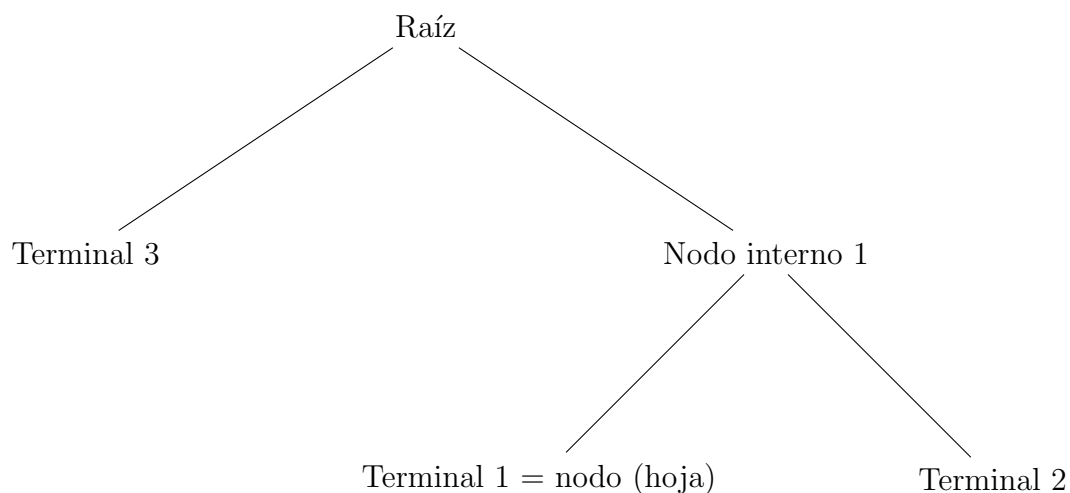
Neff (1986) [Una de las primeras discusiones sobre el manejo de pesos en caracteres].

Práctica 3

Árboles

Introducción

Una vez finalizado un análisis filogenético, el resultado es un agrupamiento que tradicionalmente se dibuja como un árbol donde los taxa usados estan como *hojas* o *terminales* del árbol. Las diferentes ramas se unen en *nodos* o *componentes*. Así, el agrupamiento ((Terminal 1, Terminal 2), Terminal 3) puede representarse con el árbol:



Es importante diferenciar entre el **contenido** del nodo, que son los terminales que forman el nodo (el nodo interno 1 tiene dos terminales: 1 y 2), y

la **información** del nodo, que es cómo están organizados estos terminales en terminos de grupos o clados anidados Nelson (1979). En general, para muchos autores un componente representa un grupo monofilético (i.e., definición *topológica* de monofilia).

En cladística, los árboles son también conocidos como **cladogramas** y el término es usado sin distinción. En cladística (pero no en otras metodologías) los cladogramas deben tener transformaciones de caracteres en nodos (nodos soportados), en caso de no tener transformaciones, el nodo no está soportado o es un nodo de longitud cero, es decir, es un artefacto de los programas usados para obtener los árboles (véase Coddington & Scharff (1994); Goloboff (1998)). El número de transformaciones es un estadístico importante que hay que tener en cuenta y se usa como criterio para seleccionar entre diferentes cladogramas, se escoge la topología que sugiere el menor número de transformaciones. En los métodos estadísticos generalmente es importante la longitud de las ramas, esta suele representarse como una probabilidad o frecuencia de transformaciones y suelen presentarse los árboles con una escala que muestra la longitud de ramas; sólo la dimensión que va de las ramas a los terminales tiene importancia en ese caso, el "ancho" del árbol se usa para acomodar los diferentes terminales usados. En los cladogramas, las dimensiones no tienen ningún significado especial.

Técnicas

Una forma de expresar los árboles en formato de texto (por ejemplo para usar con programas o para el resumen de un artículo) es la notación de paréntesis, donde los paréntesis limitan los nodos; dependiendo de los autores o los programas, los grupos hermanos son separados por espacios, comas o símbolos de suma, por ejemplo (a (b c)), es equivalente a (a+(b+c)) y a (a, (b, c)), tenga en cuenta que en los árboles filogenéticos el orden de los terminales, sin cambiar la topología, no altera el contenido del árbol, por ejemplo (a (b c)) es exactamente igual a ((c b) a) y a ((b c) a).

Existen diferentes programas para manipular y generar árboles; la mayoría de ellos permiten interacción matriz-árbol, mientras que otros solo utilizan árboles. En general, los programas de análisis filogenético permiten manipular árboles en un esquema gráfico rudimentario. Algunos programas permiten guardar información adicional a la topología en el mismo árbol como la longitud de la rama que conduce al nodo o terminal o una etiqueta.

Los programas que pueden trabajar independientemente de la matriz suelen estar diseñados para la impresión o exportación gráfica de los árboles, soportando cambios de topología, longitudes y etiquetas de las ramas, pero no transformaciones de caracteres.

Dependiendo de la plataforma que trabaje, usted tiene disponible distintos programas: **Mesquite**, **R**¹ y **Figtree**² son válidos para todas las plataformas (adicionalmente son gratuitos). **Mesquite** es muy lento si su máquina es lenta (requiere de la máquina virtual de Java). Sobre Windows usted cuenta con **WinClada** (requiere la matriz de datos), su interfaz de impresión requiere mucha práctica (no es intuitiva). Otras alternativa, si usa Linux, pueden ser **NJPlot** o **GNUplot**. Para la impresión final del árbol para publicación, una opción puede ser **R**, que aunque es poco intuitivo al iniciar, logra resultados fieles de mejor calidad que otros programas.

En general, la manipulación de arboles se activa con un menú (en **WinClada** **Edit/Mouse**, en **FigTree** directamente como barra de herramientas en la ventana; en **Mesquite** se usa el menú de herramientas de la ventana de edición de **TreeView**). Las acciones se realizan al seleccionar con el ratón la rama (**WinClada**) o arrastrando las ramas (para moverlas en **Figtree**). Más que comandos, en esta práctica es importante manipular el ratón.

1. Desde **WinClada**+**NONA**, **Mesquite** o **TNT**

- a) Revise los archivos de datos con editor de texto y con el programa seleccionado abra una matriz de datos que contenga tanto datos como árboles.
- b) Establezca la forma de obtener información sobre los árboles, en un principio los costos o longitud del árbol.
- c) Cambie de posición algunos nodos o terminales de la topología, y observe cómo cambia la longitud del cladograma y los estados asignados a los nodos o la forma como los caracteres son mapeados, pruebe también moviendo ramas completas.

2. En **Mesquite** y **TNT** (para Windows) usted puede mover ramas completas; en **TNT** (para Windows) use el menu **Trees - view** y seleccione con el botón izquierdo del rat

¹cran.r-project.org

²tree.bio.ed.ac.uk/software/figtree/

ón el clado a mover (el clado queda marcado en rojo), señale con el ratón el destino usando el botón derecho.

3. Desde WinClada o TNT (para Windows), haga una búsqueda de cladogramas para la matriz de vertebrados que contruyó en la práctica de matrices, use los parámetros por omisión (menú **Analyze / heuristics, traditional search**).
4. En Mesquite o Figtree:
 - a) Cambie el orden de los terminales sin cambiar la topología (las relaciones entre terminales) del árbol.
 - b) Coloque etiquetas en los nodos, y salve el árbol.
 - c) Revise ese archivo en un editor de texto para determinar cómo se colocaron tanto las etiquetas como las longitudes.
 - d) Abra el árbol con longitud de ramas.
5. En Mesquite o WinClada abra el archivo "datos.conarbol.dat":
 - a) Indique la cantidad de árboles presentes y la longitud de los mismos.
 - b) Examine las agrupaciones obtenidas e identifique los caracteres que soportan los nodos.
 - c) Mapee los caracteres en los árboles usando distintos tipos de optimizaciones: en WinClada ACCTRAN (=fast), DELTRAN (=slow) y no ambigua, En TNT (para Windows) (**Optimize > Character > Reconstructions** y seleccione un solo árbol y algunos caracteres) y en Mesquite: Dollo e Irreversible (C-S); revise cómo cambian las optimizaciones en los distintos nodos.
 - d) Revise de nuevo el efecto en la longitud de los árboles de las siguientes modificaciones:
 - 1) Cambie la raíz del árbol.
 - 2) Cambie de posición nodos y terminales.
 - 3) Seleccione un nodo y colapselo.
6. R:

- a) Lea el archivo con árboles "arbol.r.tre" use las instrucciones:

```
> library(ape); arbol <- read.tree("arbol.r.tre")
```

- b) Grafique el objeto arbol use las instrucciones, que por defecto dibuja el árbol con longitud de ramas o como un filograma:

```
> plot(arbol)
o
> plot.phylo(arbol)
```

- c) Para graficar el árbol sin incluir la longitud de ramas use las instrucciones:

```
> plot(arbol, use.edge.length=FALSE)
```

- d) Para asignar las longitudes de las ramas como rótulos de los nodos y graficarlos use las instrucciones:

```
> arbol$node.label <- arbol$edge.length
> plot(arbol, show.node.label=TRUE)
```

- e) Para graficar el árbol con título de gráfica ("Árbol 1"), colores de ramas y tipo cladograma use las instrucciones:

```
> plot(arbol, type="cladogram", main="Árbol 1",
edge.col=c("red", "blue", "cyan"))
```

- f) Para obtener la longitud del árbol, lea la matriz de datos "matriz.r.phy" y asignela a un objeto que se llame MatrizDatos y use la función parsimony():

```
> parsimony(arbol,MatrizDatos)
```

Preguntas.

- a) Al hacer una búsqueda por defecto ¿Usted y sus compañeros obtienen las mismas topologías y los mismos caracteres que soportan los grupos?.
- b) Al cambiar la aditividad/no aditividad de un caracter ¿Cambia esto la forma como se mapean? Haga este ejercicio para varios caracteres, tanto binarios como multiestado.
- c) Si cambia la raíz, ¿cambia la longitud? ¿Por qué cree que se presenta el resultado que obtuvo? ¿Es (y por qué) una ventaja o una desventaja?
- d) Examine sus caracteres y cambie su aditividad-no aditividad. ¿Cambia esto la forma como mapean? Haga este ejercicio para varios caracteres, binarios y multiestado.
- e) Si cambia la raíz, ¿cambia la longitud? ¿Por qué cree que se presenta el resultado que obtuvo? ¿Es (y por qué) una ventaja o una desventaja?
- f) Dibuje el siguiente cladograma:
(Lampreas (Tiburón (Esturión Teleósteo) (Celacanto (Pezpulmonado (Anfibio (Mamífero (Tortuga (Lagarto (Cocodrilo Ave))))))))))
- g) Si la raíz esta colocada entre Ave y Cocodrilo, ¿cómo es la topología resultante?
- h) Convierta la topología dibujada a notación de paréntesis.

Preguntas generales. Aparte de no tener un análisis explícito, existe una gran diferencia entre los árboles filogenéticos actuales y sus "equivalentes" usados por algunos taxónomos (por ejemplo, Haeckel, Romer, etc.) ¿Es usted capaz de descubrirla? Clave: intente dibujar alguno de esos árboles al estilo actual.

Literatura recomendada

Page & Holmes (1998) [El capítulo 2 esta dedicado a los árboles y presenta muy buenas ilustraciones, También puede consultar la página de docencia de Rod Page (taxonomy.zoology.gla.ac.uk/teaching/index.html)].

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

Práctica 4

Pesaje de caracteres

Introducción

Cuando se habla de pesaje de caracteres en cladística, se refiere a que algunos caracteres sugieren más información que otros. El tema es controversial; la idea de pesaje es central en autores como Neff (1986), mientras que para autores como Kluge (1997) es regresar a la subjetividad de la época de la taxonomía clásica, al imponer a la filogenia un prejuicio de "cómo" es la evolución Kluge (1997); otros autores (por ejemplo, Goloboff (1993, 1995)) defienden el uso de pesado, argumentando que es claro tras un análisis filogenético que diferentes caracteres poseen distintos niveles de información; lo cual se hace evidente al multiplicar cualquier cuantificador de homoplasia del carácter por el peso inicial que se le asignó, en la lógica de pesaje sucesivo Kluge & Farris (1969).

Existen dos formas de hacer pesado de caracteres y una tercera que es un criterio de búsqueda, que no es excluyente del pesaje de caracteres. El primero es el pesado *a priori*, antes de empezar el análisis.

En la actualidad la forma más común es disminuir el peso del tercer codón en los análisis moleculares. Aunque se han propuesto muchas formas de encontrar a partir de los caracteres un peso inicial, quienes practican pesado del tercer codon no están muy preocupados e insisten que lo importante es diferenciar un tipo de codon de los otros, por ejemplo: Springer et al. (2001); Strugnell et al. (2014).

En el pesado *a posteriori* el peso se asigna basándose en un análisis inicial de los datos; se estima la *confianza* del carácter, usando por ejemplo, el

índice de consistencia, o el de retención, y con base en esos pesos se reinicia el análisis (Farris, 1969). En general es el esquema de pesaje más usado. El pregunta clave aquí es: ¿cómo se hace este primer análisis? Y después, ¿cómo se termina?. Otro de los problemas de esta forma de pesado radica en la comparación de los cladogramas, puesto que diferentes juegos de pesos pueden producir diferentes respuestas que no son comparables (a nivel de los estadísticos de ajuste, como longitud).

La tercera forma de "pesado" no es ni *a priori*, ni *a posteriori*, ya que en realidad es un criterio de búsqueda. Esta forma es conocida como pesado implícito (Goloboff, 1993). Este procedimiento utiliza la confianza del carácter (con una función cóncava de homoplasia) y utiliza ese valor como criterio óptimo (en vez de la longitud pesada); el problema de este método, de hecho el de cualquier método, es cómo escoger entre las diferentes funciones. El pesado sucesivo es muy popular para "reducir" la cantidad de árboles más parsimoniosos (Carpenter, 1988), aplicación que no es correcta ya que es una metodología con su propia lógica, definir los pesos de acuerdo al comportamiento en las búsquedas previas, por lo que los resultados no tienen que coincidir con los de pesos iguales, ni en topología ni en número de soluciones (implícito en Goloboff (1995)).

Para iniciar las rondas de pesado, Farris (1969) propuso comenzar basándose en la compatibilidad de caracteres (caracteres que no se contradicen entre sí). En la actualidad la forma más común es iniciar con un árbol de pesos iguales. Kjer et al. (2001) propusieron comenzar con el mejor resultado de varios árboles producto de *bootstrap* (u otro método que produzca pseudoréplicas).

Dado que se espera que el análisis sea autoconsistente (Farris, 1969), dos réplicas consecutivas deben generar los mismos árboles (y por lo tanto los mismos pesos). La autoconsistencia se puede ver afectada por el hecho de tener gran cantidad de árboles óptimos, por lo que es importante limitar el número de iteraciones. Kjer et al. (2001) proponen no iterar y solo mantener los pesos dados por los árboles de las pseudoréplicas.

Farris (2001) buscó de solucionar simultáneamente ambos problemas. Propuso comenzar con un árbol donde se ha hecho *jackknife* con probabilidad de 0.5, restaurar luego todos los caracteres y comenzar a partir de los estadísticos del árbol producto de la permutación e iterar; el proceso se repite varias veces. Farris2001 argumenta que si los pesos son independientes del punto inicial, las diferentes réplicas producirán aproximadamente los mismos resultados (los resultados se muestran como un árbol consenso de la mayoría de las diferentes réplicas).

El método de Goloboff (1993) sigue la lógica de parsimonia tradicional, y aunque Goloboff1993 veía su método como un refinamiento de pesado sucesivo, este pesaje se puede usar igualmente en coordinación con pesado implícito. La forma común de pesado implícito es usar una función cóncava de la forma $\frac{k}{k+h}$, donde k es la constante de concavidad y h la homoplasia del carácter. Cuanto mayor sea la constante de concavidad, menor es la diferencia entre los caracteres muy homoplásicos y los poco o no homoplásicos, por lo que es equivalente a parsimonia lineal.

Técnicas

Para la matriz de datos (datos.pesado.dat).

1. En PAUP* se pueden definir juegos de caracteres usando el código X - .\ N, donde X es el número del carácter donde se empieza y N el número de posiciones en los que se vuelve a aplicar la opción. Por lo tanto la instrucción weights 3:all; todos los caracteres tendrán peso de 3 y con weights 1: 3 - .\ 3; cada tercer carácter, a partir del carácter 3 será pesado con 1. Usted puede chequear esto usando cstatus full=yes; que le mostrará el peso de todos los caracteres y deben verse en la secuencia 3 3 1 3 3 1..., también tiene implementado una opción para tomar pesos a partir de los árboles en memoria. Esa opción es reweight en la que se pueden modificar el tipo de pesado, qué valor se va a utilizar y la escala de los pesos. ¡Use la ayuda en línea para manipular estos valores!.

- a) Abra la matriz y realice una búsqueda simple.
- b) Repita la búsqueda pesando diferencialmente el tercer codón (por ejemplo asignándole un peso de 0), primero creamos la partición:

```
> charset terceraPosicion=3-.\ 3;
```

y luego se asigna el peso

```
> weight 0 :terceraPosicion;
```

Compare los resultados con los de la matriz sin pesado diferencial.

- c) Coloque todos los pesos iguales y realice una búsqueda con pesado sucesivo, itere un máximo de 10 veces. Chequee si los pesos se estabilizaron revisando la longitud de los arboles usando pcores;.
2. En TNT el pesado implícito se activa utilizando piwe=N; donde N es el valor de concavidad a usar. Con PAUP* se activa usando

```
> pset Goloboff=yes gk=(N-1);
```

La notación de N-1 es necesaria ya que PAUP* comienza con 0 y suma uno (técnicamente es como comenzar desde 1). Tanto en TNT como en PAUP* el límite es de 32000 (lo cual es prácticamente equivalente a parsimonia lineal). Además TNT y PAUP* usan todos los valores decimales, por lo cual es recomendable recurrir a alguna medida de soporte para los nodos, con lo que se evita la sobre-resolución. A diferencia de PAUP*, TNT usa una función a minimizar¹; para comparar los reportes del ajuste generados en PAUP*, puede obtener el ajuste usando fit*; En PAUP* use

```
> pcores gfit=yes;
```

TNT permite definir su propia función de concavidad usando piwe [A B C...;], donde A es el fit para 0 pasos extra, B para 1, C para 2 y así sucesivamente.

3. Tanto en TNT como en PAUP*:
 - a) Con los caracteres con pesos iguales, active los pesos implícitos con k=1, y haga una búsqueda.
 - b) Revise el soporte de los nodos usando *jackknife* (use pocas réplicas, máximo 100).
 - c) Repita los análisis con valores de k de 3, 6 y 10.

Preguntas generales.

¹el inverso de la usada en PIWE: $\frac{h}{k+h}$

- Compare sus resultados (topologías y caracteres en los nodos) con todos los métodos de pesado utilizados. ¿En qué se parecen y en qué se diferencian los resultados?
- Usted debió definir una forma de pesado del tercer codon. Defienda su esquema de pesado y compárelo con el de sus compañeros.
- Dados los resultados que encontró al medir el soporte de los nodos, ¿cree usted que hay sobrerresolución en los datos usados?
- Escriba un ensayo corto (de media página) donde presente su posición con respecto al pesado de caracteres, tenga en cuenta las siguientes preguntas:
 - ¿Esta a favor o en contra? ¿por qué?
 - ¿Cuáles son las ventajas y problemas que ofrece su posición?
 - ¿El pesaje contradice la lógica de agrupamiento por homología?

Literatura recomendada

FALTA

Farris (1969) [La idea original de pesar usando la información derivada de los cladogramas].

Goloboff (1995) [Una defensa de pesado implícito].

Goloboff et al. (2008) [Un análisis empírico que muestra que pesaje es un mejor predictor que parsimonia lineal].

Kluge (1997) [Un ataque a todas las formas de pesaje].

Práctica 5

Consensos

Introducción

Por las prácticas anteriores y por la literatura consultada, usted ha notado que en muchas ocasiones se produce más de un cladograma como respuesta. Para resumir la información contenida en los diferentes cladogramas se puede usar un consenso, que de información sobre los agrupamientos en los diferentes árboles iniciales. Swofford (1991) y Nixon & Carpenter (1996) ofrecen una discusión extensa sobre los árboles de consenso con dos visiones diferentes. Es importante recalcar que la topología del consenso es un resumen de los cladogramas y **no** es una **hipótesis de filogenia**.

En el consenso estricto, solamente los nodos compartidos por todos los cladogramas son incluidos en el resumen final. Este método es el más conservativo, otros métodos presentan la información de algunos nodos, como el consenso de la mayoría, presenta solo los nodos que se encuentran por encima del valor de corte. Otros dos tipos de consensos, que en general no son usados, son el de Bremer y de Nelson, sus respuestas son muy similares al consenso estricto y sólo difieren de este en casos muy particulares.

El consenso de Adams se basa en operaciones de conjuntos entre los nodos; es muy útil para mostrar cuáles son los taxa que producen inestabilidad en el cladograma, pero puede producir nodos que no se encuentran en ninguno de los cladogramas originales. Kearney (2002) ofrece una buena discusión sobre cómo combinar los resultados del consenso estricto y el de Adams.

Los consensos de la mayoría son muy populares, especialmente en los análisis moleculares, aunque su uso es **muy** discutible (Sharkey & Leathers

(2001) vea también Goloboff & Pol (2005)). En este tipo de consenso se hace un conteo de las veces que el nodo aparece en los diferentes árboles: si el nodo aparece en al menos la mitad de los árboles, o el valor de corte seleccionado, el nodo es incluido en el consenso. Es posible hacer cortes más estrictos. Es una convención colocar en forma de porcentaje la cantidad de veces en las que el nodo apareció.

Como se mencionó en la introducción de la práctica de búsquedas, los métodos de consenso también se usan para dar respuestas parciales, como búsquedas de *jackknife*, o el doble consenso de Goloboff & Farris (2001); o como resultados en el caso de los análisis bayesianos que usan el árbol obtenido en el consenso de la mayoría (Huelsenbeck & Ronquist, 2001).

Técnicas

En la mayor parte de los programas los consensos estrictos y de la mayoría están implementados (por ejemplo, TNT, PAUP*, POY, WinClada o Component). En algunos casos, la implementación del consenso de la mayoría es sólo hasta el 50 %, y el usuario decide que tan estricto hace el corte, eliminando los nodos que estén por debajo del valor de corte (un consenso estricto sólo retendría los nodos con soporte del 100 %).

En otros casos, los programas pueden elaborar el consenso de la mayoría, pero no reportan los porcentajes de aparición en los nodos (por ejemplo en TNT), o simplemente los visualizan pero no salvan el árbol con los porcentajes incluidos (TNT). Es importante que si va a usar esta clase de consenso, evalúe cómo recuperar los reportes de frecuencias si usa TNT.

Finalmente hay que alertar un poco acerca de la resolución de los cladogramas usados para generar los consensos. La mayor parte de los programas usa los árboles perfectamente dicotómicos, por lo que pueden incluirse ramas no soportadas; así, al hacer un consenso es necesario eliminar tales árboles con ramas no soportadas. La mayoría de los programas actuales incluyen opciones para controlar la salida: incluir o no incluir árboles con nodos de longitud cero (por ejemplo TNT, NONA, WinClada, MacClade y PAUP*), otros no (Component).

A partir de la matriz de datos: `datos.consenso.dat`,
en WinClada, PAUP*, POY o TNT:

1. Abra la matriz de datos, y realice una búsqueda convencional (ver práctica 3, página 15). Guarde los árboles encontrados.

2. Realice y salve un consenso estricto, un consenso de la mayoría al 50 %, y en caso de que el programa lo permita, incluya 75 % y 90 % de corte. En todos los casos reporte el número de nodos presentes en el árbol, y compare los grupos encontrados. Recuerde anotar las frecuencias de los grupos.
3. Para **WinClada** use los menús correspondientes en la sección de **win-clados**; el menú **Trees** tiene la entrada **consensus compromise**, donde hay la posibilidad de hacer consenso estricto, consenso estricto eliminando nodos no soportados (nelsen, que no debe confundirse con el consenso de Nelson) y consenso de la mayoría¹. Para salvar los árboles siga las instrucciones previas usadas en la práctica XXXX.
4. En **PAUP***:
 - Siga el mismo esquema, realice una búsqueda y calcule primero el consenso por defecto y posteriormente los consensos estricto y de Adams; guárdelos en un archivo, use la instrucción

```
> contree
```

- En caso de duda el nombre del comando y el signo ?: `contree ?` para que evalúe las opciones del comando; si desea guardar el consenso directamente en un archivo de árboles **debe** hacerlo desde esta instrucción.

5. En **TNT** use:

```
> nelsen
```

- para obtener el consenso estricto, si desea que el consenso sea el último árbol, use `nelsen*`
- Para el consenso de la mayoría, use la instrucción `majority` y `majority*` respectivamente.

¹una vez calculado el consenso de la mayoría, **WinClada** tiene algunos problemas posteriores en la forma como los gráficos, por ejemplo, los *Hashmarks* no se pueden activar (no son dibujados), y modificar la topología produce cambios inesperados en las frecuencias de los nodos.

- Para guardar los árboles de consenso directamente, después de calculados use

```
> save {strict}
```

o

```
> save {majority}
```

- En todos los casos **debe** haber abierto previamente el archivo de árboles. En la versión de menú puede usar **Trees** y la entrada **consensus** y escoger los diferentes tipos de consenso.

6. En POY, se pueden realizar consensos con la instrucción

```
> report(consensus)
```

- para el consenso estricto y `report(consensus:x)` para consensos de la mayoría, donde x es un entero mayor que 50.
- Si desea guardar los resultados en un archivo debe colocar el nombre entre comillas dobles. Por ejemplo:

```
> report ("mayoria.txt", consensus:75)
```

guarda el consenso de la mayoría al 75 % en el archivo `mayoria.txt`.

7. En R, cargue primero las bibliotecas: *ape* y *jrich*.

- En PAUP* haga un búsqueda y guarde los árboles iniciales en formato Newick en un archivo `arboles.phy` con las instrucciones:

```
> hsearch; savet file=arboles.phy format=Newick;
```

- Asigne los árboles salvados a un objeto denominado `arboles`:

```
> arboles <- read.tree("arboles.phy")
```

- Si lo desea, puede graficarlos con el comando `plot`.
- construya y grafique el consenso de la mayoría y el consenso de la mayoría pesado Sharkey et al. (2013).

```
> plot(wconsensus(arboles,collapse=F));  
plot(wconsensus(arboles,collapse=T))
```

- compare las dos topologías y mire los grupos que fueron colapsados.

Preguntas generales.

- Compare las topologías de los árboles encontrados: ¿difieren los resultados entre los programas? Revise si las frecuencias de los nodos comunes son iguales en los diferentes resultados.
- Al asignar las transformaciones (sinapomorfías) de cada nodo sobre el consenso, ¿cómo lo haría usted? Compare su aseveración con la implementada en los programas (recuerde la práctica de matrices).
- ¿Recomendaría usted el uso de consensos de mayoría como herramienta para resumir la información de los árboles iniciales?

Literatura recomendada

Goloboff & Farris (2001) [Presenta las técnicas de consensos rápidos].
 Miyamoto (1985) [presenta una crítica hacia la interpretación de los consensos en clasificaciones].
 Nixon & Carpenter (1996) [Una discusión clásica sobre consensos].
 Sharkey & Leathers, 2001 [Una crítica al uso y abuso del consenso de mayoría].

Bibliografía

- Carpenter, J. M. 1988. Choosing among multiple equally parsimonious cladograms. *Cladistics* 4:291–296. [doi:10.1111/j.1096-0031.1988.tb00476.x].
- 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].
- Farris, J. 1969. A successive approximations approach to character weighting. *Systematic Biology* 18:374–385.
- Farris, J. 2001. Support weighting. *Cladistics* 17:389–394. [doi:10.1111/j.1096-0031.2001.tb00133.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. 1993. Estimating character weights during tree search. *Cladistics* 9:83–91. [doi:10.1111/j.1096-0031.1993.tb00209.x].
- Goloboff, P. 1995. Parsimony and weighting: A reply to turner and zandee. *Cladistics* 11:91–104. [doi:10.1111/j.1096-0031.1995.tb00006.x].
- Goloboff, P. 1998. Tree searches under sankoff parsimony. *Cladistics* 14:229–237. [doi:10.1006/clad.1998.0068].

- Goloboff, P. & J. Farris. 2001. Methods for quick consensus estimation. *Cladistics* 17:S26–S34.
- Goloboff, P. & D. Pol. 2005. Parsimony and Bayesian phylogenetics. 148-162 Oxford.
- Goloboff, P. A., J. M. Carpenter, J. S. Arias, & D. R. Miranda-Esquivel. 2008. Weighting against homoplasy improves phylogenetic analysis of morphological data sets. *Cladistics* 24:758–773. [doi:10.1111/j.1096-0031.2008.00209.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.
- Huelsenbeck, J. & F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Kearney, M. 2002. Fragmentary taxa, missing data, and ambiguity: mistaken assumptions and conclusions. *Systematic Biology* 51:369–381.
- Kjer, K. M., R. J. Blahnik, & R. W. Holzenthal. 2001. Phylogeny of trichoptera (caddisflies): characterization of signal and noise within multiple datasets. *Systematic Biology* 50:781–816.
- Kluge, A. 1997. Sophisticated falsification and research cycles: Consequences for differential character weighting in phylogenetic systematics. *Zoologica Scripta* 26:349–360. [doi:10.1111/j.1463-6409.1997.tb00424.x].
- Kluge, A. & J. Farris. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18:1–32. [doi:10.1093/sysbio/18.1.1].
- 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.
- Miyamoto, M. M. 1985. Consensus cladograms and general classifications. *Cladistics* 1:186–189. [doi:10.1111/j.1096-0031.1985.tb00421.x].
- Neff, N. 1986. A rational basis for a priori character weighing. *Systematic Zoology* 35:110–123.
- Nelson, G. 1979. Cladistic analysis and synthesis: Principles and definitions, with a historical note on adanson's familles des plantes (1763–1764). *Systematic Biology* 28:1–21. [doi:10.1093/sysbio/28.1.1].
- Nixon, K. & J. Carpenter. 1996. On consensus, collapsibility, and clade concordance. *Cladistics* 12:305–321.
- Page, R. & E. Holmes. 1998. *Molecular Evolution: A Phylogenetic Approach*. Evolutionary Biology Series Wiley.
- 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.
- Sharkey, M. & J. Leathers. 2001. Majority does not rule: the trouble with majority-rule consensus trees. *Cladistics* 17:282–284.
- Sharkey, M. J., S. Stoelb, D. R. Miranda-Esquivel, & B. J. Sharanowski. 2013. Weighted compromise trees: a method to summarize competing phylogenetic hypotheses. *Cladistics* 29:309–314. [doi:10.1111/cla.12000].
- Sokal, R. 1983. A phylogenetic analysis of the caminalcules. i. the data base. *Systematic Zoology* 32(2):159–184.
- Springer, M. S., R. W. DeBry, C. Douady, H. M. Amrine, O. Madsen, W. W. de Jong, & M. J. Stanhope. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution* 18:132–143.
- Strong, E. & D. Lipscomb. 1999. Character coding and inapplicable data. *Cladistics* 15:363–371.
- Strugnell, J. M., M. D. Norman, M. Vecchione, M. Guzik, & A. L. Allcock. 2014. The ink sac clouds octopod evolutionary history. *Hydrobiologia* 725:215–235. [doi:10.1007/s10750-013-1517-6].
- Swofford, D. 1991. When are phylogeny estimates from molecular and morphological data incongruent? *Phylogenetic Analysis of DNA Sequences*. Oxford.

- Wiens, J. 2001. Character analysis in morphological phylogenetics: problems and solutions. *Systematic Biology* 50:689–699.
- Wiley, E. & B. Lieberman. 2011. John Wiley & Sons.