

## Evolución y Filogenia de Arthropoda

Sección I: Conceptos y métodos en el estudio de la filogenia

### MÉTODOS DE ANÁLISIS EN LA RECONSTRUCCIÓN FILOGENÉTICA

Emilio J. López Caballero y Gonzalo Pérez Suárez

Departamento de Biología Animal. Universidad de Alcalá  
Ctra N II, Km 33'6; 28871. Alcalá de Henares. Madrid. España  
emilioj.lopez@alcala.es

#### Resumen

Se realiza una breve exposición general de la historia de la clasificación biológica, describiendo los procesos intelectuales y científicos que, en cada época, marcaron el pensamiento y los criterios para clasificar los organismos. Se describen los principales caracteres utilizados en taxonomía, la variación de su importancia a lo largo del tiempo y su utilidad para cada sistema de clasificación y se hace un breve estudio crítico de los tres principales métodos taxonómicos: Fenético, cladista y la sistemática ortodoxa o evolutiva.

Se ha realizado una breve exposición de la metodología de reconstrucción filogenética y su forma de considerar los caracteres. Se define el concepto de homología y los criterios para su comprensión así como el concepto de parsimonia y los algoritmos de cálculo de los árboles filogenéticos utilizados por los principales programas informáticos de reconstrucción.

**Palabras clave:** Clasificación, Taxonomía, Caracteres taxonómicos, Fenética, Cladismo, Sistemática evolutiva, Homología, Parsimonia, Cladograma, Árbol filogenético, Fenograma, Polarización, Compatibilidad.

#### Analysis methods in phylogenetics reconstruction

##### Abstract

We carried out a brief general exhibition of the history of the biological classification, describing the intellectual and scientific processes that marked the thought and the approaches to classify the organisms, in each time. The main characters used in taxonomy are described, the variation of their importance along the time and their utility for each classification system and a brief and criticise study is made of the three main taxonomic methods: Fenetist, cladist and the one orthodox or evolutionary systematic.

We carried out a brief exhibition of the methodology of phylogenetic reconstruction and their form of considering the characters. The homology concept and the approaches for their understanding as well as the concept of parsimony and the algorithms of calculus of phylogenetic tree, used by the main computer programs of reconstruction, is defined.

**Key words:** Classification, Taxonomy, Taxonomic characters, Phenetic, Cladism, Evolutionary systematic, Homology, Parsimony, Cladogram, Phylogenetic tree, Phenogram, Polarity, Compatibility.

### 1.- BREVE HISTORIA DE LA CLASIFICACIÓN. LOS CARACTERES TAXONÓMICOS

Mientras que todas las ciencias se encuentran hoy día en una situación comprometida debido a la multiplicidad y cantidad de información, las ciencias biológicas han estado en esa situación desde hace mucho tiempo. Esto se debe a que muchos organismos se encuentran en muy variadas situaciones y porque gran cantidad de ellos han atraído la atención del hombre. Bastante antes de la invención de la escritura y, por ello, del comienzo de la historia, el hombre prehistórico debió haber acumulado un conocimiento enciclopédico de su medio orgánico en términos de qué era vivo y porqué, qué era amigable o enemigo, alimento o veneno.

La existencia de conceptos tales como alimento, implica que los organismos se han agrupado y esto implica una clasificación. Simpson (1961) señaló que la necesidad de juntar seres en categorías es una característica general de los organismos vivos e indica que *Amoeba*, de forma semejante al hombre,

no puede existir sin la apreciación de la categoría "alimento". El pensamiento humano, en general, como se refleja en el lenguaje, parece depender fuertemente del reconocimiento de grupos, particularmente los grupos de objetos a los que damos nombres colectivos. Tenemos que notar que el nombre "insecto" implica dos procesos mentales, primero el de considerar, de forma "subconsciente", las semejanzas que poseen "todos" los insectos y segundo, considerar las diferencias entre los "insectos" y los "no insectos".

La necesidad de una clasificación taxonómica fue señalada convincentemente en 1970 por Savory: "encontramos animales tan diferentes en el mundo que no podemos tratarlos separadamente e, incluso, si quisiéramos hacerlo, la tarea estaría más allá de la capacidad de la mente y de la memoria humana. La clasificación nos viene impuesta por las limitaciones del cerebro". En el contexto en que se expresa Savory y en

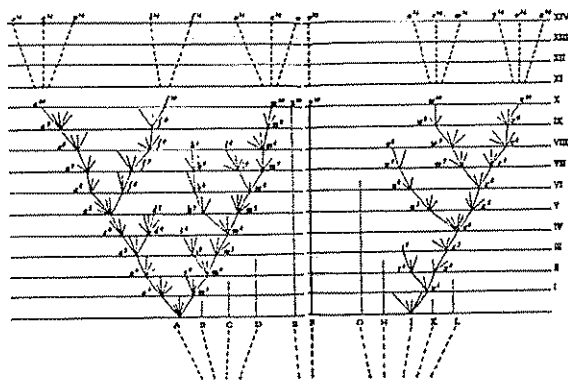


Fig. 1. Diagrama de Darwin (Tomado del original)

el de muchos otros autores (Mayr et al., 1953, p. e.), el concepto de clasificación biológica está restringido al agrupamiento de organismos en taxones – desde el Phylum hasta el género y la especie – por sus atributos estructurales.

Los taxónomos empíricos no pueden proporcionar una explicación causal para el hecho de ser capaces de agrupar las especies de acuerdo con “relaciones” o “afinidades”. Cuando Strickland (1840) definió afinidad como “la relación que hay entre dos o más miembros de un grupo natural, o en otras palabras, una coincidencia en caracteres esenciales”, dejó las palabras clave, “natural” y “esencial”, sin definir. Darwin vino a rellenar ese hueco y mostró por qué hay grupos naturales y por qué comparten caracteres “esenciales”. Como señaló Simpson (1961), hay dos tipos de relaciones que se pueden utilizar para hacer un agrupamiento general, a saber, la “asociación por semejanza” y la “asociación por contigüidad”, siendo conocida la primera generalmente como “sistemática” y la segunda como “taxonomía”. Las definiciones aceptadas de estos términos incluirán, lógicamente, esos conceptos y así la definición de Sistemática sería “El estudio científico de los tipos y diversidad de los organismos y de cualquiera y de todas las relaciones entre ellos”, mientras que la Taxonomía sería “El estudio teórico de la clasificación, incluyendo sus bases, principios, procedimientos y reglas”.

Darwin fue por ello, también, quien proporcionó la teoría básica de la clasificación biológica, aunque Linneo sentara unas primitivas bases sobre métodos. Nadie antes de Darwin había señalado tan inequívocamente que los miembros de un taxón son semejantes porque son descendientes de un ancestral común. Esta idea, no obstante, no era enteramente nueva, ya que Buffon, el propio Erasmus Darwin y algunos evolucionistas alemanes, habían jugado con la posibilidad que especies similares, tales como caballo y asno, o todos los gatos, podían haber derivado de una especie ancestral. Linneo, en sus últimos años, había sugerido que los miembros de un taxón más alto pudieran ser producto de hibridación. Ni Linneo ni Buffon convirtieron estas especulaciones en una teoría de clasificación ni de evolución.

Muy pocos comprenden que Darwin fue el fundador de todo el campo de la taxonomía evolutiva. Como señaló correctamente Simpson “la taxonomía evolutiva proviene explícitamente y casi exclusivamente de Darwin”. Pero esto significa que no sólo se explica automáticamente la teoría de la descendencia común para muchos de los grados de semejanza entre los organismos, sino que Darwin desarrolló también una teoría bien fundamentada con una declaración detallada de métodos y dificultades (consultar el capítulo decimotercero de la primera edición del *Origin*...). Él hace

referencia a su famoso diagrama (Fig. 1), en el que cada una de tres especies cogenéricas del Silúrico (A, F e I), tienen descendientes modernos de rango muy diferente. La línea derivada de la especie F ha cambiado tan poco que todavía contiene un único género, mientras que sus dos grupos hermanos, A e I, constituyen, en la actualidad, diferentes familias o incluso ordenes. Para la determinación del “rango” Darwin, generalmente, aplicó el principio del grado de divergencia, más que el de proximidad, al punto de ramificación. En particular, rechazó la idea, tan ampliamente extendida entre los botánicos de los siglos XVII y XVIII, y también entre los zoólogos, desde Cuvier, de que cuanto más importante es una estructura para la supervivencia y la perpetuación de un organismo, más importante es también esa estructura para su clasificación, lo que ilustra con varios ejemplos como el del valor altamente desigual que tienen las antenas como un carácter taxonómico en diferentes familias de insectos y para rechazar aquella idea señala explícitamente: “Que la mera importancia fisiológica de un órgano no determina su valor para la clasificación se puede demostrar por el hecho de que en grupos relacionados, en los que el mismo órgano... tiene prácticamente el mismo valor fisiológico, su valor para la clasificación es ampliamente distinto”.

El consejo de Darwin no constituye un rechazo de la importancia de la selección natural, lo que dice es que las adaptaciones especiales pueden implicar solamente una porción limitada de la dotación genética de un grupo y por ello ser menos informativas que, p. e. el “hábito” general. Además, las adaptaciones especiales pueden ser adquiridas independientemente en varias líneas evolutivas no relacionadas, o sea que están sujetas a convergencia.

Para la evaluación de los caracteres el propio Darwin propone ciertas reglas. Considera de elevada importancia taxonómica aquellos que son **constantes** en varios grupos grandes. Llama la atención sobre los errores que pueden cometer los taxónomos al considerar “semejanzas espúreas” causadas por **evolución convergente** (semejanzas análogas o adaptativas). Otras dos recomendaciones son particularmente importantes, como es el rechazar las semejanzas debidas a una descendencia de semejanzas superficiales producidas por convergencia, y la otra se refiere al “peso” de los caracteres. Esta segunda es importante ya que algunos caracteres tienen un contenido de información mayor que otros. Este peso viene definido por la correlación de ese carácter con las partes de la clasificación establecidas de manera más firme (por varios métodos de comprobación). En la actualidad se combina este método con varias técnicas de computación para determinar esa importancia (“peso”).

En lo que se refiere a la metodología de la clasificación, la revolución darvinista sólo tuvo un pequeño impacto. Es evidente que el punto crucial para el giro en la historia de la taxonomía fue el abandono del esencialismo y de la “clasificación descendente”, ocurrido antes de 1859. La decisiva contribución hecha por Darwin a la taxonomía fue doble: por su teoría de la descendencia común que proporcionó una explicación de la existencia de la jerarquía linneana, y por la homogeneidad de los taxones en una clasificación “natural”, restaurando el principio de continuidad entre los grupos de organismos rechazado por Cuvier y por los *Naturphilosophen* alemanes en su teoría de los arquetipos.

A partir de 1880 hubo un declive, gradual y visible, en cuanto al interés por los estudios taxonómicos y filogenéticos, debido a varias razones. Quizá el factor más importante para este declive fuese la competencia creciente de otras ramas de la biología experimental en las que se realizaron excitantes descubrimientos como la citología, genética mendeliana,

fisiología, bioquímica, etc., que condujeron a los biólogos más jóvenes hacia esos campos de investigación, por lo que el espléndido inicio que hizo Darwin al desarrollar una teoría y una metodología de la taxonomía fue ampliamente ignorado en el período post-darvinista, entrándose en una etapa en la que no se respetaron esquemas anteriores, ni reglas de nomenclatura, ni criterios sobre la validez de los caracteres, etc. (Mayr, 1982, p.220).

Hacia la década de 1930 el interés resurgió y fue finalmente estimulado al inicio de los 40 por dos libros de J. S. Huxley, "La nueva sistemática" de 1940 y "La evolución, la síntesis moderna" de 1942. La tesis defendida es que las unidades sistemáticas eran las poblaciones y no los individuos y que las especies no deberían estar ligadas al ejemplar (o ejemplares) tipo definidos por el autor original. Se estudió más intensamente la variabilidad intraespecífica y se propusieron métodos estadísticos para la distinción entre especies y "dentro" de las especies. Ciertamente, gran cantidad de estas ideas no eran realmente nuevas y, por ejemplo, el deseo de conservar grandes series de ejemplares de cada especie, había sido ya aplicado desde hacía tiempo en los museos y sirva de ejemplo de lo señalado lo que Mayr et al. 1953 recogen de Baird, 1854, relativo a la Smithsonian Institution: "El objetivo de la Institución al hacer sus colecciones no es simplemente poseer las diferentes especies, sino también determinar su distribución geográfica, por lo que es de importancia poseer de cada localidad series tan completas como sea posible..."

La obra "La nueva sistemática" fue valiosa con relación a que dirigió la atención hacia los mecanismos de especiación y a que ligó las actividades de los genéticos y biogeógrafos con la corriente principal de la biología sistemática. No obstante la taxonomía falló en relación con recuperar su vigor primitivo y, el impacto creado por "La nueva sistemática" fue menor de lo que sus proponentes habían esperado. Posiblemente el "biólogo medio" se asustó por la estadística, aunque no hubiera necesidad de conocer sus implicaciones y podía ser usada con precaución y de manera simple. Los conocimientos a este respecto se limitaban en ese momento a la interpretación de una curva bimodal, histogramas, diagramas de dispersión, desviaciones estándar, chi-cuadrado y coeficientes de diferencias medias.

Parte de la relativa falta de interés por "La nueva sistemática" podía ser debida a un cambio en la racionalidad de los criterios para la identificación de las especies. Pocos parecían interesados en el análisis exhaustivo de las afinidades de los taxones y muchos consideraban la identificación como un medio para otro fin, por ejemplo, en ecología y no hay duda de que este tipo de estudios dependían completamente de los expertos en taxonomía. Chace (1969) señalaba correctamente: "...un problema importante y, virtualmente insuperable, que arrastra cualquier estudio ecológico de las aguas tropicales hoy día, es el imperfecto conocimiento de la fauna existente...". En nuestros días, con el fervor por la conservación del medio, de las especies, de la biodiversidad, es fantástico que solamente aquellos investigadores que nos puedan decir qué especie ocupa un determinado medio, sean un pequeño grupo (y cada vez menor) de taxónomos mal considerados, aunque, eso sí, dedicados con fervor a su trabajo.

Quizá el excelente desarrollo constructivo de aquel período fue dar un significado ecológico a los taxones más elevados. Se reconoció que los taxones más altos estaban formados por especies que, en su conjunto, ocupaban un nicho específico o una "zona adaptativa" (en palabras de Mayr, 1982). Como "La nueva sistemática" se concentró ampliamente sobre el nivel específico, no ofreció soluciones a las necesidades de la macrotaxonomía y la ayuda tenía que venir

de otra parte. De manera independiente se propusieron dos soluciones drásticamente distintas: la **Fenética numérica** y la **Cladística**. Ninguna de ellas se propuso como reforma a lo que había sino como un reemplazo revolucionario de los procedimientos existentes, manteniendo su vigencia la **Metodología evolutiva** o tradicional, desarrollada sin solución de continuidad por una escuela y cuyos principios se pueden encontrar en Mayr (1969).

Antes de realizar una breve descripción de cada una de estas aproximaciones al conocimiento de la estructura de las relaciones en el mundo de los seres vivos, veamos los caracteres taxonómicos, tradicionales y nuevos, y qué uso hacen de ellos estos tres métodos de clasificación.

La base de estos tres métodos es el análisis y la evaluación de los caracteres taxonómicos. Una insuficiencia de caracteres informativos es la razón más frecuente de porqué no se puede resolver la falta de coincidencia entre las clasificaciones propuestas por uno u otro método. No deja de sorprender que la queja más frecuente que expresa un taxónomo, es que el grupo de animales con los que trabaja no presenta suficientes caracteres para permitir una decisión inequívoca sobre sus relaciones. Dos fenómenos contribuyen a esta dificultad, primero el hecho, bien conocido, de que el fenotipo, en determinados grupos de organismos, está notablemente "estandarizado" (como en los cientos de especies de *Rana* o en los miles de especies de *Drosophila*) y por ello proporciona solamente unos pocos indicios morfológicos de las relaciones, y segundo, que una desviación del tipo estándar generalmente sólo afecta a un complejo funcional simple, correlacionado con alguna adaptación especial "ad hoc". El cambio a una nueva fuente de alimento o la adopción de un nuevo conjunto de señales de cortejo, puede, a veces, tener como efecto una notable adquisición morfológica que puede ser dividida en un número considerable de caracteres individuales. Considerar éstos como caracteres separados nos podría conducir a error, puesto que hablando en términos filogenéticos, son puramente un reflejo de un único cambio funcional. Darwin ya alertó frente a muchas de las especializaciones "ad hoc".

Una dificultad hallada por los taxónomos con mucha frecuencia está representada por el conflicto que se establece entre las conclusiones basadas sobre diferentes estructuras. El estudio de las extremidades puede indicar que un taxón *b* está más claramente relacionado a otro *a*, mientras que los rasgos del tracto intestinal sugieren que el taxón *b* está más próximo al taxón *c*. En estos casos el estudio de rasgos adicionales de las extremidades o del tracto intestinal raramente proporcionan una solución satisfactoria. Se pueden encontrar numerosos ejemplos de este tipo de casos en cada uno de los taxones más elevados y debido a ello, los taxónomos, en las últimas décadas han dedicado gran atención a la búsqueda de nuevos caracteres taxonómicos. Aunque los análisis morfológicos cuidadosos proporcionan continuamente nuevos caracteres, los caracteres no-morfológicos juegan un papel, que se incrementa día a día, en el establecimiento de clasificaciones y entre ellos cabría señalar caracteres tales como componentes del comportamiento, ciclo de vida y anual, fisiología, ecología (por ejemplo, utilización del nicho), parásitos y cualesquiera otros atributos de cada organismo. Muchos de ellos son útiles para la discriminación de especies y otros son indicativos de relaciones entre grupos de especies.

La distribución geográfica proporciona, a menudo, pistas inesperadas. La búsqueda del grupo relacionado más próximo en un área geográfica fácilmente accesible no se puede utilizar en el caso de algunos grupos con unas facilidades de dispersión muy elevadas o en grupos que han quedado relegados a áreas muy restringidas, pero sí es muy útil en otros

casos (como ha demostrado Simpson en numerosas ocasiones en los animales). Una combinación de análisis cladísticos y biogeográficos, como demostró Hennig y sus seguidores, a veces es particularmente reveladora.

El desarrollo tecnológico está permitiendo, en la actualidad, llegar a límites insospechados en la estimación de la estructura **fenética** de un organismo. Como ejemplo podemos citar la microscopía electrónica de barrido (Scanning) que permite apreciar microestructuras en relieve con una gran capacidad de definición, y la gran variedad de técnicas bioquímicas para medir indirectamente la "distancia genética" entre los organismos. Estas técnicas han alcanzado gran popularidad entre los sistemáticos y en ellas se incluyen la hibridación de ADN, electroforesis, electroenfoque, isoelectroenfoque, cromatografía líquida de alta presión (HPLC), análisis de las actividades enzimáticas e índices de semejanza inmunológica, etc. que permiten detectar los polimorfismos de proteínas existentes en un órgano o estructura de un organismo determinado.

Estos y otros tipos de datos (recogidos de forma más general en Tabla I) son utilizados a diario por los taxónomos. La base fundamental para la anatomía comparada (o la bioquímica comparada, o la etología comparada, etc.) es el concepto de **homología**. Veamos en un cuadro comparativo algunas de las características generales de los tres métodos de clasificación y el uso que cada uno hace de una serie de atributos.

Una relación, no exhaustiva, de los diferentes tipos de caracteres taxonómicos se recoge en la Tabla I.

**Tabla I**

Relación (no exhaustiva) de los caracteres taxonómicos  
(Tomado de Mayr, 1969 modificado)

**1.- MORFOLÓGICOS**

- morfología general externa
- estructuras especiales (p. e. genitalia)
- morfología interna (= anatomía)
- embriología
- cariología (y otras diferencias citológicas)

**2.- FISIOLÓGICOS**

- factores metabólicos
- secreciones corporales
- factores de esterilidad génica
- diferencias serológicas y en general en proteínas
- otras diferencias bioquímicas

**3.- ECOLÓGICOS**

- hábitats y hospedadores
- alimento
- variaciones estacionales
- parásitos
- reacciones del hospedador

**4.- ETOLÓGICOS (COMPORTAMIENTO)**

- mecanismos de aislamiento reproductor
- patrones de comportamiento
- otros mecanismos de comportamiento

**5.- BIOGEOGRÁFICOS**

- tipos de distribución geográfica generales
- relaciones simpátricas-alopátricas entre las poblaciones

En la Tabla II se pueden comprobar las relaciones entre los tres principales sistemas de clasificación, atendiendo a la consideración que cada uno de ellos tiene para algunos de los tipos de caracteres/atributos que podemos considerar en los taxones.

## 2.- BREVE DESCRIPCIÓN Y COMPARACIÓN CRÍTICA DE LOS TRES SISTEMAS DE CLASIFICACIÓN

La **Fenética**, denominada a veces también como **Taxonomía numérica** o **Taximetría**, se desarrolló como consecuencia de la disponibilidad creciente de computadores al final de los años 50. La base filosófica para la "fenética numérica" es el argumento de que como nunca es posible conocer con certeza cual de las diferentes filogenias en competencia es la correcta, la descripción real de los individuos de un grupo ayudará a conocer la evolución de ese grupo, pero nada más. Por lo anterior, los organismos deberían ser clasificados estrictamente en función de la conveniencia, en lugar de establecer clasificaciones basadas sobre reconstrucciones hipotéticas de la historia filogenética de un grupo de animales o plantas, o sea, como los libros en una biblioteca.

Los "fenetistas" analizan los datos mediante uno o más programas de ordenador que harán que se agrupen las OTU's (Unidades Taxonómicas Operativas, concepto introducido por Peter H. A. Sneath y Robert R. Sokal, 1973) en función de índices de semejanza, o de diferencia, global y producen un diagrama ramificado (dendrograma, Fig. 2) denominado **fenograma**. Este se puede transformar en una clasificación y esta transformación requiere una considerable subjetividad, al no existir una guía clara que determine qué nivel de semejanza (o de diferencia) es suficiente para establecer los diferentes niveles en una clasificación.

Durante los años 60, muchos científicos fueron atraídos por la fenética numérica, pero hacia mitad de los 70, su popularidad decreció de forma notable por varias razones. En primer lugar y aunque se desarrollaron una gran cantidad de programas para analizar los datos, cuando el mismo conjunto de datos se analizaba con distintos programas producían, con frecuencia, diferentes fenogramas. En segundo lugar los datos se manipulan de tal manera, en la mayoría de los programas, que el ideal de producir un análisis lo menos distorsionado posible resulta inviable. En tercer lugar, los fenogramas producidos tienden a generar grupos polifiléticos al no tener en cuenta la posible "convergencia" de los caracteres, ni considerar las homologías. Por último, muchos "fenetistas" sufrieron el síndrome de la "caja negra", al desconocer lo que el ordenador estaba haciendo con sus datos. Quizá, uno de los errores más importantes de los "fenetistas" sea el de haber considerado a los organismos como "objetos" inanimados cuando, coincidiendo con G. G. Simpson, "los miembros de un taxón son semejantes porque comparten una herencia común y no se ubican en un taxón porque sean semejantes".

La **sistemática cladística** (cladismo o **sistemática filogenética**) tuvo su origen en el libro publicado en 1950 por Willi Hennig (traducción castellana de 1968). Desde entonces su popularidad ha ido en aumento y con los años el cladismo ha ido mucho más allá del contexto planteado por Hennig originalmente. La meta del cladismo es producir hipótesis "comprobables" de las relaciones genealógicas entre grupos monofiléticos de organismos. Como metodología está basado completamente sobre la "descendencia común", o sea, la genealogía estricta. El dendrograma usado por los cladistas es denominado **cladograma** o **árbol** (Fig. 3) y está construido, únicamente, para mostrar la genealogía, es decir, las

Tabla II

Relaciones entre los tres principales sistemas de clasificación, según tipos de caracteres/atributos por taxon.

ATRIBUTOS	FENÉTICA	CLADÍSTICA	SISTEMÁTICA ORTODOXA
Muestra las relaciones mediante un árbol o una clasificación	Semejanza o desemejanza global	Genealogía	Genealogía + semejanza/desemejanza global (afinidad genética)
Semejanza evolutiva <sup>a</sup>	<b>Usados todos los tipos</b>	Solamente apomorfías	Solamente homologías
Peso de los caracteres	No usado	Generalmente no usado	Usado
Homología	No considerada	<b>De importancia capital</b>	Importante
Fósiles	No usados	Pueden ser considerados pero no de más importancia que las especies vivas	<u>Pueden ser muy importantes</u>
Datos ecológicos y evolutivos	No usados	Raramente usados	<u>Pueden ser muy importantes</u>
Tasas de evolución	No consideradas	No consideradas	<b>Muy importantes</b>
Transformación del árbol en una clasificación	Sin reglas generales; para delimitar los taxones se escogen niveles arbitrarios de semejanza/desemejanza global	La clasificación muestra precisamente modelos de ramificación o cladogramas	La clasificación refleja tanto modelos de ramificación como grados de diferencia entre taxones

<sup>a</sup> Apomorfias, Plesiomorfias, Convergencias, paralelismos, evolución regresiva

relaciones ancestral-descendiente. Como se expresa en el cuadro expuesto anteriormente, el valor principal en los análisis se asigna a las homologías y, de ellas, el peso recae sobre el concepto de caracteres homólogos primitivos frente a derivados o más recientes, identificándose pues las homologías estrictamente como plesiomorfias y apomorfias, respectivamente. Todos los demás conceptos relacionados los definiremos más adelante.

Un análisis cladístico clásico consta de: la identificación de los caracteres homólogos de los organismos que estemos estudiando; establecimiento de la “dirección del cambio” del carácter, o evolución del carácter (“análisis de la polaridad”); construcción del cladograma de los taxones que poseen el carácter analizado. El paso final debería ser la conversión del cladograma en un esquema de clasificación.

Una de las críticas más fuertes que se pueden hacer al cladismo es que el proceso fundamental de la especiación sólo lo considera como la separación de una especie ancestral en dos especies hermanas, a pesar de la existencia de otros numerosos modelos de especiación. Para la cladística, una vez que aparece una especie, se debe iniciar una rama en el cladograma y las dos (o más) líneas representan grupos hermanos, a pesar de que la especie original pueda continuar existiendo. Aceptar esta suposición de trabajo es necesaria para facilitar los métodos de construcción de los cladogramas, como se verá inmediatamente.

La **sistemática ortodoxa (evolutiva)**. Tanto la fenética como la cladística han contado, la primera, y cuenta, la segunda, con numerosos seguidores, pero la mayoría de los taxónomos, aún habiendo aceptado los avances de uno o de otro sistema, han conservado la metodología tradicional de la clasificación. Ésta consiste en intentar representar en la clasificación no sólo la ramificación de las líneas filéticas sino también su posterior divergencia y esto se puede hacer indicando en la disposición de distintos taxones si se han vuelto, o no, radicalmente diferentes de los que deberían ser sus grupos hermanos, por la invasión de un nuevo nicho o zona adaptativa. El resultado es la conversión del cladograma en un **filograma** en terminología de Mayr (1969) o “árbol evolutivo” (Fig. 4). El uso del término “evolutiva” es quizá desafortunado pues parece que el utilizar la teoría evolutiva fuese de uso particular para esta filosofía de clasificación, aunque se hace atendiendo a que se siguen los postulados de

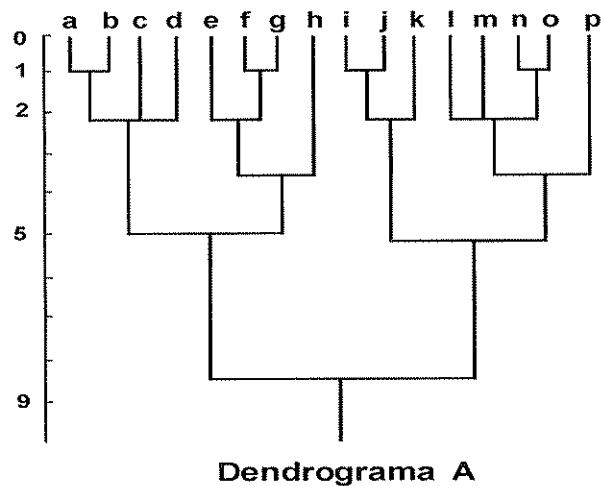
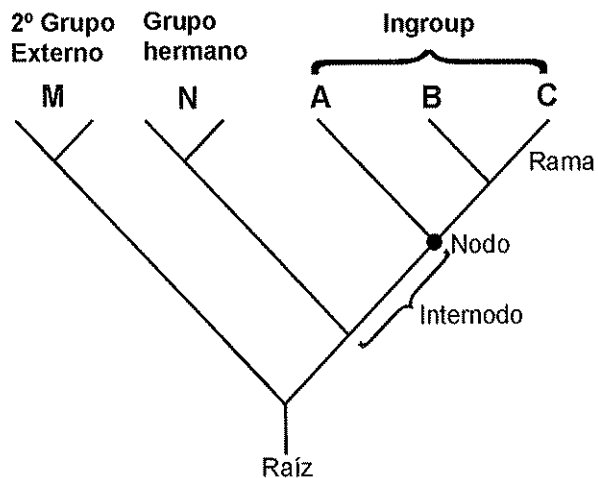


Fig.2.- Dendrograma hipotético de una hipotética matriz de datos



Árbol (enraizado) para el grupo ABC y dos de sus Grupos Externos N (Grupo hermano) y M

Fig. 3.- Representación de un cladograma con algunos términos de nomenclatura.

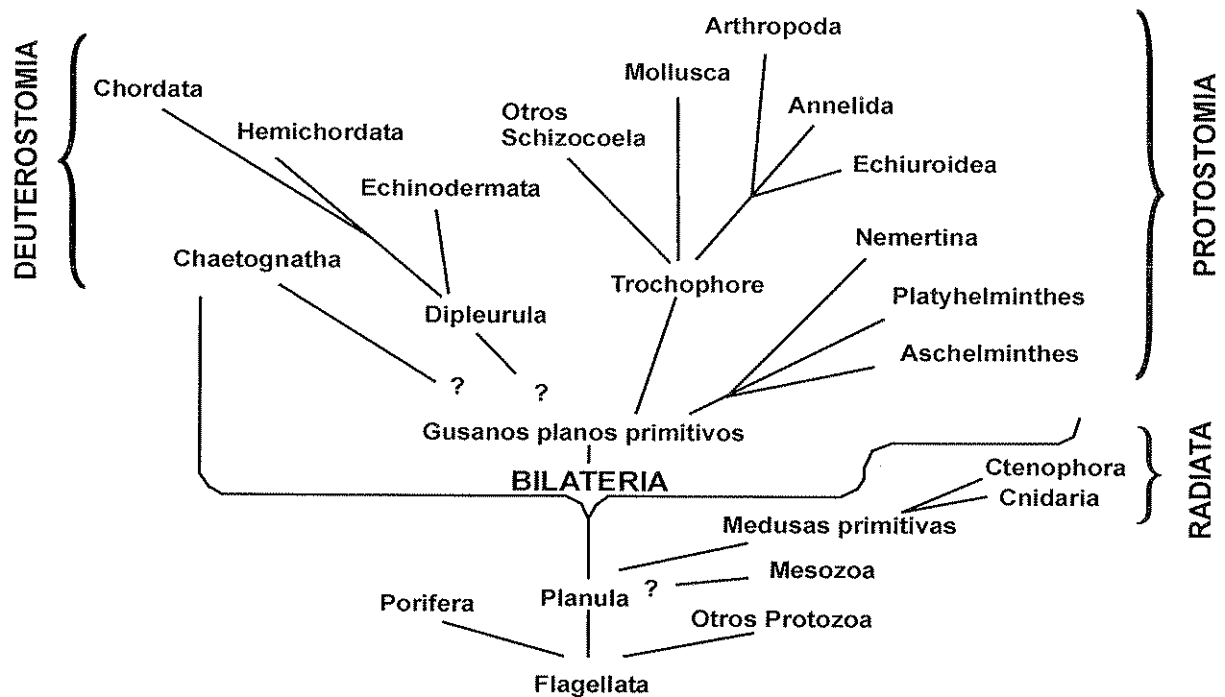


Fig. 4.- Árbol evolutivo según Hyman, 1940 (tomado de Brusca y Brusca, 1990).

Darwin en mayor o menor medida. La principal diferencia entre este método y la cladística, está en el considerable peso que se le da a las "autapomorfias" (caracteres derivados adquiridos por un grupo hermano pero no por otro). Frente al término "clade" de la cladística, la sistemática ortodoxa emplea el término "grado" que representa niveles de cambio evolutivo que deben quedar reflejados en la clasificación.

Muchos de los contenidos filosóficos de esta metodología sistemática se han expuesto al principio, al comentar el bagaje histórico general de la problemática de la clasificación en Biología. A pesar de lo indicado conviene señalar que la premisa básica de este método es la teoría de la evolución por selección natural y aunque se ha desarrollado gradualmente, más allá de la tradición linneana, las metodologías o reglas específicas que se utilizan, de forma general, para crear clasificaciones, no han sido nunca expresamente formuladas. Muchos de los procesos se basan en la intuición o en hipótesis, expresadas muy a menudo de forma ambigua, sobre el proceso evolutivo. Por todas estas razones, los "árboles" producidos por este método no se prestan bien para ser comprobados.

No deberíamos finalizar esta primera parte sin volver a referirnos a los aspectos más "cuantitativos" de la Cladística. Los procedimientos metodológicos "numéricos" enfocados bajo una perspectiva filogenética tuvieron como marco fundamental el trabajo de Kluge y Farris (1969), primero de los estudios que trataron de formalizar el uso del "análisis de parsimonia" en la cladística, basada, precisamente, en los conceptos propuestos por Wagner (1961), que era uno de los fenetistas más señalados. Este estudio generó algunos de los conceptos básicos de los que comenzó a denominarse como "análisis cladístico", conceptos tales como **parsimonia**, **árboles óptimos**, **índices de consistencia (CI)**, y otros, etc. todos ellos se sumaron al arsenal conceptual y metodológico ya propuesto por Hennig. Posteriormente Farris (1970) influyó profundamente en la discusión sobre el método al generalizar el uso de la parsimonia en los análisis no enraizados (sin referencia a la polaridad de los caracteres) y este trabajo acabó de sentar las bases de la cladística numérica.

El paso siguiente en este camino fue el desarrollo de varios programas de ordenador que implementaban los análisis de parsimonia, pero, para la sistemática, esto no fue una novedad ya que los programas de análisis sistemático eran comunes entre los fenetistas. La abundancia y optimización de los programas de análisis cladísticos puede hacer parecer que este tipo de análisis, hoy día, es una actividad simple, casi automática. Nada más lejos de la realidad, como señalan recientemente Ferrarezzi y Marques (1997) "el uso de los métodos numéricos... debe ser mirado exclusivamente, más como una herramienta para ejecutar esa parte del análisis, ahorrando tiempo y proporcionando fiabilidad en cuanto a la precisión de los resultados obtenidos".

Un análisis de las relaciones filogenéticas de un grupo determinado se realiza en dos etapas:

1. Construcción de la lista de caracteres a analizar (que incluye todos los razonamientos anteriores al análisis numérico de la matriz de datos, como son el establecimiento de homología, su codificación y ordenación, etc.).
2. Elección del árbol óptimo para este conjunto de datos, según el criterio adoptado.

Los aspectos críticos de esta "cladística numérica" se pueden sintetizar, siguiendo a los autores antes citados, en las siguientes preguntas:

- ¿Cuándo utilizar un programa de análisis?
- ¿Cuál o cuáles de los diversos programas son los más apropiados?
- ¿Qué tipos de análisis son posibles?
- ¿En qué circunstancias los estamos realizando?
- ¿Cómo se deben interpretar sus resultados?

En el siguiente apartado se tratará de contestar a estas y otras preguntas, así como establecer la metodología básica y los conceptos acerca del tratamiento de los caracteres y de su análisis.

### 3.- MÉTODOS DE RECONSTRUCCIONES FILOGENÉTICAS

La reconstrucción filogenética se sustenta en el concepto darwiniano de "descendencia con modificación", de forma que los caracteres que se observan en las distintas especies, heredados a partir de un ancestro común, son indicativos de una relación genealógica. El postulado del que se parte es que la semejanza puede ser interpretada de forma racional bajo la óptica de una ascendencia común. El problema, por tanto, que plantea la reconstrucción filogenética es el de la inferencia del ancestro a partir de la observación de los caracteres que presentan los taxones terminales.

Hasta la aparición de la obra de Hennig (1950), las construcciones filogenéticas se establecían a partir de llamado "triple paralelismo" (*Threefold parallelism*), consistente en el apoyo de la anatomía comparada, la ontogenia y la paleontología. Su objetivo era encontrar las similitudes o semejanzas entre los taxones basándose en el análisis de los caracteres estudiados bajo este criterio de triple paralelismo, de forma que entre aquellos taxones que más se asemejaran, más próximo era su parentesco. La dificultad estriba en que la semejanza no nos es dada, sino que es interpretada. De los distintos criterios de esta interpretación surgen los diferentes métodos de análisis filogenéticos, y la consecuente controversia entre los biólogos.

Desde el punto de vista evolutivo, la semejanza puede ser dividida, simplemente, en homología y en homoplasia. La homología es el parecido heredado a partir de un ancestro común, mientras que la homoplasia es un parecido que no es consecuencia de una herencia común (Simpson, 1961: pág. 78). La homoplasia a su vez se subdivide en convergencia y en reversión. La convergencia, también llamada paralelismo, es la aparición independiente en dos o más especies de un mismo carácter; este carácter no había aparecido antes. La reversión es la aparición, de nuevo, del carácter ancestral, lógicamente este carácter ya había aparecido.

Tal como se ha comentado anteriormente, para los sistemáticos taximétricos (Sokal y Sneath, 1963) las relaciones entre taxones solamente pueden ser expresadas basándose en la similitud global manifestada a partir del cálculo de matrices de distancias e índices de similitud, para ello no se hace ningún tipo de distinción entre los caracteres, mezclando caracteres homólogos y homoplásicos.

Para los sistemáticos evolutivos (Simpson, 1961; Mayr, 1969) la similitud global propuesta por los taximétricos no puede generar reconstrucciones filogenéticas debido a las falsas similitudes aportadas por las homoplasias. Sólo las similitudes basadas en homologías pueden permitir establecer las filogenias.

Para los sistemáticos filogenéticos o cladistas (Hennig, 1950, 1966, 1968; Eldredge y Cracraft, 1980; Wiley, 1981; Nelson y Platnick, 1981) el concepto de homología debe diferenciarse en plesiomorfía (estado primitivo) y apomorfía (estado derivado), de forma que sólo el compartir apomorfías (sinapomorfías) por distintas especies, es indicativo de un parentesco común y los grupos así constituidos son monofiléticos.

En la actualidad dos son los campos de actuación en las reconstrucciones filogenéticas, resultado de sus dos distintas concepciones de la similitud. Una de ellas basada en la similitud global: concepción fenética, y la otra basada en el análisis de los caracteres, diferenciados en plesiomorfos, apomorfos y homoplásicos: concepción cladista.

### 4.- DEFINICIÓN Y CRITERIOS DE HOMOLOGÍAS

La definición de homología es clara y concisa, por homología se entiende todo aquello que se hereda a partir de un ancestro común, de forma que un carácter compartido por diferentes especies es homólogo si se heredó a partir de un ancestro común, compartido por dichas especies. En todo momento debe recordarse que la homología es una hipótesis sobre la genealogía. Así, uno puede reconocer una semejanza en determinados caracteres morfológicos mediante la anatomía comparada, y establecer la deducción evolucionista de que tal observación es debida a una ascendencia común. Por tanto, la homología no se observa, sino que se propone una hipótesis de homología a partir de una observación. Tal como expresa Fitch, en Lewin 1987: "Lo importante es distinguir entre la observación y la conclusión". La definición de homología y el criterio de reconocimiento de homología son esencialmente diferentes. Si bien la definición de homología es del todo explícita, se plantea el problema de ¿Cómo reconocer la homología sin conocer *a priori* la filogenia?.

El criterio para establecer la homología es triple (Patterson, 1982): criterio de semejanza, de no-coexistencia y de congruencia.

El criterio de semejanza, íntimamente ligado al principio de las conexiones o de idéntica posición, utilizado en la anatomía comparada: un órgano es homólogo en dos o más especies si con independencia de su forma o función posee idénticas conexiones con otros órganos.

El criterio de la no-coexistencia, permite distinguir la verdadera homología de la llamada homología seriada: dos caracteres genealógicamente homólogos no pueden existir en un mismo organismo. Patterson lo expone en un ejemplo didáctico sobre los ángeles: "la teoría según la cual el brazo del hombre y las alas de las aves son genealógicamente homólogos, debería ser rechazada si los ángeles fuesen entidades reales, pues presentan a la vez brazos y alas".

El criterio de congruencia que permite superponer los mismos árboles construidos a partir de diferentes caracteres. Los caracteres homólogos son congruentes. Este principio de congruencia se sustenta en el principio de parsimonia.

### 5.- EL MÉTODO CLADISTA

#### 5.1. El principio de parsimonia

El método cladista puede ser calificado como un método hipotético deductivo. La hipótesis versa sobre el sentido de las transformaciones de los caracteres, y las deducciones sobre las afinidades filogenéticas caracterizadas por el método. El principio de parsimonia (Edwards y Cavalli-Sforza, 1963, 1964) es el criterio metodológico adoptado por el cladismo para decidir, en un análisis filogenético que presenta incongruencias entre los caracteres, cual de los cladogramas posibles, el llamado árbol más corto, constituye la reconstrucción más probable que es representativa de la filogenia del grupo. El principio de parsimonia sostiene que la explicación más simple es preferible a las más complejas, es decir, aquella que requiera un menor número de cambios evolutivos, permitiendo tanto las reversiones como las convergencias de los caracteres. Este principio no sólo es de aplicación en análisis cladistas (Camin y Sokal, 1965; Kluge y Farris, 1969) sino también en análisis de distancias (Cavalli-Sforza y Edwards, 1967; Fitch y Margoliash, 1967).

La asunción del criterio de la parsimonia de que las transiciones en un carácter son intrínsecamente poco probables, es consecuencia de la tacañería del cambio evolutivo.



Este principio cumple con la finalidad a la que todo análisis filogenético debe estar sometido: la construcción de un esquema racional, y por tanto, no arbitrario. Tal esquema debe estar en todo momento disponible a su refutación o a su confirmación, al introducir nuevos caracteres o taxones. El árbol más corto es el que nos va a permitir tal tipo de control (Darlú y Tassy, 1993: pp 44).

## 5.2. Codificación, ordenación y polarización de los caracteres

La codificación es simplemente la aplicación de un símbolo, generalmente numérico, para cada estado diferente de un determinado carácter o serie de transformación. Los caracteres se codifican de forma que puedan ser analizados comparativamente. En el caso de caracteres con más de dos estados, su codificación puede además contener informaciones adicionales sobre sus conexiones evolutivas; en este caso diremos que los caracteres están ordenados. Si a esta ordenación se le asigna una orientación del cambio evolutivo, la serie de transformación estará polarizada, estableciéndose la relación apomórfica y plesiomórfica relativa de los estados de los caracteres. Al contrario de lo que frecuentemente se piensa, la incorporación de esta información de direccionalidad evolutiva en las series de transformación mediante la polarización *a priori* de los caracteres es, en general, totalmente innecesaria en el análisis cladista, salvo en casos de imposición de determinadas restricciones del criterio de parsimonia.

### 5.2.1.- Caracteres binarios

Aquellos que están constituidos por dos estados. Son los de codificación más simple pues automáticamente constituyen un par ordenado, representado bajo la forma de 0-1 o a-b, en el que uno de ellos representa el estado plesiomorfo y el otro el apomorfo, pero *a priori* no indican una orientación particular. Las transformaciones  $a \rightarrow b$  o  $b \rightarrow a$  tienen ambas la misma probabilidad, y se les asigna un paso evolutivo en su transformación. Ello no resta que en estudios en que se establezca previamente la polarización de los caracteres, concomitante a su codificación, se indique la dirección de la transformación. En estos casos el binomio  $a \rightarrow b$  se define como polarizado y por conveniencia se asigna el código 0 al estado plesiomorfo y el 1 al apomorfo.

### 5.2.2.- Caracteres multiestado

Los caracteres multiestado son aquellos que presentan más de dos estados en su serie de transformación, y cuyas relaciones o conexiones de transformación de un estado a otro pueden ser de dos tipos:

- a. **Serie de transformación desordenadas:** aquellas en que cada estado puede transformarse directamente en otro estado, contándose como un paso en cualquiera de las transformaciones. a-b, b-c, b-a, c-d, etc. En estos casos hablamos de series no aditivas y lógicamente éstas nunca son de naturaleza lineal. Este tipo de codificación se aplica a secuencias de proteínas o de nucleótidos. En ese último caso los cuatro estados los constituyen las bases A, C, G y T, pudiendo cada una de las bases ser reemplazada por el resto de forma equivalente.
- b. **Serie de transformación multiestado ordenadas.** También llamadas aditivas, y que a su vez pueden ser lineales o no lineales

### b.1. Serie lineal no polarizada: $0 \rightarrow 1 \rightarrow 2$

Una serie de transformación (carácter que presenta estados múltiples) es lineal cuando puede pasarse sucesivamente de un estado a otro. Cada transformación se contabiliza como un paso.

### b.2.- Serie lineal polarizada: $0 \rightarrow 1 \rightarrow 2$

Serie en que queda establecida la condición relativa de apomorfía y plesiomorfía. 1 es apomorfo respecto a 0 pero plesiomorfo respecto a 2. Debe quedar clara la diferencia entre carácter ordenado (concerniente al orden de la transformación) y carácter polarizado (concerniente al sentido de la transformación), de forma que la ordenación de los estados de una serie es independiente de su dirección, si bien ambos conceptos están relacionados, pues la ordenación de una serie de transformación restringe su polarización o secuenciación.

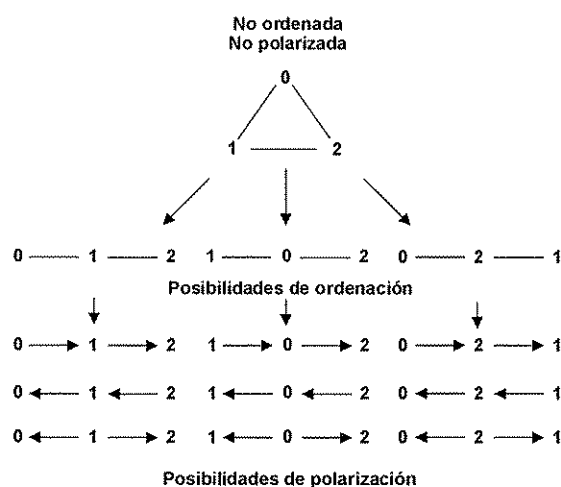


Fig. 5.- Relaciones entre ordenación y polaridad. Serie no ordenada y no polarizada, con las posibilidades de ordenación (tres) y de polarización (nueve). Tomado de Ferrarezzi y Marques, 1997.

### c. Series ramificadas.

La serie ramificada manifiesta las relaciones o conexiones entre los estados múltiples bajo la forma de bifurcaciones:



Las ordenaciones de las series de transformación son de gran importancia en el análisis de los caracteres, pues dichas series son tratadas de forma distinta por los diversos algoritmos del análisis de parsimonia, según se consideren ordenadas o desordenadas.

## 5.3. Modelos de parsimonia

Aunque los métodos de parsimonia siempre buscan minimizar el número de pasos en las transformaciones de un carácter, existen distintas modalidades de parsimonia que difieren con relación a las restricciones que se imponen a dichas transformaciones, afectando a la topología final del árbol así como a su longitud total.

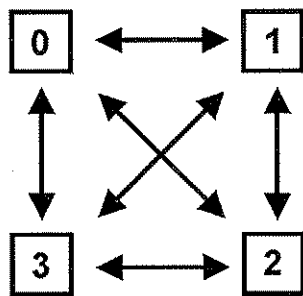


- **Parsimonia de Wagner** (Wagner, 1961; Kluge y Farris, 1969; Farris 1970).



Este modelo permite *a priori* las convergencias y las reversiones ( $0 \rightarrow 1$ ,  $1 \rightarrow 0$ ). Este criterio es apropiado para el tratamiento de caracteres multiestado ordenados codificados aditivamente, es decir, se contabiliza un paso para cada transformación adyacente, pero en el caso de transformaciones no adyacentes, el número de pasos se corresponde con la suma de las transformaciones implicadas, p.e.  $0 \rightarrow 2 = 2$  pasos.

- **Parsimonia de Fitch** (Fitch, 1971).



Este criterio es una generalización del método de Wagner para caracteres multiestado desordenados tales como las secuencias de proteínas o de nucleótidos. Siempre son tratados bajo la opción de no ser aditivos, ya que cada estado puede derivarse directamente de cualquier otro y en cualquier secuencia. Se contabiliza como máximo un paso por transformación al aplicarlo a la longitud del árbol.

- **Parsimonia de Camin-Sokal** (1965).

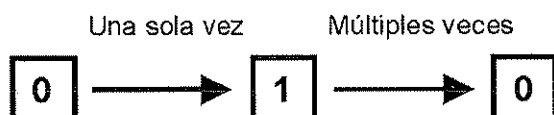
Este modelo sólo permite las convergencias ( $0 \rightarrow 1$ ), las reversiones son excluidas y el estado ancestral (0) debe ser conocido de antemano.



Este criterio impone fuertes restricciones a la dirección de las series de transformación al asumir la irreversibilidad de los estados de los caracteres. Este criterio apenas se utiliza en los análisis cladistas, siendo inadecuado tanto para los datos morfológicos como para los moleculares.

- **Parsimonia de Dollo** (Le Quesne, 1972; Farris, 1977).

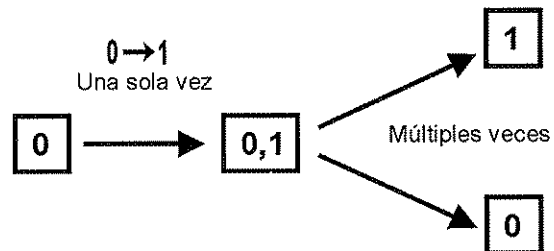
Este modelo sólo acepta las reversiones y excluye las convergencias:



Este criterio surge de la aplicación del "concepto de carácter único" de Le Quesne (1972). Según este autor, es más fácil perder un carácter (retorno a su estado inicial) que adquirirlo en paralelo (convergencia). La expresión "parsimonia de Dollo" resulta engañosa, pues la "ley de Dollo", del paleontólogo belga Louis Dollo, expresa lo contrario, que el retorno al estado ancestral es imposible. Este criterio se aplica a determinados caracteres en que la probabilidad de que acontezca una pérdida es mucho mayor que la de orígenes independientes, caso de los mapeos de los sitios de restricción por endonucleasas del ADN mitocondrial. Puede aplicarse también a caracteres, tanto ordenados como desordenados, siempre que cada estado del carácter evolucione sólo una vez.

- **Método de parsimonia de polimorfismos** (Farris, 1978; Felsenstein, 1979).

Un carácter polimorfo es aquel que está presente en su forma apomórfica y plesiomórfica.



El método de parsimonia de polimorfismos admite que el estado plesiomórfico evolucione sólo una vez, originando un polimorfismo para ambos estados ( $0 \rightarrow 0,1$ ), y permite las pérdidas  $0,1 \rightarrow 0$ ,  $0,1 \rightarrow 1$ , en cada estado del polimorfismo, es decir, su posible fijación en el estado plesiomorfo o apomórfico. Este criterio es de utilidad para caracteres tales como reorganizaciones cromosómicas, polimorfismos enzimáticos o sitios de restricción de endonucleasas.

#### 5.4. Búsqueda del árbol más parsimonioso

En los apartados precedentes se ha visto que existen distintos tipos de métodos que pueden ser elegidos para realizar un análisis numérico de los datos, tales como ordenar o no ordenar los estados de los caracteres, escoger un determinado criterio de parsimonia, o también la posibilidad de asignarles un determinado peso. Todos estos criterios influyen en la búsqueda del árbol más parsimonioso (AMP), "*most parsimonious tree*", es decir, aquel o aquellos árboles que minimizan el número de pasos requeridos para explicar una matriz o conjunto de datos. El problema es que con  $n$  taxones terminales se pueden construir  $(2n-3)! / [2^{n-2} (n-2)!]$  árboles dicotómicos (para  $n=10$  existen aproximadamente  $3,5 \times 10^7$  árboles). De ello se concluye que la comparación de tal número de árboles resulta una operación tediosa, sino imposible.

##### ■ Algoritmos exactos y heurísticos

Los criterios de resolución para establecer el AMP a partir de un conjunto de datos han ido modificándose a lo largo de su corta historia. Inicialmente, el procedimiento inspirado por Hennig, consistía en la adición secuencial de carácter por carácter en el contexto de una topología inicial irresoluta para el conjunto total de los taxones a analizar, la cual iba progresivamente resolviéndose con dicha

adición secuencial. Con la aparición de los algoritmos numéricos los análisis se llevan a cabo de forma inversa. En este caso, son los taxones los que se adicionan secuencialmente en el contexto del conjunto de la totalidad de los caracteres. Estos procedimientos se realizan en árboles sin raíz, pues el número de sus topologías es menor que el de sus correspondientes árboles enraizados. El enraizamiento se realiza una vez finalizado este análisis según distintos criterios, p.e. la comparación con un grupo externo.

Los algoritmos numéricos pueden dividirse en algoritmos exactos y en algoritmos heurísticos. El criterio de aplicación de uno u otro dependerá del número de taxones a analizar, del criterio de parsimonia adoptado y del grado de coherencia de los datos.

#### ■ Algoritmos exactos

Los algoritmos exactos son aquellos por los que se garantiza la obtención de todos los AMPs (árboles más parsimoniosos) correspondientes a una matriz dada. Su aplicabilidad está limitada a unos 20 taxones terminales. A partir de ese rango se manifiesta el problema de los denominados "*NP-complete*" (*non deterministic polynomial*). Un algoritmo se dice polinomial si su ejecución demanda un número mínimo de operaciones, limitada por una función polinomial del tamaño de los datos de entrada. En estos casos el problema puede ser resuelto. Por el contrario, para un problema de *NP-complete*, no existe tal algoritmo polinomial y su resolución no es tratable.

Los algoritmos exactos pueden ejecutarse por medio de un análisis exhaustivo o por la técnica de *Branch and bound*. Los métodos exhaustivos evalúan todos los árboles posibles, siempre que el conjunto de especies no sea superior a 10. Este método permite examinar la distribución de frecuencias de las longitudes de los árboles. El método "*branch and bound*" (Hendy y Penny, 1982; Swofford y Olse, 1990) consiste en un algoritmo exacto que se utiliza cuando el número de taxones está comprendido entre 20 y 30. Esta técnica permite una búsqueda exhaustiva, pero referida a un árbol determinado, el cual puede ser elegido al azar, o calculado a partir de un algoritmo heurístico, p.e. el algoritmo de Wagner. A partir de la selección de este árbol se determina el número de pasos, de forma que el árbol "mínimo" no pueda exceder la longitud de este árbol de referencia. Para ello utiliza métodos heurísticos aplicados a las ramas parciales, y así obtiene el límite superior del valor del criterio de optimización.

#### ■ Algoritmos heurísticos

Estos algoritmos se utilizan cuando resulta imposible aplicar los algoritmos exactos, y por tanto, sólo resta la posibilidad de una aproximación al mejor resultado posible. Si bien poseen la ventaja de obtener un resultado en un tiempo de cálculo razonable, sin embargo, no se tiene la certeza de que sean los más parsimoniosos, ni que todos los AMPs hayan sido localizados.

Entre los métodos heurísticos más utilizados está la "adición paso a paso" (*stepwise addition*), cuyo resultado está condicionado por el orden en que se introduzcan los taxones, y los denominados "reordenamiento de ramas" (*branch swapping*), que intenta paliar los efectos del anterior método y mejorar el árbol inicial mediante el desplazamiento de sus ramas. Si una reordenación genera un árbol más corto, éste es el utilizado para una nueva reordenación, de esta forma se puede llegar a alcanzar el árbol más corto.

### 5.5. Polarización por grupo externo. Polarización *a priori*

Un grupo externo es un taxón, o grupo de taxones, filogenéticamente emparentado con el grupo en estudio. La polarización por el método del grupo externo pretende la reconstrucción de un ancestro hipotético, ya que la identificación previa de un "verdadero" ancestro en todo análisis filogenético raramente se encuentra. La polarización por comparación con el grupo externo es simple y clara: dado un grupo monofilético, se infiere que el estado ancestral más parsimonioso es aquel que acontece en el grupo externo, pues cumple la hipótesis de parsimonia de exigir un menor número de pasos para explicar la transformación de los estados de los caracteres en apomorfos y plesiomorfos, entre los componentes del grupo interno. El problema práctico es que entre los caracteres del grupo(s) externo(s) elegido y el grupo interno surgen incoherencias o contradicciones filogenéticas. Para minimizar este problema Maddison et al. (1984) propone analizar los grupos externos de forma separada al del grupo interno, tratando de optimizar los caracteres y asumiendo una relación previa entre los grupos externos. Las etapas de esta aplicación consisten en asumir el monofiletismo del grupo de estudio o interno, definiendo dos nodos: uno correspondiente al grupo interno y que representa el ancestro común más reciente al grupo interno y otro, el llamado nodo del grupo externo, que corresponde al ancestro común relativamente más antiguo común al grupo interno y a uno de los grupos externos más próximo. A partir del establecimiento de estos nodos y del conocimiento de las relaciones existentes entre los grupos externos entre sí y con el grupo interno, se define la topología de los grupos externos. El problema de este método es que si no se resuelven o conocen las relaciones entre los grupos externos y de éstos con el grupo interno, puede ser imposible inferir la polaridad de los caracteres.

Por tanto, la polarización de los caracteres como paso inicial en el análisis cladista, además de ser innecesario, puede conducir a resultados que no sean globalmente parsimoniosos, por pretender validar una parsimonia en dos fases sucesivas y disyuntas (Nixon y Carpenter, 1993; Ferrarezzi y Marques, 1997).

### 5.6. Enraizamientos y grupo(s) externo(s). Polarización *a posteriori*

Una forma más segura y práctica de evitar los problemas anteriormente mencionados, es la inclusión de todos los taxones en el análisis filogenético, dejando que los ancestrales hipotéticos sean dilucidados *a posteriori* como resultado de la aplicación de dicho análisis. Esto se debe a que la longitud de un árbol es independiente de la posición de la raíz o de la elección de cualquier brazo terminal como grupo externo, no siendo necesaria la polarización previa de los caracteres. El enraizamiento y la consecuente polarización de los caracteres se obtendrá en un punto intermedio entre el, o los, grupo(s) externo(s) y el grupo interno.

Lundberg (1972) propone otro método de enraizamiento, resolviendo en primer lugar la topología del grupo interno a través de un análisis de Parsimonia no enraizada y posteriormente valida cual es la posición en la que un ancestro hipotético puede ser insertado de forma más parsimoniosa al árbol obtenido. Se aconseja la aplicación de este método en aquellos casos en que no es posible obtener grupos externos adecuados (Nixon y Carpenter, 1993).

### 5.7. Medidas de la homoplasia

Al comparar distintos árboles, éstos pueden diferir en su diseño. La decisión para establecer cual de esos árboles es el más representativo, el por qué lo es y en que medida, puede ser argumentada a partir de la medida de la cantidad de información filogenética contenida en los árboles y obtenida mediante el estudio de una serie de índices que muestren la carga de homoplasia.

- **Índice de consistencia IC** (*consistency index*, Kluge y Farris, 1969).

Este índice es una estima de la cantidad de sinapomorfias y por tanto del grado de homoplasia. El índice de consistencia de un árbol es igual a la relación entre el número mínimo (R) de transformaciones necesarias para los estados de todos los caracteres y el número efectivo de transformaciones (L) en el árbol considerado. En el árbol más parsimonioso L representa la longitud mínima del árbol.  $IC = R / L$

Valores igual a 1 indican ausencia de homoplasia, la diferencia L-R representa el número de homoplasias y el inverso del IC, refleja el número medio de transformaciones por carácter, si estos son binarios.

Este índice presenta algunos inconvenientes. Uno de ellos, es que tiende a 1 al aumentar el número de autapomorfias, el otro es que su valor mínimo no es igual a 0, sino a la relación entre R y el valor máximo de L, ajustándose en torno al valor de 0,2 al distribuirse los estados de los caracteres de forma aleatoria sobre las ramas terminales. Por tanto, aquellos árboles con un IC entre 0,25 y 0,30, poseen una elevada cantidad de homoplasia y la información filogenética de los datos es de escasa relevancia. Una forma de corregir estos inconvenientes es no considerar las autapomorfias o excluir los caracteres no informativos, con el propósito de reducir el valor del índice y reflejar mejor la verdadera proporción de homoplasia.

- **Exceso relativo de homoplasia** (*Homoplasy excess ratio*, Archie, 1989 a y b).

Este índice intenta determinar el "ruido de fondo", sin significado filogenético, responsable de aportar al IC un valor siempre mayor de 0.

$$HER = (M-L) / (M-R)$$

M se obtiene transformando la matriz de datos, al distribuir de forma aleatoria los estados de los caracteres en los distintos taxones. A continuación se busca la longitud del árbol más parsimonioso a partir de esos datos aleatorizados. Esta operación se repite un gran número de veces con el objetivo de estimar la longitud media, M, que representa el número medio de transformaciones observadas en los árboles de longitud mínima, obtenida ésta por la aleatorización de los datos.

- **Índice de retención IR** (*Retention index*, Farris, 1989).

Una forma de realizar un cálculo aproximado del anterior índice (HER), y evitar el proceso descrito, es reemplazar M por G, el máximo de transformaciones requeridas por los datos para cualquiera de los árboles. El índice que se obtiene es el *Homoplasy excess ratio maximum* (Archie, 1989 a y b):

$$HERM = (G-L) / (G-R)$$

G es siempre mayor que M, de hecho HERM es una sobreestimación del HER y subestima la homoplasia. Este HERM se corresponde con el Índice de Retención IR (*Retention index* de Farris, 1989).

### 5.8. Árbol de consenso

El árbol de consenso trata de resolver el problema de la comparación entre árboles que tienen una misma longitud o igualmente parsimoniosos, tratando de optimizar las diferentes topologías y no de optimizar la matriz de datos. La búsqueda de este árbol, se condensa en una única representación en la que se diferencian claramente las partes concordantes entre los distintos árboles, de las de las partes discordantes.

Los métodos de consenso más generalizados son: el método estricto o de Nelson (Sokal y Rohlf, 1962, 1981; Nelson, 1979), semiestricto (Bremer, 1990), de Adams (Adams, 1972) y de consenso mayoritario (Margush y McMorris, 1981). Estos métodos difieren en el criterio por el que se forman los subgrupos de taxones en el árbol de consenso. En general, el árbol de consenso de un conjunto de árboles es menos resolutivo, de mayor longitud, que los árboles iniciales y menos informativo. Ello se debe al hecho de que los caracteres sobre el árbol de consenso tienden a cambiar más veces que en un árbol cualquiera, en que los cambios se minimizan. Debe destacarse que ningún árbol de consenso es la hipótesis más idónea para el conjunto de los datos, sino que debe interpretarse como el grado de acuerdo o desacuerdo entre varios árboles obtenidos después de una búsqueda y no como un árbol filogenético.

## 6. EL MÉTODO DE COMPATIBILIDAD

El método de compatibilidad es, en cierta medida, una variante de los métodos de parsimonia, pues si bien utiliza el principio de parsimonia, se diferencia de ellos en cómo trata la homoplasia.

La compatibilidad se basa en los trabajos de Le Quesne (1969, 1972), en los que define los caracteres mutuamente compatibles como aquellos cuyos cambios de estado sobre un árbol determinado pueden ser explicados sin recurrir a la homoplasia, es decir, que el árbol sólo presenta una transformación por carácter. Al conjunto de caracteres mutuamente compatibles se les denomina "clique" (Estabrook et al., 1977). Los caracteres que no son compatibles son, por tanto, homoplásicos.

El método de compatibilidad consiste simplemente en buscar el árbol para el cual el "clique" es mayor, despreciando los caracteres que son homoplásicos. El rechazo por este método a usar caracteres homoplásicos en la búsqueda del árbol filogenético, conlleva una serie de problemas. Así, puede ser que existan caracteres sometidos a homoplasia, pero que sin embargo sea diagnóstico de un grupo monofilético situado en el interior del grupo estudiado. Su supresión supondrá la pérdida de una información filogenéticamente útil. Por ejemplo, la homeotermia se considera una sinapomorfia, por una parte en las aves y, por otra, en los mamíferos. Este carácter ha aparecido dos veces por convergencia. Si esta ocurrencia es correcta, el carácter homeotermia no será considerado en un análisis de compatibilidad entre los amniotas, rehusando explicar como ha aparecido este carácter y suprimiendo información filogenética útil.

## BIBLIOGRAFÍA

- ADAMS, E. N., 1972. Consensus techniques and the comparison of taxonomic trees. *Syst. Zool.*, **37**: 27-29.
- ARCHIE, J. M., 1989 a. A randomization test for phylogenetic information in systematic data. *Syst. Zool.*, **38**(3): 239-252.
- ARCHIE, J. M., 1989 b. Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematic and a critique of the consistency index. *Syst. Zool.*, **38**(3): 243-269.
- BREMER, K., 1990. Combinable component consensus. *Cladistics*, **6**: 369-372.
- BRUSCA, R. C. & BRUSCA, G. J., 1990. *Invertebrates*. Sinauer Associates, Inc. Sunderland, Massachusetts, USA. 922 pp.
- CAMIN, J. H. & SOKAL, R. R., 1965. A method for deducing branching sequences in phylogeny. *Evolution*, **19**: 311-326.
- CAVALLI-SFORZA, L. L. & EDWARDS, A. W., 1967. Phylogenetics analysis: models and estimation procedures. *Am. J. Hum. Genet.*, **19**: 233-257.
- CHACE, F. C., 1969. Unknown species in the sea. *Science*, **163**: 1271.
- DARLU, P. & TASSY, P., 1993. *La reconstruction phylogénétique. Concepts et méthodes*. Masson (Paris). 245 pp.
- EDWARDS, A. W. & CAVALLI-SFORZA, L. L., 1963. The reconstruction of evolution. *Ann. Hum. Genet.*, **27**: 104-105.
- EDWARDS, A. W. & CAVALLI-SFORZA, L. L., 1964. *Reconstruction of evolutionary tree*. Systematics Association Publication. 6: 67-76.
- ELDRIDGE, N. & CRACRAFT, J., 1980. *Phylogenetic patterns and the evolutionary process*. Columbia University Press. New York.
- ESTABROOK, G. F., STRAUCH, J. G. & FIALA, J. K., 1977. An application of compatibility analysis to the Blackiths' data on orthopteroid insects. *Syst. Zool.*, **26**: 269-276.
- FARRIS, J. S., 1970. Methods for computing Wagner trees. *Syst. Zool.*, **19**: 83-92.
- FARRIS, J. S., 1977. Phylogenetics analysis under Dollo's law. *Syst. Zool.*, **26**: 78-88.
- FARRIS, J. S., 1978. Inferring phylogenetics trees from chromosome inversion data. *Syst. Zool.*, **27**: 275-284.
- FARRIS, J. S., 1989. The retention index and homoplasy excess. *Syst. Zool.*, **38**(4): 406-407.
- FELSENSTEIN, J., 1979. Alternative methods of phylogenetic inference and their interrelationship. *Syst. Zool.*, **28**: 49-62.
- FERRAREZZI, H. & MARQUES, A. C., 1997. Análise cladística numérica e recursos computacionais. 164-186. En: A MORIN, D. de S. (Ed.) *Elementos Básicos de Sistemática Filogenética* (2ª ed.). Holos Editora y Sociedade Brasileira Entomologia, Ribeirão Preto y São Paulo. Brasil. 276 pp.
- FITCH, W. M., 1971. Distinguishing homólogos from analogous proteins. *Syst. Zool.*, **20**: 406-416.
- FITCH, W. M. & MARGOLISH, E., 1967. Construction of phylogenetic trees. *Science*, **155**: 279-284.
- HENDY, M. D. & PENNY, D., 1982. Branch and bound algorithms to determine minimal evolutionary trees. *Math. Biosci.*, **59**: 277-290.
- HENNIG, W., 1950. *Grundzüge einer Theorie der Phylogenetischen Systematik*. Deutscher Zentralverlag, Berlin. 370 pp.
- HENNIG, W., 1966. *Phylogenetic Systematics*. University of Illinois Press. 263 pp.
- HENNIG, W., 1968. *Elementos de una sistemática filogenética*. Editorial Universitaria de Buenos Aires. Buenos Aires, Argentina. 353 pp.
- HUXLEY, J., (ed.), 1940. *The New Systematics*. Clarendon Press., Oxford, 583 pp.
- HUXLEY, J., 1942. *Evolution, the modern synthesis*. Allen and Unwin, London, 645 pp.
- KLUGE, A. G. & FARRIS, J. S., 1969. Quantitative phyletics and evolution of anurans. *Syst. Zool.*, **18**(1): 1-32.
- LE QUESNE, W. J., 1969. A method of selection of characters in numerical taxonomy. *Syst. Zool.*, **18**(2): 201-205.
- LE QUESNE, W. J., 1972. Further studies based on uniquely derived character concept. *Syst. Zool.*, **21**(3): 281-288.
- LEWIN, R., 1987. When does homology mean something else? *Science*, **237**: 1570.
- LUNDBERG, J. G., 1972. Wagner networks and ancestors. *Syst. Zool.*, **21**: 398-413.
- MADDISON, W. P., DONOGHUE, M. H. & MADDISON, D. R., 1984. Outgroup analysis and parsimony. *Syst. Zool.*, **33**: 83-103.
- MARGUSH, T. & MCMORRIS, F. R., 1981. Consensus n-trees. *Bull. Math. Biol.*, **43**: 239-244.
- MAYR, E., 1969. *Principles of Systematic Zoology*. McGraw-Hill, New York, 428 pp.
- MAYR, E., 1982. *The Growth of Biological Thought*. Harvard Univ. Press., Cambridge, Mass. 974 pp.
- MAYR, E., LINSLEY, E. G. & USINGER, R. L., 1953. *Methods and Principles of Systematics Zoology*. McGraw-Hill, New York, 328 pp.
- NELSON, G., 1979. Cladistic analysis and synthesis: Principles and definitions, with a historical note on Adamson's *Familles des Plantes* (1763-1764). *Syst. Zool.*, **28**: 1-29.
- NELSON, G. & PLATNICK, N., 1981. *Systematic and biogeography: cladistics and vicariance*. Columbia University Press, New York. 567 pp.
- NIXON, K. C. & CARPENTER, J. M., 1993. On outgroups. *Cladistics*, **9**: 413-426.
- PATTERSON, C., 1982. Morphological characters and homology. 21-74. In: JOYSEY, K. A. & A. F. FRIDAY (Eds). *Problems of phylogenetic reconstruction*. Academic Press. London. 442 pp.
- SAVORY, T., 1970. *Animal Taxonomy*. Heinemann Educational, London, 101 pp.
- SIMPSON, G. G., 1961. *Principles of Animal Taxonomy*. Columbia Univ. Press., New York, 247 pp.
- SNEATH, P. H. A. & SOKAL, R. R., 1973. *Numerical Taxonomy. The principles and practice of numerical classification*. Freeman, San Francisco, California. 573 pp.
- SOKAL, R. R. & ROHLF, F. J., 1962. The comparison of dendograms by objective methods. *Taxon*, **11**: 33-40.
- SOKAL, R. R. & ROHLF, F. J., 1981. Taxonomic congruence in the Lepopodomorpha reexamined. *Syst. Zool.*, **30**: 309-325.
- SOKAL, R. R. & SNEATH, P. H. A., 1963. *Principles of numerical taxonomy*. W. H. Freeman and Co. San Francisco. 359 pp.
- STRICKLAND, H. E., 1840. Observations upon the affinities and analogies of organized beings. *Ann. Mag. Nat. Hist.*, **4**: 219-226.
- SWOFFORD, D. L. & OLSE, G. J., 1990. Phylogeny reconstruction. In: HILLIS D. M. & MORITZ, C. (Eds). *Molecular Systematic*. Sinauer Ass. (Sunderland, Massachusetts): 411-501.
- WAGNER, W. J. Jr., 1961. Problems in the classification of ferns. In: *Recent advances in Botany*. University of Toronto Press., Montreal: 841-844.
- WILEY, E. O., 1981. *Phylogenetics. Theory and practice of phylogenetic systematic*. John Wiley and sons. New York. 439 pp.