EL ANÁLISIS FILOGENÉTICO: MÉTODOS, PROBLEMAS Y PERSPECTIVAS

LUIS E. EGUIARTE¹, VALERIA SOUZA¹, JUAN NUÑEZ-FARFÁN¹
Y BLANCA HERNÁNDEZ-BAÑOS²

¹Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70-275 CU, Coyoacán 04510, D.F., México.

²Museo de Zoología, Facultad de Ciencias, Universidad Nacional Autónoma de México, Apartado Postal 70-399 CU, Coyoacán 04510, D.F. México.

Resumen. En el siglo pasado, Darwin y Haeckel fueron los primeros en proponer árboles filogenéticos. Las primeras filogenias fueron construidas intuitivamente, pero a partir de los años cincuenta diferentes investigadores intentaron dar reglas para su reconstrucción. La escuela evolucionista propuso utilizar los principios de la evolución, la escuela de la taxonomía numérica (o fenética) dió origen a los "métodos de distancia", mientras la escuela cladista propone el "método de parsimonia". Los métodos de distancia más utilizados son el UPGMA y el Neighbor-Joining. Los métodos de parsimonia tratan de encontrar el árbol que requiere el menor número de cambios. Existen otros métodos que toman en cuenta a las fuerzas evolutivas, como el de máxima verosimilitud (Felsenstein, 1981). En este artículo discutimos los problemas asociados a la reconstrucción filogenética, las ventajas de los diferentes métodos, y de algunos de los programas de computadora para la reconstrucción filogenética; presentamos asimismo una perspectiva de los usos de las filogenias.

Palabras clave: Evolución, filogenia, reconstrucción filogenética

Abstract. In the last century, Darwin and Haeckel were the first to propose the phylogenetic trees. The first phylogenies were made intuitively, but in the fifties several researchers tried to propose rules for their reconstruction. The evolutionary school intended to use the principles of evolution, the numerical (or pheneticist) school advanced the "distance methods", while the cladistic school proposed the "parsimony method". The UPGMA and the Neighbor-Joining are the more commonly used distance methods. The parsimony method tries to find the tree that requires the smallest number of changes. There are other methods that take into account the evolutionary forces, like the maximum likelihood (Felsenstein, 1981). In this paper we discuss some of the problems associated to phylogeny reconstruction, the advantages of the different methods and of some of the computer available programs, and finally we offer a perspective on the uses of the phylogenies.

Key words: Evolution, phylogeny, phylogenetic reconstruction

El Santo Grial de muchos biólogos evolutivos desde tiempos de Darwin y Haeckel ha sido la obtención de las filogenias. La reconstrucción filogenética se podría considerar como un objetivo central en la biología, ya que la filogenia le da sentido y perspectiva a la información que generan las diferentes ramas de la biología. Sin el marco de referencia de la filogenia, la biología se vuelve una larguísima colección de anécdotas más o menos inconexas. Adicionalmente, la filogenia ocupa un lugar central en la sistemática y la taxonomía, al ayudar a resolver problemas de posición y nomenclatura (Forey, 1992). Por otra parte, las filogenias son una herramienta básica en estudios ecológicos y evolutivos, al utilizar por ejem-

plo, el "método comparativo" (Harvey y Pagel, 1991; Harvey *et al.*, 1996). La reconstrucción filogenética también es relevante al servir como un puente de unión entre la micro y la macroevolución (Avice, 1994).

Recientemente se ha renovado el interés en la reconstrucción filogenética. Esto se debe, en parte, al desarrollo de metodologías explícitas y formales para la reconstrucción de las filogenias, y a su disponibilidad en programas de computadoras "amistosos", que pueden ser usados por cualquier científico. Por otra parte, el interés actual en la filogenia se debe al desarrollo impresionante de nuevos métodos moleculares que permiten, en un tiempo relativamente corto, obtener gran cantidad de datos para muchos taxa.

Sin embargo, el problema de la reconstrucción filogenética no es sencillo, y se encuentra rodeado de un aura de misterio, generado en parte por una serie de amargas e inútiles discusiones que rodearon al campo desde los años sesenta y hasta muy recientemente (para una revisión ver Hull, 1988). En este artículo queremos orientar a las personas interesadas en el amplio campo de la reconstrucción filogenética, señalando algunos puntos sobre porqué es tan difícil reconstruir la filogenia real, y cuáles métodos se han propuesto desde una perspectiva evolutiva y práctica.

Los orígenes

La noción básica de la reconstrucción filogenética es la idea intuitiva de todos los naturalistas desde tiempos de Aristóteles y que subyace a todo proceso de clasificación biológica: que el grado de similitud entre dos taxa nos habla de sus relaciones evolutivas (Mayr, 1982; Hull, 1988). Los conceptos de filogenia y árboles filogenéticos datan del siglo pasado. Generalmente se considera que fueron propuestos por Haeckel en 1866 (figura 1), aunque anteriormente algunos autores habían propuesto arreglos jerárquicos (ver Hull 1988). Sin embargo, la estructura jerárquica de la vida fue propuesta originalmente por Darwin (1859) usando un árbol filogenético hipotético (figura 2). El análisis filogenético de los productos de la evolución es de fundamental importancia en la teoría de Darwin, porque constituye un esquema magnificado de los procesos que ocurren dentro de las poblaciones. Darwin presenta al análisis de las relaciones genealógicas en varios contextos. Por una parte, dentro del proceso de la especiación y los mecanismos que la producen (entre los cuales él incluye a la selección natural). Por otra parte, desde una perspectiva evolutiva más amplia, que incluye lo que hoy denominamos macroevolución. Para ello, Darwin desarrolla el *Principio de Divergencia*. Su visión de una jerarquía de relaciones ancestros-descendientes esta presente en varias partes del *Origen* (1859, Cap. IV):

"The affinities of all the beings of the same class have sometimes been represented by a great tree. I believe that this simile largely speaks the truth. The green and budding twigs may represent existing species; and those produced during each former year may represent the long succession of extinct species." ... "As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fill with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications."

Aunque una preocupación fundamental de Darwin era demostrar que la diversidad de formas y especies es explicada mejor por la evolución (descendencia con modificación) que por creación (Hull, 1988), Darwin vislumbró las implicaciones que la filogenia tiene en una posible clasificación natural:

"We can understand, on these views, the very important distinction between real affinities and analogical or adaptive resemblances." ... "On my view of characters being of real importance for classification, only so far as they reveal descent, we can clearly understand why analogical or adaptive character, although of the utmost importance to the welfare of the being, are almost valueless to the systematist." ... "We can also understand the apparent paradox, that the very same characters are analogical when one class or order is compared with another, but give true affinities when the members of the same class or order are compared with another" (Darwin 1859, Cap. XIII).

No obstante, existe discusión sobre si Darwin fue exitoso en convencer a los taxónomos que la clasificación puede ser eminentemente genealógica, ya que permanece, para ellos, el problema del reconocimiento de las analogías y las homologías (rasgos "primitivos" y "derivados") (Hull, 1988). Sin embargo, muchos cladistas consideran que algunos de los principios de Darwin se mantienen vigentes (Hull, 1988).

A partir de Darwin y Haeckel, gran cantidad de evolucionistas propusieron filogenias para muchos grupos biológicos. Sin embargo, estas filogenias fueron construidas sin reglas, ni principios, y básicamente se elaboraron según las ideas intuitivas que estos investigadores tenían sobre su grupo de organismos (Hull, 1988). Por esta razón, a lo largo del siglo XX la reconstrucción de filogenias fue progresivamente cayendo en desprestigio, y las filogenias producidas fueron tomadas solamente como bonitas historias y metáforas para ilustrar libros infantiles y museos, pero no dignas de un interés científico formal.

Sin embargo, partir de los años cincuenta diferentes grupos de investigadores intentaron dar reglas formales que permitieran, por un lado, la reconstrucción filogenética y, por otro, unir la taxonomía/sistemática con la biología evolutiva (Mayr, 1982). Así, surge la escuela evolucionista que propuso utilizar, de manera más o menos intuitiva, los principios de la teoría neodarwinista y la genética de poblaciones (Mayr, 1969, 1981, 1982; Simpson, 1961). Por otra parte, un grupo de investigadores intentaron aplicar diferentes métodos cuantitativos y estadísticos para formalizar la relación de la similitud entre los organismos, escuela a veces llamada de taxonomía numérica o taxonomía fenética (Michener y Sokal, 1957; Sokal y

Figura 1. La primera propuesta de un árbol filogenético, según el diagrama original de Haeckel (1866).

Michener, 1958; Sneath y Sokal, 1973). De este ímpetu inicial surgen los métodos de reconstrucción filogenética que revisaremos posteriormente como “métodos de distancia” (Felsenstein, 1982). Por último está la escuela cladista, surgida de las ideas del entomólogo alemán W. Hennig y sus alumnos en los años cincuenta, (Hennig, 1965), y reforzada por varios taxónomos norteamericanos (Farris, 1982; Hull,

1988). Esta se constituye como una escuela de clasificación biológica con un énfasis en la filogenia.

Es importante mencionar que el análisis filogenético, independientemente de la escuela de clasificación que se siga, debe utilizar caracteres cuantificables, los cuales pueden ser morfológicos (medidas de las estructuras, número de las mismas, formas de un carácter, tipos de compuestos químicos, etc.), o ge-

néticos (tipos y números de cromosomas, similitud de reacciones inmunológicas, hibridación de ADN, presencia o ausencia de isoenzimas o su frecuencia, RFLPs, secuencias de ADN tanto nucleares como de cloroplasto, mitocondria y plásmidos) (Hillis *et al.*, 1996). A partir de estos caracteres, el objetivo de los estudios filogenéticos será reconstruir la historia evolutiva de un grupo, para lo cual deben aceptarse ciertos supuestos. La mayoría de los investigadores interesados en las filogenias asumen explícitamente dos supuestos (Farris 1986):

1] La evolución es "descendencia con modificación"

2] Las similitudes observadas son explicables por herencia y ancestría común.

A partir de aquí, el objetivo fundamental será el encontrar las similitudes de valor sistemático entre taxa, las *homologías*, y rechazar las *analogías*. Cuando una homología entre taxa posee un significado para la reconstrucción filogenética se dice que es un carácter *sinapomórfico*. Los criterios para valorar un rasgo como una homología son: 1] varios caracteres homólogos no concurren en un mismo organismo (*conjunción*); 2] que posea similitud en *topología*, *composición* y *ontogenia*; y finalmente 3] que muestre *congruencia* con otras homologías (Eggleton y Vane-Wright 1994).

Los caracteres homólogos pueden ser de dos tipos (para un grupo particular de organismos): 1] los sinapomórficos (i.e., caracteres derivados compartidos) que indican relaciones filogenéticas y están limitados a un subgrupo de los taxa estudiados; y 2] los simplesiomórficos (i.e., caracteres primitivos compartidos) que se encuentran en todos los miembros del grupo y son irrelevantes al problema de reconstrucción de las relaciones genealógicas dentro del grupo (i.e., entre subgrupos); además, pueden estar presentes en otros organismos fuera del grupo estudiado, por lo que es posible que sean útiles en el establecimiento de relaciones filogenéticas a otro nivel (i.e., géneros, familias, etc., como lo señaló Darwin).

Dependiendo del tipo de carácter empleado en la reconstrucción filogenética es posible obtener grupos *monofiléticos* (definidos por caracteres sinapomórficos), *parafileticos* (presencia de caracteres simplesiomórficos), o bien *polifiléticos* (presencia de analogías). Un *autapomorfismo* es un carácter sinapomórfico, único, que define un subgrupo monofilético dentro de los taxa estudiados (usualmente las "puntas" de los árboles) (Henning 1965).

Evidentemente, existe dificultad en reconocer cuándo un rasgo homólogo es apomórfico o simplesiomórfico, y para ello es necesario establecer cuál

ha sido la dirección del cambio evolutivo. Existen varias formas de conocer la dirección del cambio (*polaridad de un carácter*): 1] A través de la *ontogenia* de los organismos en cuestión (los estados más generales de un rasgo se consideran simplesiomórficos). 2] Comparación con un *grupo externo* ("outgroup"); usualmente se usa un taxón fuera del grupo (del "ingroup"), y los caracteres compartidos entre outgroup e ingroup son plesiomórficos, mientras que los exclusivos al ingroup son apomórficos. Adicionalmente existe el criterio de *comunalidad*, en el cual se asume al carácter más común y más raro como plesiomórfico y apomórfico, respectivamente (Eggleton y Vane-Wright 1994). Generalmente, el método del outgroup es el más usado para conocer la polaridad de un rasgo.

Puede entonces notarse que el análisis de reconstrucción filogenética reside de manera importante en dos principios darwinianos: descendencia con modificación, que hace que los rasgos cambien, diverjan o converjan, en el transcurso de la evolución, y que las similitudes entre caracteres reflejan herencia (i.e., distinción del tipo de homología de que se trata).

En una filogenia podemos distinguir varias partes: Los *nodos terminales* que son los taxa actuales en los que se analizaron los caracteres, y los *nodos internos*, que se podrían interpretar como taxa ancestrales, en los cuales se puede intentar la reconstrucción de sus caracteres. La posición de estos nodos en la filogenia constituye la *topología*. El largo de las ramas que conectan a estos nodos puede representar el número de cambios o mutaciones que sucedieron entre un nodo y otro. Por último, se debe de tener en cuenta la forma final como se grafica el árbol, ya que un mismo análisis se puede representar de manera gráfica de muchas maneras diferentes, y algunas de ellas pueden sugerir cosas distintas o falsas para la gente no entrenada (Swofford *et al.*, 1996). Así, se pueden graficar los mismos árboles filogenéticos con o sin raíz. El largo de las ramas puede ser 1] proporcional al tiempo divergencia, 2] al número de cambios acumulados entre linajes, o 3] arbitrario (que no contenga ninguna información). Por último, se pueden graficar los árboles de diferentes formas alternativas, pero que contienen la misma información ("network", rectangular, angular, circular, etcétera).

Problemas del análisis filogenético

La reconstrucción filogenética de un grupo de organismos puede ser algo tan elusivo y decepcionante como la búsqueda del Santo Grial por los Caballeros de la Mesa Redonda. Entre las razones que contribuyen a que la reconstrucción filogenética sea un problema muy difícil citamos las siguientes:

Figura 2. Diagrama original de Darwin (1859) proponiendo una filogenia hipotética.

1] La extensión del “bosque” filogenético. El principal obstáculo para la obtención de la filogenia verdadera lo constituye el hecho de que existen una gran cantidad filogenias alternativas, pero sólo una es la correcta. Por ejemplo, si se quieren analizar 10 organismos, existen más de 2 millones de árboles sin raíz, y más de 34 millones de árboles con raíz (Felsenstein, 1988; Soltis y Soltis, 1996). Así, es muy probable que la filogenia que obtengamos no sea la filogenia verdadera. Otro problema asociado es que varios métodos de reconstrucción filogenética requieren, en principio, de la evaluación de todos los árboles posibles para elegir al que consideran mejor a partir de ciertos criterios. Si se analizan muchos organismos, se necesitan muchos cálculos y, en consecuencia consume mucho tiempo pudiendo en la práctica ser imposible. Esto ilustra también el porqué no es confiable una filogenia basada en la intuición del taxónomo.

2] La heterogeneidad entre las ramas. Otro obstáculo importante es que diferentes caracteres morfológicos o genéticos pueden cambiar a distintas tasas, dentro y entre linajes evolutivos (Gaut *et al.*, 1992; Eguiarte *et al.*, 1994; Li, 1997). Si existen heterogeneidades en las tasas de cambio, significa que no está operando el reloj molecular, y esto ocasiona que los diferentes métodos de reconstrucción filogenética no sean capaces de reconstruir la filogenia real (Hillis *et al.*, 1994; Eguiarte *et al.*, 1994).

3] La convergencia. Adicionalmente, algunos caracteres pueden estar afectados por selección natural y otros procesos genéticos y evolutivos, por lo que puede llegarse a convergencias (*homoplasias*) que sugerirían similitudes evolutivas, aunque los taxa no estén realmente emparentados filogenéticamente (Felsenstein, 1988).

4] Los sesgos de estimación. Por otra parte, los mismos métodos de reconstrucción filogenética pueden tener distintos sesgos en la reconstrucción, que no necesariamente son evidentes (Nei, 1991; Hillis *et al.*, 1994; Eguarte, 1995). Por ejemplo, el método de UPGMA tiende a unir entre sí a los linajes con tasas de cambio muy rápidas por un lado, y a los linajes con tasas muy lentas por el otro (Nei, 1991). Desafortunadamente, otros métodos pueden tener sesgos, pero no son bien conocidos (Nei, 1991; Hillis *et al.*, 1994). Por otra parte, los métodos de reconstrucción filogenética tienen una serie de supuestos sobre la genética y el comportamiento de los caracteres, y sobre cómo operan las diferentes fuerzas evolutivas en ellos (Felsenstein, 1982; 1988; Swofford *et al.*, 1996). Por lo tanto, es posible que algunos métodos y sus supuestos no sean los adecuados para los taxa y caracteres que se quiere analizar.

5] Los nodos y puntas perdidos. Otro posible problema al reconstruir la filogenia es que no siempre se tienen los datos de todos los miembros del grupo bajo análisis (*i.e.*, especies o poblaciones). Por lo tanto, es posible que la topología del árbol filogenético sea distinta si se incluye a todos los miembros del grupo). Las especies extintas podrían también ser centrales para entender la evolución del grupo, pero usualmente es imposible o muy difícil incorporarlas al análisis (Hillis *et al.*, 1996).

De manera análoga al problema de la no inclusión de todos los miembros de un grupo, existe la posibilidad de reconstruir una filogenia no incluyendo caracteres de importancia genealógica. Por ejemplo, pueden existir polimorfismos difíciles de detectar en las poblaciones por su baja frecuencia. Este problema se relaciona con el de la distinción entre árboles de genes y árboles de especies (Nei, 1987; Li y Graur, 1991).

6] Flujo génico e hibridación. Un último punto que complica el análisis filogenético es el hecho de que puede existir flujo génico entre las poblaciones o especies que estamos analizando. Este flujo génico puede confundir el patrón evolutivo, ya que no existen buenos métodos para separar en la filogenia la herencia vertical de la transferencia horizontal (Sober, 1988; Slatkin, 1994; Eguarte y Souza, 1993)

Bases genéticas, fuerzas evolutivas y elección de caracteres

Para poder llevar a cabo un análisis filogenético adecuado, consideramos que es importante entender cómo se comportan genéticamente los caracteres elegidos, y cómo actúan sobre ellos las fuerzas evo-

lutivas. Por lo tanto la teoría, técnicas y metodología derivadas de la genética de poblaciones deberían poder explicar a las filogenias (Eguarte, 1988; Felsenstein 1988; Hartl y Clark, 1989). Desde el punto de vista de la genética de poblaciones, si los genes son totalmente neutros, la filogenia sería simplemente el resultado entre la deriva génica y mutación, que incrementan la diferenciación genética entre las poblaciones, y el flujo génico, que disminuye el aislamiento entre las poblaciones (Kimura, 1983; Hartl y Clark, 1989).

Otras fuerzas evolutivas, como la selección natural, la endogamia, la recombinación (inter e intragélica), y conversión génica/evolución concertada, etc., pueden ser importantes a diferentes niveles del análisis (Hart y Clark, 1989, Li y Graur, 1991, Li, 1997). Esto debe ser tomado en cuenta si se quiere llevar a cabo un análisis evolutivo adecuado.

La elección de los caracteres es crítica. Los caracteres útiles para la reconstrucción filogenética serán los caracteres homólogos. Se necesita tener caracteres homólogos lo suficientemente variables: que no sean demasiado variables, ya que rápidamente se pierde la "señal" evolutiva y no nos permiten distinguir las afinidades evolutivas, ni que sean muy poco variables, ya que si ese es el caso, casi todos los organismos van a ser iguales, y por lo tanto los caracteres no nos dicen nada (Hartl y Clark, 1989, Li y Graur, 1991). Los mejores caracteres para la reconstrucción filogenética serán aquellos poco afectados por el ambiente (*i.e.* son muy heredables), y por la selección natural.

Métodos de reconstrucción filogenética de distancia

La idea fundamental de los métodos de distancia es que los organismos más parecidos deben de ser los filogenéticamente más emparentados. El problema es que el parecido nos puede engañar. La solución que se ha propuesto es utilizar una gran cantidad de caracteres: algunos serán influenciados por la selección, pero la mayoría reflejarán los patrones evolutivos (Sneath y Sokal, 1973; Hull, 1988; Sokal, 1985). Un posible problema es que muchos caracteres morfológicos diferentes pueden estar codificados por un solo gen.

En esencia, los métodos de distancia calculan un índice de distancia entre pares de taxa, tomando en cuenta cuáles son los caracteres que comparten y, a partir de ellos, siguiendo una serie de reglas obtienen un árbol filogenético. En otras palabras, reducen todos los caracteres analizados a un solo número (la distancia "evolutiva") entre un par de taxa.

La elección del método para obtener las distancias evolutivas es importante, ya que los diferentes estima-

dores de distancia se construyen a partir de distintos supuestos del proceso evolutivo. En algunos casos, a cierta escala (ecológica o geológica) o tipo de caracteres (isoenzimas, secuencias de ADN, etc.), unos estimadores de distancia pueden ser adecuados y en otros casos pueden ser realmente muy malos (Nei, 1987; Cavalli-Sforza et al., 1994). A una escala ecológica, las herramientas de la genética de poblaciones describen bien la diferenciación genética. La teoría clásica de la genética de poblaciones (la de Wright, Fisher y Haldane) puede ser usada fácilmente y sin modificaciones con datos de isoenzimas (Hedrick, 1983; Eguiarte, 1990) y, con algunas modificaciones, en el análisis de secuencias de ADN (Nei, 1987). Para medir el grado de diferenciación entre las poblaciones, se usan comúnmente varios índices: 1) para pares de taxa (sitios, poblaciones, especies, etc.) y marcadores con comportamiento codominante, como isoenzimas, se emplean las distancias e identidades genéticas de Nei e índices relacionados (ver Hedrick, 1983; Nei, 1987). 2) Para secuencias de ADN, se pueden usar las K (proporciones corregidas de diferencias en nucleótidos, Nei, 1987, Li y Graur, 1991). Otro índice popular son los estadísticos F , en particular los estadísticos G_{st} / F_{st} (Hedrick, 1983; Eguiarte, 1990). Estos estadísticos pueden ser usados para estimar a Nm (tamaño efectivo por la tasa de migración). A mayor diferenciación genética, menor identidad y mayor F_{st} (Eguiarte, 1990, Eguiarte et al., 1993).

Una vez transformada la matriz de caracteres original a una matriz de distancia entre pares de taxa, se requiere elegir el método de reconstrucción filogenética. Existen gran cantidad de métodos, que tienen diferentes supuestos y diferentes capacidades. Entre los más utilizados tenemos al UPGMA y al Neighbor-joining (Unión de vecinos).

UPGMA (unweighted pair group method with arithmetic mean): Este fue uno de los primeros métodos cuantitativos para la reconstrucción filogenética y fue propuesto originalmente por Sokal y Michener (1958). Este es uno de los pocos métodos que dan directamente una raíz (la mitad de la distancia entre la última rama y el resto del árbol), y las distancias de las ramas son simplemente las medias entre las distancias de cada grupo de organismos. La ventaja más grande que tiene este método es que es rápido, y puede ser bastante eficiente en la reconstrucción de la filogenia, si las tasas de cambio son relativamente constantes (e.g., entre linajes bacterianos dentro de una población, ver Souza et al., 1994). Si existe heterogeneidad en la tasa de cambio evolutivo entre linajes, el método de UPGMA puede conducir a errores muy fuertes en la reconstrucción del árbol filogenético (Nei, 1991; Eguiarte et al., 1994, Eguiarte, 1995).

Neighbor-Joining (Unión de vecinos de Saitou y Nei 1987). Este es uno de los métodos de distancia más extensivamente usados actualmente (Eguiarte, 1995). Es mucho más robusto a los problemas de diferencias en las tasas de sustitución que el UPGMA y otros algoritmos, y es razonablemente rápido en la búsqueda (Hillis et al., 1994; Eguiarte et al., 1994, Eguiarte, 1995). En este método, se buscan secuencialmente los taxa ("vecinos") que minimizan el largo total del árbol. Cada vez que une un par de taxa, se vuelve a ajustar el largo de las ramas del árbol para cada par de nodos, basándose en la divergencia promedio entre los pares de nodos. Por lo tanto no requiere que exista un reloj molecular estricto (Li, 1997).

Existen otros métodos de distancia, pero se pueden considerar variaciones a los anteriores, como el de Fitch-Margoliash (con o sin reloj molecular, ver Nei, 1987; Swofford y Olsen, 1990), el de Evolución Mínima (Cavalli-Sforza y Edwards, 1967) y el de distancias transformadas de Farris (1977), entre otros.

Adicionalmente, es posible estimar la confiabilidad de los patrones de ramificación o sus intervalos de confianza con diferentes métodos, siendo los más empleados los de pseudomuestreo, como son el "bootstrap" o el "jackknife" (Felsenstein, 1988, Eguiarte et al., 1994; Swofford et al., 1996; Li, 1997).

Las principales ventajas de los métodos de reconstrucción de distancia es que son sencillos y eficientes, por lo que corren muy rápido aún cuando se analiza un gran número de taxa, aún sin equipo de cómputo muy sofisticado (Eguiarte et al., 1994, Eguiarte, 1995). Por otra parte, se ha demostrado que si se emplea el análisis de distancia adecuado, el método de distancia puede ser tan bueno o mejor que los de parsimonia para encontrar el árbol filogenético verdadero (Nei, 1991; Hillis et al., 1994).

Sin embargo, los métodos de distancia han sido criticados porque se consideran poco realistas y porque no representan bien al proceso evolutivo. Por ejemplo, varios caracteres fuertemente correlacionados no evolucionan de manera independiente, por lo que no se obtendría una filogenia real (Felsenstein, 1988). Por otra parte, si las tasas de cambio son muy heterogéneas entre linajes, es posible que una especie se parezca mucho a su ancestro en varias características, sin que realmente se demuestre que esta especie sea el pariente vivo más cercano a la especie ancestral. Además, es común que si se tienen varias especies muy diferentes o que presentan altas tasas de cambio, los algoritmos de reconstrucción por distancia las presenten como relacionadas, aunque en la realidad podrían ser de las más alejadas en la filogenia (Nei, 1991; Eguiarte et al., 1994). Por último, como ya señalamos, estos métodos usan de manera poco eficiente el to-

tal de información disponible, al transformar todos los caracteres a un solo número.

Métodos de reconstrucción filogenética cladistas y de parsimonia

La idea fundamental de los métodos cladistas es que se tiene que analizar a todos los caracteres de manera individual, en vez de reducir toda la información a un índice de distancia genética, como en los métodos anteriores. Las ideas cladistas originales de Henning (1965) sugerían que a partir de un análisis cuidadoso de los caracteres, se debían encontrar ciertos caracteres idóneos que permitirían reconstruir de manera clara la historia evolutiva de un grupo. Estos caracteres serían caracteres no sólo homólogos, sino “derivados compartidos” (sinapomórficos), mismos que definirían las relaciones filogenéticas (Swofford y Olsen, 1990; Harvey y Pagel, 1991). Los caracteres “derivados compartidos” son los homólogos que sólo se encuentran en dos grupos hermanos. En otras palabras, Henning consideraba que si se analizaban con cuidado los caracteres, se encontrarían aquellos caracteres sinapomórficos que definen de manera clara cada uno de los eventos evolutivos de ramificación, sin problemas de convergencia (Henning, 1965; Felsenstein, 1982). En la práctica, encontrar estos caracteres idóneos es muy difícil, ya que generalmente algunos caracteres apoyan un árbol filogenético, y algunos a otro. Usualmente no existe una solución sencilla para elegir cuáles caracteres apoyan al árbol real, y cuáles, por convergencia evolutiva, generan árboles espurios. En particular, para los datos de secuencia de ADN es muy común que una serie de caracteres nos sugieran una filogenia y otros caracteres otra, debido a que para cada carácter sólo existen cuatro estados posibles (GATC), por lo que las mutaciones que generan convergencia son muy comunes. Así pues, tenemos el llamado “dilema de Henning” (Harvey y Pagel, 1991): ¿Cómo decidir cuáles caracteres usar (y cuáles desechar)? o en otras palabras, ¿cuál de todas las filogenias es la verdadera?

Para resolver el dilema de Henning, siguiendo las ideas cladistas, existen dos posibilidades. Por un lado, están los *Métodos de Compatibilidad (Clique)*, que tratan de identificar el árbol que es apoyado por el mayor número de caracteres (Le Quesne, 1969; Felsenstein, 1982, 1988). Esta metodología suena lógica, pero es poco utilizada en la literatura (para un ejemplo desarrollado, ver Li y Graur, 1991, pp. 124-125), ya que muchas veces produce resultados cuestionables (L.E. Eguarte, *obs. pers.*). Adicionalmente, los algoritmos que utilizan los métodos de compatibilidad son muy lentos, ya que tienen que evaluar todos los árboles

posibles, y cuál árbol apoya cada uno de los caracteres.

La segunda opción metodológica para resolver el dilema de Henning es el *Análisis de Parsimonia*. Los análisis de parsimonia requieren que se tengan caracteres no sólo polimórficos, sino informativos: caracteres que favorezcan un árbol sobre los otros. En principio, se requiere analizar a todos los árboles posibles con los taxa, y evaluar cuántos cambios se requieren en cada uno “reconstruyendo” ancestros en los nodos intermedios (Maddison y Maddison, 1992). El árbol que mejor representa a la filogenia verdadera será aquel que requiera el menor número de cambios (o mutaciones). Así, en un análisis de parsimonia, el criterio para decidir cuál es la mejor reconstrucción de la filogenia verdadera, se reduce a considerar la similitud entre los taxa, analizando cada uno de los caracteres, siguiendo reglas lógicas bien definidas.

Para redondear las ideas expuestas anteriormente, podemos revisar los pasos que se requieren para llevar a cabo un análisis de reconstrucción filogenética de parsimonia: En primer lugar se requiere obtener, para una serie de taxa (poblaciones, especies, géneros, etc.), varios caracteres que parezcan ser buenos, o sea que consideramos que están siendo poco afectados por la selección o el ambiente, y que se puedan codificar en categorías claras, con poca o ninguna sobreposición; y en los cuales se puedan definir los estados de carácter. En términos generales, como regla de dedo, podemos señalar que para que se pueda hacer un buen análisis de parsimonia se *necesita analizar más caracteres que taxa*. En segundo lugar, se analizan todos los árboles filogenéticos posibles para estos taxa y, para cada uno de los árboles, se cuenta el número de cambios necesarios para que a partir de todos los caracteres presentes en un taxón, se pueda obtener los caracteres presentes en todos los otros taxa. El árbol que requiera menos cambios (o pasos), será el que se considere que representa mejor la filogenia verdadera. Debido a la gran cantidad de árboles posibles, este es un proceso muy tardado. Para encontrar el árbol más corto, se han propuesto una serie de algoritmos de búsqueda, llamados “métodos heurísticos”. Estos pretenden encontrar, a partir de los cambios sobre una filogenia preliminar, el árbol más corto sin tener que analizar todos los árboles posibles (ver Swofford *et al.*, 1996). El problema de estos métodos heurísticos es que no siempre encuentran el árbol más corto, especialmente si se analizan muchos taxa. Este fenómeno se conoce como las “islas de parsimonia” (Maddison, 1991; Chase *et al.*, 1994; Soltis y Soltis, 1996). Para tratar de encontrar el árbol más corto, se sugiere repetir búsquedas heurísticas varias veces, a partir de árboles generados al azar.

Existen varias decisiones que se deben realizar en un análisis de parsimonia. En primer lugar, se debe decidir si los estados de los caracteres están ordenados o desordenados. Son estados de carácter ordenados (*e.g.*, 1,2,3,...,n) cuando se tiene que pasar por estados intermedios para llegar a otro, en una serie (por ejemplo, cambio de un estado 1 al 2 para llegar al 3; involucrando dos cambios). Mientras que se consideran caracteres desordenados, si se asume que éstos pueden cambiar libremente a cualquiera de los diferentes estados de carácter (por ejemplo, del estado 1 al 3, sin pasar por el 2; todos los cambios implican un igual número de pasos) (Kitching, 1992; Maddison y Maddison, 1992).

Una segunda decisión consiste en dar peso diferencial a ciertos cambios o a ciertas mutaciones (para tomar en cuenta, por ejemplo diferente posibilidad de cambio entre transiciones y trasversiones en el ADN). En un análisis de parsimonia también se puede decidir el método de reconstrucción de los nodos intermedios (caracteres de los ancestros; Maddison y Maddison, 1992).

Existen diferentes versiones de algoritmos de parsimonia. Estas versiones se reducen a 1] permitir que los caracteres puedan cambiar libremente de estado (parsimonia de Wagner); o 2] considerar que es muy difícil que un carácter que ha cambiado regrese al estado original (parsimonia de Dollo), o 3] que esto es simplemente imposible (Camin-Sokal) (Swofford *et al.*, 1996).

Un problema interesante y potencialmente grave de los análisis de parsimonia es que rara vez se obtiene un solo árbol más parsimonioso. Esto es común si no se tienen suficientes caracteres. El problema se puede resolver tratando de incluir un mayor número de caracteres informativos, pero esto no es posible siempre. Para tratar de resolver este problema, se han propuesto métodos para darles un peso diferencial a ciertos cambios (mayor peso a los más informativos, menor peso a los que presentan tasas de cambio demasiado altas o donde se sospecha demasiada convergencia). De esta manera se puede intentar obtener una mejor estimación de la filogenia real (Albert *et al.*, 1993). Alternativamente, se pueden obtener los árboles *consenso*, que son una especie de árbol promedio de todos los árboles más parsimoniosos obtenidos. La obtención del árbol consenso no es trivial y su uso ha sido criticado por la comunidad cladista, en particular porque podrían obtenerse reconstrucciones de ancestros no presentes en ninguno de los árboles más parsimoniosos (Swofford, 1991a). Programas como el PAUP (Swofford, 1991b) ofrecen diversas opciones para la obtención de árboles consenso.

Para analizar qué tan bien funcionan los caracteres en la reconstrucción filogenética, se han propuesto varios índices, como el *Índice de Consistencia* y el *Índice de Retención* (Siebert, 1992; Eggleton y Vane-Wright 1994). Por otra parte, para ver qué tan buena es la reconstrucción, se pueden utilizar métodos de pseudomuestreo, como "jackknife", "bootstrap", ya sea reemplazando los caracteres o los taxa (Felsenstein, 1988, 1988; Swofford, 1996; Li, 1997). Existen otros tipos de análisis, como el llamado "decay analysis". En este análisis se obtienen árboles con 1,2,3,...,n pasos más largos que el árbol más parsimonioso. Se considera que si una ramificación se conserva en árboles varios pasos más largos que el árbol más parsimonioso, esta ramificación está bien apoyada por los datos, y por lo mismo es confiable (Bremer, 1988; Eguiarte *et al.*, 1994).

El método de parsimonia es el favorito de muchos investigadores, ya que presenta varias ventajas, como la de ser un método lógicamente claro, en el cual el investigador puede entender la opciones y los supuestos del análisis. Por otra parte, es muy robusto ya que recupera los árboles verdaderos en muchos casos (Hillis *et al.*, 1994; Swofford *et al.*, 1996; Li, 1997), y en ocasiones funciona igual que los métodos con procesos evolutivos explícitos, como el de máxima verosimilitud (Sober, 1988; Felsenstein, 1988). Por otra parte, el método de parsimonia se considera un método superior a los de distancia, al utilizar la información individual de cada uno de los caracteres y no condensarla en un solo valor de distancia, del cual no puede rescatarse la información original posteriormente.

Sin embargo, los métodos de parsimonia tienen una serie de problemas potenciales. El más grave, es que la evolución no sigue un proceso necesariamente parsimonioso: los cambios en los caracteres de los organismos no siempre siguen una ruta del menor cambio posible. Por ejemplo, en algunos ambientes, algunos caracteres cambian más rápido debido a que son resultado de selección natural u otro proceso evolutivo. No existe ningún principio biológico o evolutivo que garantice que todos los caracteres están evolucionando según un proceso mínimo de cambio, que sería el supuesto explícito de la parsimonia. Los cladistas argumentan que es un problema de inferencia, y que el árbol más parsimonioso representa, en términos generales, al árbol "más probable" (Farris, 1982; Felsenstein, 1982; Sober, 1988).

Otro problema del método de parsimonia, en este caso puramente logístico, es que los algoritmos utilizados son muy lentos y se puede requerir de meses de cómputo si se analizan matrices de datos grandes (Chase *et al.*, 1993; Eguiarte, 1995; Soltis y Soltis, 1996).

Por otra parte, como ya mencionamos, se pueden llegar a obtener muchos (millones o más) árboles igualmente parsimoniosos. Por último, no se puede hacer estadística para comparar árboles, sino sólo contar el número de pasos (Felsenstein, 1988; Swofford *et al.*, 1996; Li, 1997).

Métodos de reconstrucción filogenética a partir de procesos evolutivos explícitos

Desarrollar métodos que consideren de manera explícita y realista a las fuerzas evolutivas y las características genéticas de los caracteres, es una prioridad para poder reconstruir de manera objetiva a la filogenia verdadera (Felsenstein, 1982, 1988; Swofford *et al.*, 1996; Li, 1997). Este es un problema en el que están trabajando de manera activa varios grupos de evolucionistas. El método más conocido actualmente es el de "maximum likelihood" (máxima verosimilitud) de Felsenstein (1981, 1982, 1988) para secuencias de ADN. Este método considera el ajuste entre un modelo del proceso evolutivo (las probabilidades de mutación de una base en otra, la importancia de la deriva génica, etc.), los datos (secuencias de ADN por ejemplo), y cada uno de árboles filogenéticos posibles. Así, se encuentra el mejor árbol filogenético, o sea el que da el valor de *likelihood* más adecuado. Este método es considerado uno de los mejores para la reconstrucción filogenética, ya que es muy resistente a heterogeneidades en las tasas de sustitución, y generalmente es el mejor para encontrar, en la mayoría de los casos, el árbol original (Hillis *et al.*, 1994; Li, 1987). Su problema es que es muy lento, ya que requiere estimar todos los árboles (aunque se pueden usar algoritmos de búsqueda heurísticos), y los cálculos para cada árbol son laboriosos, por lo que en la práctica muy difícil usarlo para más de 10 taxa (ver Duvall *et al.*, 1993 a y b).

Programas de reconstrucción filogenética

Existen una gran cantidad de programas de reconstrucción filogenética disponibles. Una lista, con las direcciones en donde pueden obtenerse, se encuentra en el apéndice de Swofford *et al.*, (1996, pp. 510-514). Brevemente, los más usados son: BYOSIS (Swofford y Selander, 1981), que se utiliza para analizar datos de frecuencias alélicas, como los de genética de poblaciones, y puede producir matrices de distancias y reconstrucción filogenéticas con diferentes métodos; sólo disponible en formato DOS para PC. El programa clásico de reconstrucción filogenética es el PHYLIP de Felsenstein (1991), que incluye gran cantidad de documentación, y opciones que van desde calcular

matrices de distancias con varios métodos, hasta la reconstrucción con métodos de distancias, parsimonia, compatibilidad y máxima verosimilitud. También incluye opciones para obtener análisis de "bootstrap" y para graficar los árboles. El programa está disponible tanto en formato DOS como Macintosh. El PAUP de Swofford (1991), que en sus versiones pasadas es un excelente y claro programa para análisis de parsimonia, con gran cantidad de opciones. Existe una versión rudimentaria para PC, pero las versiones más conocidas son para Macintosh. Se ha anunciado una versión más poderosa, que incluye opciones de distancia y maximum likelihood tanto para PC como para Macintosh. Otro programa es el MacClade, de Maddison y Maddison (1993), que es un bonito programa para computadoras Macintosh, que permite la construcción de matrices y la exploración de diferentes árboles filogenéticos, siguiendo ideas cladistas. Se puede usar muy bien como complemento del PAUP. Otro programa de reconstrucción filogenética que ha adquirido popularidad es el MEGA, del grupo de Nei (Kumar *et al.*, 1993), que incluye varias opciones para análisis molecular, alineación y reconstrucción filogenética; está disponible en formato PC.

Discusión y perspectivas

Con el desarrollo de métodos más económicos y eficientes de secuenciación de ADN, y con la disponibilidad de computadoras más poderosas, rápidas y baratas, junto con mejores algoritmos y programas de computación, es claro que cada vez vamos a obtener más y mejores reconstrucciones filogenéticas. Consideramos que no se debe perder el tiempo en largas y estériles controversias, como la discusión entre cladistas *vs.* feneticistas, que dominó el campo hasta muy recientemente, sino que se necesita reconocer que existen diferentes metodologías complementarias, cada una con ventajas y desventajas. Por otra parte, no debemos olvidar que la evolución molecular es mucho más que sólo la reconstrucción filogenética (ver Hartl y Clark, 1989; Li y Graur, 1991; Selander *et al.*, 1991; Takahata y Clark, 1993; Avise, 1994; Hillis *et al.*, 1996; Harvey *et al.*, 1996; Li, 1997), que las filogenias por sí mismas no pueden representar la totalidad del proceso evolutivo, y que en muchos casos la reconstrucción filogenética no es un fin, sino sólo una herramienta (tal vez muy importante) dentro del análisis evolutivo. Así, nosotros proponemos un enfoque práctico y ecléctico, en el que se utilicen los diferentes métodos y herramientas disponibles, y que lo que debe tenerse claro es que existen principios genéticos y diferentes fuerzas evolutivas actuando sobre el material genético de las poblaciones.

Actualmente es un momento excitante dentro de la biología general, ya que los datos moleculares y las filogenias nos permiten conocer muchos patrones y procesos que no estuvieron disponibles para Darwin ni los biólogos de la síntesis evolutiva. Con los datos moleculares y de filogenia podemos conocer el parentesco entre todos los organismos y, a partir de calibraciones cuidadosas del reloj molecular, conocer cuándo se separaron, aún sin poseer un buen registro fósil. Esto habla de la posibilidad de realizar detallados estudios ecológicos y evolutivos de los organismos, de sus caracteres, y de sus adaptaciones, a partir de comparaciones estadísticas cuidadosas conocidas mejor como el "método comparativo".

Aparentemente, ese Santo Grial biológico, la reconstrucción de la historia genealógica de todas las líneas biológicas, puede ser una meta alcanzable en los próximos años, y posiblemente constituirá realmente una nueva síntesis biológica, y permitirá la realización de las metas trazadas en la revolución Darwiniana (Clegg *et al.*, 1991). Esperamos que los biólogos comiencen a familiarizarse no sólo con los métodos y principios de la reconstrucción filogenética, sino con los resultados y herramientas que ofrece la evolución molecular y la genética de poblaciones. Asimismo, será necesario que los usen, de manera rigurosa, en el análisis de los problemas que están estudiando.

Es probable que las muchas hipótesis filogenéticas producidas hasta ahora sean incorrectas, pero constituyen el punto de partida de nuevas investigaciones. Regresando a las palabras de Darwin:

"Finally, the several classes of facts which have been considered in this chapter seem to me to proclaim so plainly, that the innumerable species, genera, and families of organic beings which this world is peopled have all descended, each within its own class or group, from common parents, and have all been modified in the course of descent, that I should without hesitation adopt this view, even if it were unsupported by other facts or arguments". Ch. Darwin (1859, cap. XIII)

Agradecimientos

Jordan Golubov revisó cuidadosamente una versión previa de este manuscrito. La investigación se ha realizado con apoyo de los proyectos CONACYT (N9312-3675) y PAPIIT de la DGAPA, UNAM (IN205894) a LEE.

Literatura citada

- Albert V.A., Chase M.W. y Mishler B.D. 1993. Character weighting for cladistic analysis of protein-coding DNA sequences. *Annals of the Missouri Botanical Garden* 80:752-766.
- Avice J.C. 1994. *Molecular markers, Natural History and Evolution*. New York. Chapman Hall.
- Bremer K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42:795-803.
- Cavalli-Sforza L.L. y Edwards W.F. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 32:550-570.
- Clegg M.T., Felsenstein J., Fitch W., Goodman M., Hillis D., Riley, M., Ruddle F., Sankoff D., Arzberger P., Courtney M., Harriman P., Lynch C., Plesset J., Weiss M. y Yates, T. 1992. Position paper on molecular evolution. *Molecular Phylogenetics and Evolution* 1:84-85.
- Chase M. W. Soltis D.E., Olmstead R.G., Morgan Les, D.H., Mishler B.D., Duvall M.R., Price R.A., Hills H.G., Qiu Y.L., Kron K.A., Rettig, J.H., Conti E., Palmer J.D., Manhart, J.R., Systma K.J., Michaels H.J., Kress W.J., Karol K.G., Clark W.D., Hedrén M., Gaut B.S., Jansen R.K., Kim K., Wimpee C.F., Smith J.F., Furnier G.F., Strauss S.H., Xiang Q., Plunkett G.M., Soltis P.S., Swensen S.M., Williams S.E., Gadek, P.A., Quinn C.J., Eguiarte L.E., Golenber, E., Learn, Jr. G.H., Graham S., Barrett S.C.H., Dayanandan S. y Albert V.A. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80:528-580.
- Darwin, C. 1859. *On the origin of species by means of natural selection*. John Murray, London, first edition.
- Duvall, M. R., Clegg M.T., Chase M.W., Clark, W.D., Kress, J.W., Hills H.G., Eguiarte L.E., Smith F.J., Gaut B.S., Zimmer E.A. y Learn Jr. G.H. 1993. Phylogenetic hypothesis for the Monocotyledons constructed from *rbcL* sequence data. *Annals of the Missouri Botanical Garden* 80:607-619.
- Duvall M.R., Learn Jr. G.H., Eguiarte L.E. y Clegg M.T. 1993. Phylogenetic analysis of *rbcL* sequences identifies *Acorus calamus* as the primal extant monocotyledon. *Proceedings of the National Academy of Sciences USA* 90:4641-4644.
- Eggleton P. y R.I. Vane-Wright. 1994. Some principles of phylogenetics and their implications for comparative biology. En: Eggleton P., y Vane-Wright R.I. Edrs. *Phylogenetics and ecology*. Academic Press, London. 345-366.
- Eguiarte L.E. 1988. Una guía para principiantes en la genética de poblaciones. En: Olea-Franco A. Edr. *Polémicas Contemporáneas en Evolución*, AGT editores, México, D.F. 41-56.
- Eguiarte L.E. 1990. *Genética de poblaciones de Astrocaryum mexicanum Liebm.* en Los Tuxtlas, Veracruz. Tesis de doctorado, Centro de Ecología, UNAM.
- Eguiarte L.E. 1995. Hutchinson (Agavales) vs. Huber y Dahlgren (Asparagales): análisis moleculares sobre la filogenia y evolución de las familias Agavaceae *sensu* Hutchinson dentro de las monocotiledóneas. *Boletín de la Sociedad Botánica de México* 56:45-56.
- Eguiarte L.E., Búrquez A., Rodríguez J., Martínez-Ramos M., Sarukhán J., y Piñero D. 1993. Direct and indirect esti-

- mates of neighborhood and effective population size in a tropical palm, *Astrocaryum mexicanum*. *Evolution* **47**:75-87
- Eguiarte L.E. y Souza V. 1993. Evolución experimental en bacterias: Diez mil generaciones del experimento de Irvine, California. En: Nuñez-Farfán J, Cordero C. Edrs. *Tópicos de Biología Evolutiva I: Diversidad y Adaptación*. Centro de Ecología, UNAM, México, 129-153.
- Eguiarte L.E., Duvall M.R., Learn, Jr. G.H. y Clegg M.T. 1994. The systematic status of the Agavaceae and Nolinaceae and related Asparagales in the Monocotyledons: An analysis based on the *rbcl* gene sequence. *Boletín de la Sociedad Botánica de México* **54**:35-56.
- Farris, J. 1977. On the phenetic approach to vertebrate classification. En: Hecht M.K., Goody P.C. y Hecht B.M., Edrs. *Major patterns in vertebrate evolution*. Plenum, New York, 823-850.
- Farris J. 1982. The logical basis of phylogenetic analysis. En: Platnick, N. , Funk, V. Edrs. *Advances in cladistics: Proceedings of the second meeting of the Willi Henning Society*. Columbia University Press, New York, 7-36.
- Felsenstein, J. 1982. Numerical methods for inferring evolutionary trees. *Quarterly Review of Biology* **57**:379-404.
- Felsenstein J. 1988. Phylogenies from molecular sequences: Inferences and reliability. *Annual Review of Genetics* **22**:521-565.
- Felsenstein J. 1991. *Phylogeny Inference Package*. Univ. Wash. Press, Seattle.
- Forey P.L. 1992. Formal classification. En: Forey P.L., Humpries C.J., Kitching I.J., Scotland R.W., Siebert D.J. y Williams D.W. Edrs. *Cladistics. A practical course in systematics*. Oxford, Claredon Press, 124-136.
- Gaut, B.S., Muse S.V., Clark W.D. y Clegg M.T. 1992. Relative rates of nucleotide substitution at *rbcl* locus in monocotyledonous plants. *Journal of Molecular Evolution* **35**:292-303.
- Hartl D.L. y Clark A.G. 1989. *Principles of population genetics*. Second edition. Sunderland, Mass: Sinauer.
- Haeckel, E. 1866. *Generelle morphologie der organismen*. 2 vols. Georg Reimer, Berlin.
- Harvey P.H. y Pagel M.D. 1991. *The comparative method in evolutionary biology*. Oxford University Press. Oxford.
- Hedrick P.W. 1983. *Genetics of Populations*. Science Books International. Boston.
- Henning W. 1965. Phylogenetic systematics. *Annual Review of Entomology* **10**:97-116.
- Hillis D.M., Huelsenbeck J.P. y Cunningham C.W. 1994. Application and accuracy of molecular phylogenies. *Science* **264**:671- 677.
- Hillis, D.M., Moritz C. y Mable B.K. 1996. *Molecular systematics*, second edition, Sinauer Associates, Sunderland, Mass.
- Hull D.L. 1988. *Science as a process*. University of Chicago press, Chicago.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge, University Press.
- Kitching I.J. 1992. Tree building techniques. En: Forey P.L., Humpries C.J., Kitching I.J., Scotland R.W., Siebert D.J., y Williams D.W. Edrs. *Cladistics. A practical course in systematics*. Oxford: Claredon Press, 44-71.
- Kumar S., Tamura K. y Nei M. 1993. *MEGA: Molecular Evolutionary Genetic Analysis, version 1.0*. Pennsylvania State University, University Park, PA.
- Le Quesne W. 1969. A method of selection of charcteres in numerical taxonomy. *Systematic Zoology* **18**:201-205.
- Li W.H. 1997. *Molecular Evolution*. Sinauer. Sunderlandnd, Mass.
- Li W.H. y Graur D. 1991. *Fundamentals of Molecular Evolution*. Sunderland, Mass: Sinauer.
- Maddison D.R. 1991 The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Zoology* **40**:315-328.
- Maddison W.P. y Maddison D.R. 1992. *MacClade, version 3.0.*, Sinauer, Suderland, Mass.
- Mayr E. 1942. *Systematics and the origin of species*. Cambridge, Cambridge Univeristy Press.
- Mayr E. 1969. *Principles of systematic zoology*. McGraw Hill, New York.
- Mayr E. 1981. Biological classification:towards a synthesis of opposing methodologies. *Science* **214**:510-516.
- Mayr E. 1982. *The growth of biological thought*. Bleknep Press, Cambridge, Mass.
- Michener C.D. y Sokal R.R. 1957. A quantitative approach to a problem in classification. *Evolution* **11**:130-162.
- Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- Nei M. 1991. Efficiencies of different tree-making methods for molecular data. En: Miyamoto M.M. y Cracraft J. Edrs. *Phylogenetic analysis of DNA sequences*. Oxford University press, New York. 90-128
- Saitou N. y Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. **4**:406-425.
- Selander R.K., Clark A.G. y Whittam T.S. Edrs.1991. *Evolution at the molecular level*. Sinauer, Sunderland, Mass.
- Siebert D.Y. 1992. Tree statistics; trees and "confidence"; consensus trees; alternatives to parsimony; character weighting; character conflict and its resolution. En: Forey P.L., Humpries C.J., Kitching I., Scotland R.W., Siebert D.J. y Williams D.W. Edrs. *Cladistics. A practical course in systematics*. Claredon Press, Oxford. 72-88
- Simpson G.G. 1961. *Principles of animal taxonomy*. Columbia University Press, New York.
- Slatkin M. 1994. Gene flow and population structure. En Real L.A. Edr. *Ecological Genetics*. Princeton University Press, Princeton, New Jersey, 3-17.
- Sneath, P.H.A. y R.R. Sokal. 1973. *Numerical taxonomy*. W.H. Freeman, San Francisco.
- Sokal R.R. 1985. The continuing search for order. *American Naturalist* **126**: 729-749.
- Sokal R.R. y Michener C.D. 1958. A statistical method for

- evaluation systematic relationships. *University of Kansas Scientific Bulletin* **28**:1409-1438.
- Sober, E. 1988. *Reconstructiong the past: Parsimony, evolution and inference*. The MIT press. Cambridge, Mass.
- Soltis, A.S. y Soltis D.E. 1996. Phylogenetic analysis of large data sets. *Boletín de la Sociedad Botánica de México* **59**: 99-113.
- Souza V., Eguiarte L., Avila G., Cappello R., Gallardo C., Montoya J. y Piñero D. 1994. Genetic structure of *Rhizobium etli* biovar *phaseoli* associated with wild and cultivated bean plants (*Phaseolus vulgaris* and *P. coccineus*) in Morelos, Mexico. *Applied and Environmental Micorbiology* **60**: 1260-1268.
- Swofford D.L. 1991a. When are phylogenetic estimates from moelcular data incongruent. En: Miyamoto M. M., Cra-
craft, J. Edrs. *Phylogenetic analysis of DNA sequences*, Oxford, New York, 294-333.
- Swofford D.L. 1991b. PAUP 3.0s: *Phylogenetic Analysis Using Parsimony. User's Manual*. Illinois Natural History Survey, Champaign.
- Swofford D.J. y Selander R.B. 1981. BIOSYS-1: FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *Journal of Heredity* **72**: 281-283.
- Swofford D.L., Olsen G.J., Waddell P.J. y Hillis D.H. 1996. Phylogenetic inference. En Hillis, D.M., C. Moritz y B.K. Mable Edrs. *Molecular systematics*, second edition, Sinauer Associates, Suderland, Mass., 407-514.
- Takahata N., Clark A.G., Edrs. 1993. *Mechanisms of molecular evolution*. Japan Scientific Societies Press, Tokyo.