EL USO DE SECUENCIAS GÉNICAS PARA ESTUDIOS TAXONÓMICOS

DOLORES GONZÁLEZ

Departamento de Sistemática Vegetal, Instituto de Ecología, A. C.
Apartado Postal 63, Xalapa, Veracruz, 91000, México.

Resumen. Durante los últimos años el uso de caracteres moleculares en investigaciones taxonómicas ha aumentado considerablemente. Los caracteres más comunes provienen de las enzimas de restricción y la secuenciación de genes o regiones particulares de ADN. Las secuencias potencialmente presentan ventajas sobre otros caracteres moleculares. Se describen los procedimientos para muestrear la variación del ADN mediante secuencias génicas. Entre los procedimientos de secuenciación destaca la aplicación generalizada de los métodos enzimáticos y la tendencia hacia el uso de marcadores no radiactivos así como la secuenciación automatizada. Se presenta una introducción breve a las distintas fases de análisis cladísticos especialmente requeridas para las secuencias de ADN. Además, se proporciona una revisión de la diversidad de genes usados en sistemática. Los más comunes son los del ARN ribosomal (nuclear, mitocondrial y de cloroplasto) y el *rbcL* de cloroplasto. En adición a éstos, se están investigando otros genes como las globinas y la alcohol deshidrogenasa del genoma nuclear, el citocromo b y los ARNs de transferencia del genoma de mitocondria, y el "matK" y el "rpo" del genoma de cloroplasto.

Palabras clave: secuencias génicas, sistemática molecular, alineamiento de secuencias, análisis cladísticos.

Abstract. During the last years, the use of molecular characters for taxonomic research has increased notably. Characters most commonly used come from restriction enzymes and sequencing of genes or particular DNA regions. Sequences present potential advantages over other molecular characters. This paper describes sampling procedures to detect DNA variation through sequencing. Among procedures for DNA sequencing, enzymatic methods are of generalized application, and the tendency is toward the use of non-radioactive markers and automated sequencing. A brief introduction to the different stages of cladistic analysis is also included, especially those required for DNA sequences. A revision of genes used in systematics is provided. The most common are the nuclear, mitochondrial and chloroplast ribosomal RNA genes, and the *rbcL* from chloroplast. Other genes under investigation are the globins and the alcohol dehydrogenase (from the nuclear genome), the cytochrome b and the transfer RNAs (from the mitochondrial genome), and the "matK" and the "rpo" (from the chloroplast genome).

Key words: gene sequences, molecular systematics, sequence alignment, cladistic analysis.

Los caracteres morfológicos, tradicionalmente, han servido como la única base para la mayoría de los estudios taxonómicos (Stevens, 1984, 1986). Estos caracteres han generado el sistema básico de clasificación para la mayoría de los taxa. En la medida que estos sistemas se pretenden refinar, se han usado datos adicionales como: número de cromosomas, fitoquímica, micromorfología, anatomía y ultraestructura entre otros (Bisby *et al.*, 1980; Ferguson, 1980; Grant, 1984).

La incorporación de nuevos tipos de datos, particularmente los caracteres moleculares, ha sido necesaria

en taxa donde los caracteres morfológicos, por sí solos, no han resuelto los problemas taxonómicos. Estos pueden provenir de biomoléculas como proteínas y ácidos nucleicos, e incluyen reacciones inmunológicas, isoenzimas, las secuencias de proteínas, la hibridación de ADN, las enzimas de restricción ("RFLPs" [polimorfismo en la longitud de los fragmentos de restricción] y mapas génicos), los "RAPDs" (polimorfismo del ADN amplificado al azar), el "fingerprinting" (huella digital) del ADN y las secuencias de genes particulares. Los méritos intrínsecos del uso de este tipo de datos en sistemática ya se ha discutido en la

literatura (Buth, 1984; Patterson, 1987; Hillis y Moritz, 1990; Bruns *et al.*, 1991; Soltis *et al.*, 1992; Doyle, 1993). En esta revisión ya no se abordarán los argumentos para seleccionar uno u otro tipo de caracteres moleculares ni las ventajas de éstos respecto a los no-moleculares.

Entre los distintos tipos de caracteres moleculares las secuencias génicas son los más recomendables en la mayoría de los casos. En este artículo se hace énfasis en el uso de secuencias génicas para resolver problemas taxonómicos, las cuales se han usado en un gran número de organismos, desde bacterias hasta primates (Hillis *et al.*, 1990). Algunas de las razones para usar secuencias génicas en estudios sistemáticos son: 1] representan una fuente abundante de caracteres (potencialmente cada nucleótido de la secuencia), 2] tienen diferentes propiedades estructurales y funcionales y 3] presentan diferentes tasas evolutivas. Estas tres propiedades se han podido observar tanto en plantas como en animales (Ritland y Clegg, 1987; Field *et al.*, 1988; Soltis *et al.*, 1990; Giannasi *et al.*, 1992). No obstante, las secuencias génicas no son la única fuente de datos por lo que no deben reemplazar a otros, especialmente a los que no son moleculares. Por el contrario, un estudio taxonómico se debe basar en toda la información disponible, molecular y no molecular, para los organismos de interés.

El estudio de la variación taxonómica de las secuencias génicas ha sido posible gracias a los avances tecnológicos, para la extracción, amplificación y secuenciación de ácidos nucleicos, avances que han proliferado a partir de finales de la década de los años ochenta. No obstante, en México aún no se ha aplicado esta tecnología debido a la falta de infraestructura e investigadores entrenados. Sin embargo, ya existen antecedentes de trabajos hechos por mexicanos, de manera que se podrán resolver estos dos factores limitantes (González, 1992; Barrera *et al.*, 1993; Delgado-Salinas *et al.*, 1993; Dorado *et al.*, 1992; Eguiarte *et al.*, 1994; Arias y Rieseberg, 1995).

El objetivo de este trabajo es hacer una revisión tanto de los procedimientos para obtener los datos del ADN mediante secuencias, así como de la diversidad de genes más usados en sistemática. Es importante señalar que la buena calidad de los estudios en este campo no está garantizada simplemente por el uso de datos moleculares, la calidad de un estudio sistemático está dada por el manejo adecuado de la combinación "tipo de datos-tipo de análisis". Los datos moleculares, como cualquiera otros por sí solos no dan soluciones a problemas taxonómicos, y su análisis debe hacerse con procesos explícitos y abiertos a la crítica. Por otro lado, no todos los métodos de análisis de datos disponibles en sistemática (fenéticos y

cladísticos) son adecuados y recomendables (De Luna, 1995). Por esto, es necesario evaluar también los métodos de análisis a los que se someterán los datos moleculares. Estos aspectos son fundamentales, por lo que también se consideran en esta revisión, aunque de una forma somera.

Procedimientos para la secuenciación de genes particulares

La variación del ADN se puede muestrear con diferente cobertura y resolución. La cobertura puede ser tan amplia como la que se logra con la hibridación del ADN total, o tan reducida como cuando se muestrea sólo una sección de un gen particular. La resolución puede ser tan gruesa como obtener un índice de similitud entre dos moléculas de ADN o tan fina como cuando se conoce en detalle la secuencia de nucleótidos. Los procedimientos de muestreo con enzimas de restricción y secuenciación proveen poca cobertura. Por razones prácticas y de costos, generalmente sólo se explora un gen o una sección pequeña de ADN. Sin embargo, los métodos de secuenciación son los que dan la resolución más fina porque permiten muestrear en mayor detalle la variación.

En esta sección se presentan los procedimientos para la secuenciación de genes particulares. En taxonomía, estos procedimientos se han aplicado a un número limitado de genes. Históricamente se inició el muestreo de la variación en diversos genes a nivel de enzimas de restricción. Estas exploraciones continuaron con la secuenciación del subconjunto de genes que resultaron ser más útiles con niveles de variación apropiados a la categoría taxonómica. En la parte final de esta revisión, se presenta la diversidad de genes secuenciados para su uso en sistemática.

La secuenciación de ADN para propósitos taxonómicos consiste primero en seleccionar un segmento particular del genoma (nuclear o extranuclear) que sea apropiado para el grupo taxonómico de interés. El segmento puede ser un gen, un intrón o una secuencia espaciadora y puede provenir del ADN nuclear, de cloroplasto o de mitocondria. Posteriormente, el ADN total se aísla y se purifica mediante distintos procedimientos químicos que dependen del organismo y del tejido de origen.

A partir del ADN total se separa el segmento particular del genoma. El ADN primero se desnaturaliza con calor para abrir las cadenas. La región de interés se aísla mediante el uso de "primers" (iniciadores). Estos son cadenas de 20 a 30 nucleótidos cuya secuencia es específica a una región particular de ADN. Los iniciadores se hibridan flanqueando una región determinada permitiendo que una ADN polimerasa re-

plique selectivamente tal región. De este modo, entre todas las regiones del ADN total sólo se amplifican las que tienen las secuencias iniciadoras. Esta amplificación enzimática se logra con la "reacción en cadena de la polimerasa" (RCP, [o "PCR" en inglés] Mullis y Faloona, 1987; Saiki *et al.*, 1988). Este proceso de amplificación involucra la desnaturalización del ADN con calor; la hibridación de dos iniciadores con el ADN desnaturalizado y la amplificación de la porción de ADN que se encuentra entre los dos iniciadores (figura 1). Esto se logra con la ayuda de una polimerasa estable al calor y con la adición de nucleótidos. La desnaturalización, hibridación y amplificación se repite varias veces cíclicamente (25-30) de tal forma que al final se tienen suficientes copias de una región de ADN para secuenciar. La introducción de la RCP ha facilitado enormemente la secuenciación de genes o regiones particulares del ADN. En la actualidad es la técnica de amplificación más usada en sistemática molecular y ha desplazado a la de clonación de genes específicos en bacterias.

Existen otras estrategias de amplificación de ADN con la RCP para secuenciación. La amplificación asimétrica de ADN (McCabe, 1990), la que usa quinasas/exonucleasas (Higuchi y Ochman, 1989) y la amplificación selectiva de una cadena de ADN usando iniciadores biotinilados (Mitchell y Merrill, 1989). Todas estas estrategias se desarrollaron para amplificar sólo una cadena de la molécula de ADN, sin embargo, su uso es menos frecuente ya que resultan más laboriosas que la amplificación enzimática directa de las dos cadenas. Una vez amplificada la región de interés, se puede secuenciar usando el método químico (Maxam y Gilbert, 1977, 1980) o el enzimático (Sanger *et al.*, 1977).

Método químico de secuenciación. El método químico de secuenciación (figura 2) se basa en el uso de reactivos como la piperidina que rompe la doble cadena de ADN en cada sitio donde se encuentre un nucleótido metilado, protonado o sustituido por su isómero. Este método requiere primero un fragmento de ADN cortado con una enzima de restricción. Este fragmento de doble cadena se marca específicamente en una de las terminaciones (5' ó 3') con un isótopo radiactivo. En cada molécula de ADN, las dos cadenas están marcadas en la misma terminación. Se pueden seguir dos estrategias para retener sólo una cadena marcada como base de la reacción de secuenciación. La primera es cortar el ADN con una segunda enzima para originar dos fragmentos que contienen sólo una cadena marcada. Los fragmentos se separan en una electroforesis y se extrae del gel uno de ellos para secuenciar. La segunda estrategia es desnaturalizar el

fragmento de ADN para separar las dos cadenas marcadas, se corre una electroforesis y se extrae sólo una de ellas para secuenciarla.

La solución que contiene sólo las cadenas marcadas en una terminación, separadas con una enzima de restricción o por desnaturalización, se divide en cuatro muestras. Cada una se somete a una reacción química con reactivos que metilan, protonan o substituyen por su isómero a nucleótidos específicos dentro de la cadena. En la primera muestra se hace una reacción de metilación. Esto se logra al añadir dimetil-sulfato a la molécula de ADN, para metilar selectivamente guaninas. Al añadir piperidina a la reacción la cadena de ADN se rompe en los sitios donde se encuentren guaninas metiladas y origina fragmentos de longitud variable pero cuyo nucleótido terminal es guanina. La segunda muestra genera fragmentos cuyos nucleótidos terminales son adenina y guanina. Esto se realiza con ácido fórmico que protona los anillos de nitrógeno de las adeninas y guaninas. Al añadir piperidina la cadena de ADN se rompe en todos los sitios donde se encuentren los dos nucleótidos protonados. La tercera muestra origina fragmentos que terminan en timina y citosina. Esto se logra añadiendo hidrazina, la cual reacciona con la timina y citosina del ADN y sustituye estos nucleótidos por su isómero, la ribosilurea. Al añadir piperidina a esta muestra, la cadena de ADN se rompe en los sitios donde haya ribosilurea. La cuarta muestra sólo origina fragmentos que terminan en citosina. Esta muestra es tratada como la tercera pero se adiciona NaCl lo cual inhibe la reacción de la timina con la hidrazina y sólo se substituyen las citosinas por ribosilurea.

Los fragmentos generados en cada una de las cuatro muestras se someten a una electroforesis vertical en un gel de poliacrilamida para ser separados por su tamaño (figura 2). La fragmentación de la región de ADN seleccionada produce una población de fragmentos muy grande y de longitudes diversas. Esta población incluye fragmentos rotos en cada una de las posiciones de los nucleótidos. Como cada fragmento está marcado, es posible visualizarlos por medio de una autorradiografía al exponer el gel a una película sensible a la radiación o a través de un analizador de imágenes.

Método enzimático de secuenciación. El método enzimático de secuenciación (figura 3) se basa en que la ADN polimerasa termina la síntesis donde se encuentre un nucleótido análogo (los 2',3'-dideoxynucleosido 5'-trifosfatos, [ddNTP's]). Normalmente, la enzima ADN polimerasa incorpora nucleótidos regulares (los 2',3'-dideoxynucleosido 5'-trifosfatos, [dNTP's]) durante la síntesis de ADN. En la secuenciación enzimática se

Figura 1. La amplificación enzimática de ADN (RCP) es un proceso cíclico que multiplica una región específica de ADN. Para ello se requieren cuatro ingredientes en solución: el ADN total, los primers o iniciadores, la ADN polimerasa y los nucleótidos. Cada ciclo incluye: 1) La desnaturalización del ADN con calor. 2) La hibridación del iniciador en cada extremo de la región de ADN que se va a amplificar. 3) La extensión del ADN. En esta fase, la ADN polimerasa cataliza la amplificación de una región específica de ADN. Al terminar cada ciclo, la temperatura se eleva nuevamente 4), las cadenas recién formadas se separan otra vez y el proceso empieza de nuevo. Al repetir la reacción 5) se incrementa la cantidad de ADN exponencialmente. Al concluir todos los ciclos, la reacción se somete a una extensión final 6). Esto permite que la polimerasa termine la incorporación de los nucleótidos en todas las cadenas de ADN formadas. El cuadro sombreado indica la temperatura y tiempo más comunes para cada fase durante la amplificación enzimática de ADN.

utiliza un iniciador el cual se hibrida al gen o región del ADN que se va a secuenciar. La polimerasa replica la porción seleccionada incorporando dNTP's continuamente hasta que se encuentra con uno de los cuatro ddNTP's y entonces termina la reacción.

La reacción de terminación se lleva a cabo en cuatro muestras por separado con cada uno de los cuatro nucleótidos análogos. En cada muestra, la síntesis del ADN se interrumpe selectivamente en A, C, G ó T dependiendo de uno de los ddNTP's que se encuentre en solución. Los ddNTP's carecen del grupo 3'-OH, necesario para continuar la incorporación de los nucleótidos regulares. En las cuatro muestras todos los fragmentos que se sintetizaron a partir del gen o región del ADN presentan el mismo nucleótido como origen común pero terminan con diferentes nucleótidos. Dentro de cada muestra todos los fragmentos terminan en el mismo nucleótido pero son de diferente longitud. Los fragmentos se separan de acuerdo a su tamaño en una electroforesis vertical de un gel de poliacrilamida. Durante la síntesis se incluye también un isótopo radiactivo o substratos quimioluminiscentes que permiten visualizar los fragmentos por medio de una autorradiografía o con un analizador de imágenes.

La incorporación de los isótopos radiactivos en la secuenciación enzimática se puede realizar de tres formas (figura 3). La primera combina en un solo paso el proceso de marcar la cadena y las reacciones de síntesis y terminación. Esto se logra al substituir uno de los dNTP's por el mismo dNTP pero marcado con un isótopo radiactivo. La segunda, separa la reacción de secuenciación en dos pasos: primero se marca la cadena y luego se realizan las reacciones de síntesis y terminación. La cadena se marca durante la síntesis previa de una pequeña porción de ADN usando concentraciones muy bajas de los dNTP's y uno marcado radiactivamente. Las reacciones de síntesis y terminación se completan al adicionar nuevamente dNTP's y uno de los cuatro ddNTP's. La tercera forma consiste en marcar directamente el iniciador con un isótopo radiactivo antes de que se hibride con el gen o con la porción de ADN que se va a secuenciar. Este último modo de marcar el ADN durante la secuenciación es el mismo que se usa para incorporar los substratos quimioluminiscentes.

Los isótopos radiactivos más ampliamente usados en la secuenciación química y enzimática son los de azufre (alfa-³⁵S-dATP) y fósforo (alfa-³²P-dATP). El alfa-³⁵S-dATP es de baja toxicidad pero su vida media (87 días), así como el tiempo de exposición de las autorradiografías es mayor, (72-96 hrs). En contraste, el alfa-³²P-dATP es más tóxico pero su vida media (14 días) y su tiempo de exposición (18-24 hrs) es menor compara-

do con el isótopo de azufre (Chase, 1959). Los isótopos gama-³⁵S-dATP y gama-³²P-dATP también se emplean en la secuenciación enzimática pero sólo para marcar los oligonucleótidos iniciadores. Los substratos quimioluminiscentes más usados en la secuenciación enzimática son el luminol y el adamantil 1,2-dioxietano aril fosfato (AMPPD). Estos dos substratos emiten luz cuando decaen de un nivel electrónico a otro. El luminol adquiere un estado excitado cuando se oxida con un ion metálico y peroxidasas, mientras que el AMPPD se excita al desfosforilarlo con una fosfatasa alcalina.

Recientemente se han introducido los secuenciadores automatizados (Ferl *et al.*, 1991). Esta nueva tecnología facilita los trabajos que involucran la comparación de secuencias génicas. Por un lado, ya no es necesario usar radiactividad para marcar los nucleótidos y poder visualizarlos, y por otro, la captura de secuencias se hace más rápida debido a que incorporan programas que leen los nucleótidos directamente desde el gel. El método de secuenciación que se emplea con estos aparatos es el enzimático, y se utilizan como marcadores fluoróforos que pueden aplicarse tanto al iniciador como a los ddNTP's. En cada caso se usan cuatro diferentes fluoróforos que identificarán a cada uno de los nucleótidos (A, G, T y C). Los fluoróforos más usados son los derivados de la rodamina y la fluoresceína. Conforme se lleva a cabo la electroforesis los fluoróforos son excitados por un rayo laser. Cada emisión fluorescente es transmitida como una señal a una computadora donde un programa la interpreta y la codifica como un nucleótido particular. Este tipo de programas (Sequence Navigator, Applied Biosystems) archivan las secuencias de modo que pueden ser exportadas a otros programas para su posterior análisis. Esta forma de secuenciación, aunque cara, es una mejor alternativa a los métodos con isótopos radiactivos (Ferl *et al.*, 1991).

Análisis filogenético de secuencias génicas

El uso de secuencias génicas para estudios taxonómicos tiene el potencial de ofrecer información abundante para inferir historia filogenética debido a que el número de caracteres taxonómicos está limitado únicamente por el tamaño del genoma. El mayor número de caracteres potencialmente comparables puede incrementar la resolución en los análisis filogenéticos al reducir los límites impuestos por la falta de caracteres morfológicos de algunos taxa. El análisis filogenético de las secuencias, al igual que con cualquier otro tipo de caracteres, puede dividirse en tres fases: análisis de caracteres, búsqueda de cladogramas y evaluación de cladogramas. El análisis de caracteres incluye todas las operaciones que

Figura 2. Método químico de secuenciación. Este método se basa en la ruptura específica del ADN por sustancias químicas. Primero se marcan las dos cadenas en una sola terminación 1). Sólo una de las cadenas marcadas se separa mediante enzimas de restricción o por desnaturalización. La solución de cadenas marcadas se somete a cuatro tratamientos químicos para cortar selectivamente los nucleótidos. Las sustancias químicas empleadas cortan en una o dos de las cuatro bases. Tomando como referencia la secuencia de cadena sencilla mostrada en 2), la solución con dimetil-sulfato producirá una población de fragmentos heterogéneos cortados siempre en cada sitio donde exista una guanina 3). Del mismo modo sucede con las otras tres soluciones. 4) Cada columna en el gel de la electroforesis separa los fragmentos de una solución. La flecha a la derecha del gel indica la dirección de la electroforesis. La flecha a la izquierda indica la dirección de la lectura de las bases en el gel. La secuencia al lado izquierdo del gel es la que se lee siguiendo las marcas en las cuatro columnas del gel. Si sólo consideráramos las dos reacciones donde se realiza una doble ruptura habría ambigüedad entre A o G y entre T o C. Sin embargo, para resolver a cuál base corresponde se compara la información de estas columnas con una columna paralela en donde se suprime químicamente una de las rupturas.

llevan a la formación de la matriz de datos y éstas incluyen el estudio de la variación entre taxa, el reconocimiento de estados y su distribución en las unidades de estudio (codificación). En esta primera fase analítica también se incluye la elaboración de hipótesis de vecindad entre estados (orden) y las de dirección de cambio evolutivo entre estados (polaridad).

La segunda fase analítica incluye todas las operaciones que llevan a la elaboración de cladogramas. Esta búsqueda de cladogramas procede mediante algoritmos diversos como el de parsimonia Wagner, parsimonia Dollo, parsimonia Fitch, etc. La tercera fase incluye todas las operaciones que se realizan para evaluar la robustez de cladogramas en conjunto respecto a otros cladogramas o la robustez relativa de clados particulares dentro de un cladograma. De estas tres fases de análisis, la primera es la que tiene más aspectos especiales para el caso del análisis de secuencias génicas. Esto se debe a que se conocen distintos tipos de mutaciones (indels, transversiones y transiciones) en las secuencias comparadas, así como la estructura secundaria de algunos genes. Estos aspectos peculiares del análisis de caracteres comprenden: el alineamiento, la recodificación de caracteres y la asignación de peso diferencial.

Análisis de caracteres. El alineamiento es un aspecto crucial en los análisis de secuencias ya que se hipotetizan relaciones de homología (táxica o transformacional) entre los nucleótidos de cada posición lo que es el fundamento de la reconstrucción filogenética misma. Para una posición, si las bases son iguales, entonces existe una relación de homología táxica; pero si las bases son diferentes, entonces se postula una relación de homología transformacional. Debido a que a menudo la longitud de las secuencias varía entre los taxa, es difícil establecer la correspondencia entre cada nucleótido. Por lo tanto, para la misma colección de secuencias existen varios modos de compararlas. El alineamiento de las secuencias puede hacerse manualmente ("a ojo") y mediante programas de cómputo (automático). En ambos casos el alineamiento puede considerar un par de secuencias a la vez o evaluar todas las secuencias simultáneamente.

Por un lado, en los alineamientos "a ojo" los nucleótidos de un taxon se desplazan en relación a los de otro taxon para encontrar su correspondencia con la ayuda de procesadores de textos o programas sencillos que permiten la edición de las secuencias. Este criterio de alineamiento es el más común y para algunos investigadores es el mejor. Swofford y Olsen (1990) indicaron que si no se pueden alinear las secuencias "a ojo", es mejor no hacer suposiciones de homología. Por otro lado, en los alineamientos au-

tomáticos con programas de cómputo las secuencias se desplazan con la ayuda de algoritmos que optimizan su similitud. Una forma de optimizar la similitud entre las secuencias es la de asignar costos o penalizaciones a cada tipo de mutaciones (Needleman y Wunsch, 1970). Por ejemplo, para alinear la secuencia 1=CGTAAGG con la secuencia 2=CGTTAATG, dos posibles desplazamientos serían: 1=CGTA-AGG y 1=CGT-AA-GG. El programa registra todos los desplazamientos posibles y escoge el alineamiento con la menor penalización. Si las substituciones cuestan 1 y las indels de una base 2, el primer alineamiento tendría una penalización de 4 y el segundo de 6. Los costos que se asignen a cada tipo de mutación deben considerarse cuidadosamente debido a que pueden alterar significativamente el alineamiento de las secuencias. Otras formas de optimizar la similitud entre las secuencias se basan en algunos rasgos conocidos de las moléculas como son la estructura secundaria, los sitios TATA y los sitios de corte de los intrones.

En los alineamientos automáticos los programas de cómputo generalmente siguen dos estrategias. Una de ellas es la evaluación de una secuencia a la vez, y la otra el alineamiento global. En el primer caso, una secuencia se optimiza con la secuencia de referencia. A continuación, una tercera secuencia se optimiza a la primera y segunda, y así sucesivamente. Este método funciona bien cuando los desplazamientos requieren pocas indels o éstas no están muy juntas. Este alineamiento es muy sensible al orden en que se añaden las secuencias. Sin embargo, es rápido debido a que todas las secuencias se alinean una por una con referencia a la primera. Un programa que efectúa este tipo de alineamiento es Align Plus para DOS. En el segundo caso, el alineamiento global, también conocido como alineamiento progresivo, trata de optimizar la similitud sobre todas las secuencias a la vez. Un programa que realiza alineamientos globales son es el Clustal V para Macintosh.

Una vez alineadas las secuencias, cada posición representa un carácter taxonómico potencial si por lo menos dos taxa presentan un nucleótido diferente. En una posición, los estados pueden ser los cuatro nucleótidos y las indels. Esta es la forma más común de comparar la variación taxonómica de las secuencias. Sin embargo, es posible analizar esta variación bajo otros sistemas de codificación de estados, en cuyo caso los cuatro nucleótidos de las secuencias (ACGT) se recodifican antes de la búsqueda de los árboles filogenéticos. Una forma de recodificación es la asignación de un código para las purinas y otro para las pirimidinas. El resultado es que cada posición potencialmente sólo puede tener dos estados: purinas y pirimidinas, ignorándose por completo las transiciones

Figura 3. Método enzimático de secuenciación. Este método se basa en la síntesis *in vitro* de una de las cadenas de ADN con una polimerasa. 1) La síntesis se inicia en el sitio donde un iniciador se encuentra unido a la cadena de ADN. Al mismo tiempo se puede incorporar una marca radiactiva a la cadena, ya sea 2) durante la síntesis, ó 3) directamente en el iniciador. 4) La reacción de síntesis se termina al incorporar un nucleótido análogo que carece del grupo 3'-OH que es necesario para que la polimerasa catalice la incorporación de otro nucleótido. Esto se lleva a cabo en cuatro reacciones por separado, cada una de las cuales contiene los 4 dNTP's y uno de los cuatro ddNTP's. 5) Cada columna en el gel de la electroforesis separa los fragmentos de una de las cuatro soluciones. Los fragmentos tienen un origen común pero terminan en un nucleótido diferente. La flecha a la derecha del gel indica la dirección de la electroforesis. La flecha a la izquierda indica la dirección de la lectura de las bases en el gel. La secuencia obtenida es la complementaria a la cadena usada como base para la secuenciación.

(Lake, 1987). Esta forma de recodificación se aplica porque existe la tendencia a que las transiciones son más frecuentes que las transversiones en algunos genomas (Brown *et al.*, 1982). Otra forma de recodificación es la traducción de los nucleótidos a aminoácidos. En este tipo de recodificación la secuencia se divide primero en codones que posteriormente se traducen a su correspondiente aminoácido. El resultado es una matriz en la que cada carácter puede tener 20 posibles estados (aminoácidos). Cuando se recodifican los nucleótidos a aminoácidos se ignoran las mutaciones silenciosas (Goodman *et al.*, 1985). Esta forma de recodificación se aplica para disminuir el efecto de las substituciones de la tercera posición de los codones en la topología de los árboles.

Ya sea que se consideren los cuatro nucleótidos como estados o se recodifiquen, cada posición variable en la secuencia provee información sobre homología táxica y transformacional. Todas las posiciones informativas pueden considerarse igualmente importantes o algunas pueden tener mayor valor. El peso diferencial de los caracteres se puede asignar *a priori*, es decir, antes de la construcción del árbol filogenético, y *a posteriori*, después de la construcción del árbol. El primero refleja lo que se conoce de los caracteres y los estados, por ejemplo, su frecuencia o su importancia en el mantenimiento de la estructura secundaria. El segundo refleja la contribución que hacen esos caracteres a la topología resultante. Existen varios métodos para asignar pesos *a priori* o *a posteriori* como son: el combinatorio, el dinámico interactivo y el sucesivo, entre otros (Wheeler, 1990; Williams y Fitch, 1990; Knight y Mindell, 1993).

La asignación de peso diferencial a ciertos caracteres es muy controversial. El peso diferencial se usa porque existe la tendencia a encontrar en algunas secuencias génicas un tipo de mutaciones más que otro, y entonces se le da más peso a las que son menos frecuentes. Por ejemplo, se ha observado que en el ADN mitocondrial y en algunos pseudogenes las transiciones son más frecuentes que las transversiones (Brown *et al.*, 1982; Li *et al.*, 1984). Debido a esto, frecuentemente se les asigna mayor peso a las transversiones que a las transiciones en los análisis con secuencias génicas (Swofford y Olsen, 1990). Otros caracteres a los que también se les ha asignado peso diferencial son los nucleótidos que mantienen la estructura secundaria de la molécula, por ejemplo, los nucleótidos apareados y no apareados que forman las asas del ARN ribosomal (Vawter y Brown, 1993).

Búsqueda de árboles filogenéticos. Una vez codificados los estados y decididos los esquemas de peso, la matriz de datos está lista para ver el cladograma implícito.

La siguiente fase de análisis lo constituye la búsqueda de árboles filogenéticos. Los árboles filogenéticos son representaciones gráficas que resumen las relaciones evolutivas entre los organismos. Existen tres clases de métodos para construir árboles a partir de secuencias génicas: los basados en distancias, el método de "maximum-likelihood" (máxima verosimilitud) y los basados en parsimonia. Para un número particular de taxa existe un conjunto limitado de posibles árboles. Estos métodos buscan el o los árboles que minimizan (o maximizan) el criterio de búsqueda. La búsqueda de los árboles es un procedimiento que requiere el uso de programas de cómputo, algunos de ellos son PAUP, PHYLIP y HENNIG86.

Los métodos basados en distancias buscan el árbol que maximiza la correspondencia entre las distancias implícitas en el árbol y las medidas entre los taxa. Por esto, se requiere la elaboración de una matriz con las distancias para cada par de taxa. Los tres índices para el cálculo de distancias más comúnmente usados son: el de Nei, el de Roger y el de Cavalli-Sforza y Edwards (Swofford y Olsen, 1990). Una vez que se han obtenido los valores de distancia con cualquiera de estos índices hay varios algoritmos para obtener un árbol. Los métodos principales son: el de Fitch-Margoliash, el de distancias Wagner y el del vecino más cercano (Nei, 1991; Williams, 1992). La diferencia entre estos algoritmos consiste en la forma de calcular la optimización entre las distancias derivadas de un árbol y las distancias calculadas entre pares de taxa.

Los métodos de máxima verosimilitud usan las secuencias originales en lugar de medidas de distancias, y buscan el árbol que maximiza la posibilidad de que un modelo específico de cambio de nucleótidos haya generado las secuencias incluidas en el análisis a partir de una de ellas (Felsenstein, 1984). La cuantificación de máxima verosimilitud de los datos observados depende de aceptar un modelo de cambio evolutivo. Estos modelos a menudo incorporan probabilidades iguales de cambio de un nucleótido a otro como puede ser el modelo de Jukes-Cantor. Otros modelos hacen ajustes ya que se ha observado que las probabilidades de sustitución no son iguales entre las bases. El modelo de dos parámetros de Kimura permite considerar tasas desiguales entre transiciones y transversiones. Sin embargo, ninguno de estos modelos incorpora las probabilidades de cambio de las inserciones o deleciones. Todos los modelos derivan en fórmulas matemáticas que permiten calcular las probabilidades particulares de que un nucleótido inicial cambiará a otro nucleótido. Esas probabilidades para cada posición se combinan en una probabilidad general para toda la secuencia como producto de las probabilidades en cada posición. Por

esto se requiere de mucho tiempo de cómputo y sólo es posible hacerlo con un número reducido de taxa. Las hipótesis filogenéticas inferidas son el árbol que implica la mayor verosimilitud de obtener las secuencias observadas a partir de una de ellas, pero estas estimaciones difieren porque pueden derivarse de diferentes modelos de cambio de nucleótidos (Felsenstein, 1981, 1984, 1988; Swofford y Olsen, 1990).

Los métodos de parsimonia también usan las secuencias originales en lugar de medidas de distancias, pero buscan el árbol que minimiza la cantidad de cambios evolutivos requeridos para explicar las diferencias entre taxa. La cuantificación de los cambios evolutivos en un árbol está dada en términos del número total de transformaciones de estados de los caracteres bajo un modelo evolutivo particular. Algunos modelos para calcular el número de cambios de estado son: el de parsimonia Wagner, parsimonia Fitch, parsimonia Dollo, parsimonia Camin-Sokal y otros (Swofford y Olsen, 1990). La diferencia entre los modelos es la imposición de restricciones a ciertas transformaciones de estados. Por ejemplo, parsimonia Fitch permite cuantificar como un solo paso la transformación de cualquier estado a otro en un carácter multiestado, como el caso de una posición en la secuencia (Fitch, 1971). En cambio, el modelo de parsimonia Wagner contabiliza la misma transformación como varios pasos adicionales, tantos como estados intermedios existan (Farris, 1970). Bajo este modelo, la transformación de un estado a otro implica su transformación a los estados intermedios tal como se define en la hipótesis de orden de estados. Otros modelos imponen restricciones adicionales sobre ciertos tipos de transformación. Por ejemplo, el modelo de parsimonia Dollo permite que cada estado de carácter apomórfico se origine sólo una vez (Farris, 1977). El modelo de parsimonia Camin-Sokal elimina la posibilidad de reversiones de un estado derivado a su condición relativamente más plesiomórfica (Camin y Sokal, 1965). Dependiendo del modelo de transformación de estados seleccionado, el criterio de parsimonia selecciona el árbol con la longitud total mínima.

Evaluación de robustez. La tercera fase del análisis incluye los procedimientos no estadísticos y estadísticos que se realizan para evaluar la robustez de clados particulares dentro de un cladograma o de cladogramas en conjunto respecto a otros cladogramas. La solidez relativa de los clados dentro de un cladograma puede estimarse simplemente al contar el número de sinapomorfias que sustentan cada grupo. Otro método consiste en calcular el índice de decaimiento ("decay index") para cada clado en un cladograma. Este índice es el número de pasos que se pueden

incrementar en árboles sucesivamente menos parsimoniosos hasta que un clado en particular desaparece del árbol de consenso (Bremer, 1994). También se ha sugerido remover caracteres uno por uno para identificar los que son cruciales como apoyo de clados particulares (Davis, 1993). Las comparaciones entre árboles pueden ser tan sencillas como comparar uno con otro mediante los índices de consistencia (CI), de retención (RI) y el de exceso de homoplasia (HER). Estos índices describen el grado de ajuste entre un árbol y los datos en términos de los niveles de homoplasia y sinapomorfia (Archie, 1989; Farris, 1989).

La evaluación de robustez también se ha intentado por operaciones estadísticas de remuestreo y analíticas. Unos métodos se basan en el remuestreo de los datos con reemplazo, ("bootstrap") para replicar la búsqueda de árboles. Estos análisis permiten calcular la frecuencia con la que se detecta cada clado en las réplicas (Felsenstein, 1985). Otro enfoque es mediante el remuestreo sin reemplazo ("jackknife", Penny y Hendy, 1986; Felsenstein, 1988). Entre los métodos estadísticos que se basan en procedimientos analíticos destacan los límites de confianza (Felsenstein, 1984), las pruebas de significancia de las distancias internodales (Nei *et al.*, 1985) y la comparación de los valores de verosimilitud entre varios árboles (Felsenstein, 1988; Li y Gouy, 1991). Estas evaluaciones estadísticas de robustez se aplican a árboles que se derivan de búsquedas basadas en parsimonia, en distancias o máxima verosimilitud.

Genes usados en Sistemática

Las tablas 1, 2 y 3 resumen los genes usados en trabajos taxonómicos recientes y el nivel taxonómico cubierto. Las referencias se obtuvieron de cuatro revistas (*Cladistics*, *Molecular Biology and Evolution*, *Systematic Biology*, *Systematic Botany*) publicadas entre enero de 1993 y febrero de 1997. En conjunto, la cobertura taxonómica de los estudios con secuencias génicas es muy heterogénea y van desde el nivel de poblaciones hasta divisiones y reinos. La mayoría de estos estudios incluyen sólo uno o dos niveles taxonómicos. Los primeros estudios se realizaron por arriba de orden. Sin embargo, los más recientes son de niveles inferiores al de género y actualmente son los más abundantes. Las referencias se clasificaron de acuerdo al nivel taxonómico, pero con fines prácticos sólo se reconocieron cuatro clases arbitrarias. La primera clase incluye los estudios desde el nivel de población hasta el nivel de género. En la segunda se clasificaron las referencias que estudiaron grupos hasta el nivel de familia. La tercera clase considera los estudios registrados hasta el nivel de orden. La cuarta,

incluye los estudios por arriba del nivel de orden. En cada clase se incorporaron como máximo dos referencias, y sólo las más recientes.

Las secuencias utilizadas en estudios taxonómicos provienen tanto de genes nucleares, mitocondriales como de cloroplastos. Aunque todavía son muy pocos los genes usados en sistemática, las diferentes tasas evolutivas que presentan han sido de gran utilidad para proponer hipótesis filogenéticas en diversos grupos taxonómicos vegetales y animales. Además, dependiendo del gen que se estudie se ha podido obtener resolución de las relaciones filogenéticas a varios niveles taxonómicos desde especies hasta divisiones y reinos. En esta sección se presentan los genes más usados en estudios taxonómicos.

Genes nucleares. Entre los genes nucleares destacan los que codifican para el ARN ribosomal. La estructura de

estos genes incluye una región codificadora, una región no codificadora pero que es transcrita y una región espaciadora que no se transcribe (NTS). En conjunto, estas tres regiones forman el precursor del ARN ribosomal. La región codificadora se transcribe en la subunidad corta, la subunidad 5.8S y la subunidad larga. La región que no codifica para ARN ribosomal pero que es transcrita está formada por los espaciadores intergénicos (ITS) 1 y 2 los cuales son degradados durante el procesamiento del ARN. El gen del ARN ribosomal de la subunidad 5S generalmente es transcrito en forma independiente de los otros ARN (figura 4).

Los genes nucleares que codifican para el ARN ribosomal han sido los más ampliamente utilizados para estudios sistemáticos, tanto en microorganismos como plantas y animales (Olsen *et al.*, 1986; Hillis y Dixon, 1991; Bruns *et al.*, 1992; Hamby y Zimmer, 1992). Esto

Figura 4. Procesamiento de los genes de ARN ribosomal. 1) El gen para el ARN ribosomal consiste en una región NTS (no se transcribe), dos regiones ITS (se transcriben) y tres regiones codificadoras. 2) Este gen se transcribe y se cortan las regiones ITS no codificadoras, las cuales se degradan. También se separan los tres ARN ribosomales que van a formar parte de las dos subunidades de los ribosomas. 3) El ARN ribosomal 18S se incorpora a la subunidad pequeña de los ribosomas, mientras que el ARN ribosomal 5.8S y 28S se incorporan a la subunidad larga. Esta subunidad además incorpora ARN ribosomal 5S transcrito desde otra región del genoma.

Tabla 1. Genes nucleares usados en estudios taxonómicos y nivel taxonómico cubierto. Entre paréntesis se indica la región secuenciada del ARN ribosomal.

	Categoría Taxonómica			
	≤ género	≤ familia	≤ orden	> orden
ARN ribosomal	Campbell <i>et al.</i> , 1997 (ITS); Soltis <i>et al.</i> , 1996 (ITS)	Bayer <i>et al.</i> , 1996 (ITS); Vogler <i>et al.</i> , 1997	Lafay, 1995 (28S)	Winnepenninckx <i>et al.</i> , 1996 (18S); Chaw <i>et al.</i> , 1997 (18S)
Factor de elongación-1-alfa		Cho <i>et al.</i> , 1995		
Miosina (Mlc1)	Clark <i>et al.</i> , 1996			
Ependimina				Ortí y Meyer, 1996
Fosfoglucoisomerasa	Ford <i>et al.</i> , 1995; Gottlieb y Ford, 1996			
Carbamilofosfato sintetasa				Lawson <i>et al.</i> , 1996
Tubulina				Keeling y Doolittle, 1996
Fitocromo		Mathews y Sharrock, 1996		
Caseína		Gatesy <i>et al.</i> , 1996		
Globinas			Hasegawa y Adachi, 1996	
Aldolasa A y C, histona H2AF y un exon del complejo de histocompatibilidad DQA	Slade <i>et al.</i> , 1994			
alfa-amilasa	Okuyama <i>et al.</i> , 1996			
fosfoenolpiruvato carboxilasa			Friedlander <i>et al.</i> , 1966	
Vitelogenina		Winter <i>et al.</i> , 1996		
Alcohol deshidrogenasa	Russo <i>et al.</i> , 1995; Nurminsky <i>et al.</i> , 1996			Yokoyama y Harry, 1993
Apoproteína del complejo de la dinitrogenasa (nifD)	Cournoyer <i>et al.</i> , 1993			
"IRBP"			Smith <i>et al.</i> , 1996	
Esterasa-5	Babcock y Anderson, 1996			
Citocromo c				Zhang y Chinnappa, 1994

se debe a su abundancia en el genoma y a su organización y estructura. La abundancia de estos genes facilita su aislamiento lo que permite comparaciones filogenéticas a distintos niveles taxonómicos. Mientras que su organización y estructura constituyen también una buena fuente de caracteres taxonómicos (Pace *et al.*, 1986). Además el ARN ribosomal se encuentra en todos los organismos y siempre ha tenido la misma función estructural dentro de los ribosomas (Chen *et al.*, 1986; Olsen *et al.*, 1986). Otra característica de los genes del ARN ribosomal es que sus distintas regiones tienen tasas de evolución diferente. Por ejemplo, las regiones que codifican para la subunidad corta y la subunidad larga de ARN ribosomal son potencialmente informativas en grupos taxonómicos de divergencia antigua; mientras que las regiones que no codifican, como el ITS 1 y 2, son útiles a niveles taxonómicos de divergencia reciente.

Actualmente se están incorporando en estudios taxonómicos las secuencias de otros genes nucleares. Estos genes codifican para enzimas como la fosfoglucoisomerasa (PgiC), la fosfoenolpiruvato carboxilasa, la alfa-amilasa, algunas globinas y la alcohol deshidrogenasa, entre otros (tabla 1). Además, se están evaluando las secuencias de otros genes nucleares que podrían ser útiles, entre ellos se encuentran los que codifican para enzimas como la enolasa, dopa-decarboxilasa y glucosa-6-fosfato deshidrogenasa (Friedlander *et al.*, 1992; Friedlander *et al.*, 1994), la enzima calmodulina, la proteína nucleolar NO38 y la proteína C3 del sistema inmune, algunos oncogenes y factores del crecimiento (Graybeal, 1994).

Genes mitocondriales. Los genes mitocondriales más ampliamente usados son los que codifican para el ARN ribosomal de mitocondrias (Miyamoto *et al.*, 1990; Milinkovitch *et al.*, 1993). También se han usado otros genes como los que forman parte de la cadena respiratoria. Entre ellos destacan los que codifican para las subunidades de la enzima citocromo oxidasa y el citocromo b. Además, existen diversos trabajos en los que se han utilizado los genes que codifican para los distintos ARNs de transferencia, para las subunidades del complejo de la NADH deshidrogenasa y excepcionalmente algunas regiones intergénicas no codificadoras (tabla 2).

El uso de las secuencias génicas del ADN mitocondrial ha tenido su mayor impacto en el estudio de la filogenia y la genética de poblaciones de hongos, algunos protistas y sobre todo de animales (Avise *et al.*, 1987; Harrison, 1989). En estos últimos, las secuencias de genes mitocondriales han servido para inferir relaciones filogenéticas a nivel de especie (Ammerman y Hillis, 1992; Boyce *et al.*, 1994; Brown *et al.*,

1994). El ADN mitocondrial de animales presenta características prácticas que se han aprovechado para estudios filogenéticos. Tiene un patrón hereditario estrictamente materno, es fácilmente aislado de la célula y su organización y contenido genético ya se conocen. Por ejemplo, el ADN mitocondrial contiene 2 genes de ARN ribosomal, 22 genes de ARN de transferencia, 13 genes para proteínas que forman parte de las enzimas del transporte de electrones o de la síntesis de ATP. Estos genes no presentan intrones y las secuencias intergénicas son pequeñas o ausentes (Moritz *et al.*, 1987). Además, el ADN mitocondrial contiene secuencias génicas con diferentes tasas evolutivas y un arreglo de genes variable entre taxa. Estas características han sido de utilidad para establecer las relaciones filogenéticas entre primates (Aquadro y Greenberg, 1983; Hixson y Brown, 1986) y otros grupos de animales (por ejemplo, Helm-Bychowski y Cracraft, 1993; Sturmhuber y Meyer, 1993; Van Den Bussche *et al.*, 1993; Lamb *et al.*, 1994; Milinkovitch *et al.*, 1994).

En plantas, por el contrario, el ADN mitocondrial ha sido pobremente estudiado. La poca información publicada se refiere casi exclusivamente a angiospermas (Palmer, 1992). Se ha observado, que la estructura y dinámica evolutiva del ADN mitocondrial de plantas es distinta a la de otros organismos. Por ejemplo, la mayoría de los aspectos de la estructura genómica, como son, el tamaño, la configuración y el orden génico cambian más rápidamente en plantas que en animales, mientras que la secuencia primaria de nucleótidos cambia excepcionalmente lento (Palmer y Herbon 1988). Debido a este patrón evolutivo particular, las secuencias del ADN mitocondrial de plantas no han sido útiles para estudios filogenéticos en grupos taxonómicos de reciente divergencia. Por otro lado, aunque estos genes potencialmente podrían servir en grupos taxonómicos de divergencia antigua, casi no se han usado. Esto se debe a que el ADN mitocondrial en plantas es más difícil de extraer, es menos abundante y contiene un menor número de genes para comparar entre taxa.

Genes del cloroplasto. El ADN de cloroplasto contiene varios genes potencialmente útiles para estudios filogenéticos y que no han sido explorados. Por ejemplo, tiene aproximadamente 30 genes que codifican para ARN de transferencia, 55 para proteínas de función conocida en la fotosíntesis y 30 para proteínas no identificadas. Una ventaja del genoma del cloroplasto para estudios filogenéticos es que es pequeño y fácilmente aislable de la célula.

El gen de cloroplasto más ampliamente usado para estudios taxonómicos es el gen codificador de la subu-

nidad larga de la enzima ribulosa 1,5-bifosfato carboxilasa (*rbcL*). La secuencia completa del gen *rbcL* ya se conoce en muchos taxa y ha mostrado variación útil para análisis filogenéticos a varios niveles taxonómicos (Soltis *et al.*, 1990; Donoghue *et al.*, 1992; Gianasi *et al.*, 1992; Olmstead *et al.*, 1992; Chase *et al.*, 1993). Una ventaja que se tiene al secuenciar el gen *rbcL* es que existen oligonucleótidos iniciadores específicos cuyas secuencias están disponibles en la literatura. Esto facilita la secuenciación de este gen en nuevos grupos taxonómicos (Zurawski y Clegg, 1987).

Actualmente, también se están usando otros genes del ADN de cloroplasto para análisis filogenéticos en plantas como son los genes del ARN ribosomal, el gen "matK", el gen "rps" para proteínas ribosomales, el gen "trnK-psbA" del fotosistema, los genes "rpo" para las subunidades de la ARN polimerasa y los genes para las subunidades de la NADH deshidrogenasa. También se ha evaluado la utilidad de las regiones intergénicas en análisis filogenéticos (tabla 3).

Conclusiones

La tendencia reciente en estudios taxonómicos es el uso combinado de datos morfológicos y moleculares especialmente los que provienen de enzimas de restricción y secuencias de genes particulares. Los continuos desarrollos tecnológicos están haciendo posible el muestreo de un número mayor de genes y taxa a comparar. Paralelamente, el desarrollo de métodos analíticos basados en la teoría cladística ha provisto la formalidad y robustez para interpretar la variación de esos datos. La influencia de estos dos desarrollos en México aún es incipiente, pero las perspectivas para la realización de este tipo de estudios son optimistas.

Los estudios taxonómicos basados en secuencias del ADN por ahora están limitados a muestrear sólo una región pequeña del genoma, usualmente un gen, un intrón, una secuencia espaciadora o regiones intergénicas. El uso de la amplificación enzimática mediante la RCP ha facilitado la obtención de ADN para los

Tabla 2. Genes del cloroplasto usados en estudios taxonómicos y nivel taxonómico cubierto.

	Categoría Taxonómica			
	≤ género	≤ familia	≤ orden	> orden
ARN ribosomal	Murphy y Collier, 1996 (12S); Crandall y Fitzpatrick, 1996 (16S)	Hay <i>et al.</i> , 1995 (12S, 16S)	Alves-Gomes <i>et al.</i> , 1995 (12S, 16S); Frye y Hedges, 1995 (12S, 16S)	Heise <i>et al.</i> , 1995 (12S, 16S)
Subunidades de la enzima citocromo oxidasa	Funk <i>et al.</i> , 1995; Sunnucks y Hales, 1996		Hasegawa y Adachi, 1996	Ruvolo <i>et al.</i> , 1993; Xu <i>et al.</i> , 1996
"D-loop"	Lyrholm <i>et al.</i> , 1996			
Citocromo b	Murphy y Collier, 1996; Thorpe <i>et al.</i> , 1996	Friesen <i>et al.</i> , 1996; Matthee y Robinson, 1997	Yoder <i>et al.</i> , 1996; Zardoya y Meyer, 1996	Martin y Palumbi, 1993
ARNs de transferencia	Arévalo <i>et al.</i> , 1994	Sites <i>et al.</i> , 1996	Frye y Hedges, 1995	
Subunidades de la NADH deshidrogenasa	Arévalo <i>et al.</i> , 1994	Sites <i>et al.</i> , 1996	Weller <i>et al.</i> , 1994; Zardoya y Meyer, 1996	Xu <i>et al.</i> , 1996
Regiones intergénicas	Sang <i>et al.</i> , 1994; Stewart y Baker, 1994			

muestreos de secuencias. Este procedimiento ha permitido el aislamiento selectivo de una región del genoma eliminando la necesidad de clonar tal región mediante bacterias. Los dos métodos disponibles de secuenciación, el químico y el enzimático, toman ventaja de la amplificación enzimática. El método más usado es el enzimático porque es más rápido ya que se eliminan reacciones químicas de metilación, protonación y sustitución. La introducción de los secuenciadores automatizados además de la rapidez agregan la ventaja de usar marcadores fluorescentes en vez de los radiactivos.

La abundante información que potencialmente puede derivarse del estudio de la variación del ADN debe analizarse con métodos formales. El análisis filogenético de las secuencias presenta procedimientos metodológicos especiales tanto en la fase de análisis de caracteres como en la búsqueda de cladogramas y su evaluación. El alineamiento y la codificación son aspectos fundamentales ya que equivalen conceptualmente a la elaboración de hipótesis de homología táxica o transformacional. La búsqueda de árboles mediante matrices de distancias es la más sencilla pero

agrupa los taxa por su similitud total y no por sus relaciones filogenéticas. Además, el convertir las secuencias en distancias ocasiona la pérdida de información y entonces los árboles de distancia no permiten evaluar el papel de cada carácter en la formación de los grupos.

En cambio, la búsqueda basada en máxima verosimilitud o en parsimonia permite agrupar los taxa por sus relaciones filogenéticas. La diferencia en estos dos enfoques es cómo se concibe el problema de la inferencia de un evento histórico único como la filogenia. Los métodos de máxima verosimilitud presuponen que la inferencia filogenética es un problema inductivo de estimación estadística. En contraste, los métodos de parsimonia implican que la inferencia filogenética es un problema de análisis hipotético-deductivo. Todavía existe una controversia muy activa en cuanto a cuál de estas dos posiciones epistemológicas es la más robusta para la reconstrucción filogenética (Harper, 1979; Felsenstein y Sober, 1986; Felsenstein, 1988; Sanderson, 1989; Cracraft y Helm-Bychowski, 1991; Carpenter, 1992; Trueman, 1993). Sin embargo, la decisión de cuál metodología apli-

Tabla 3. Genes del cloroplasto usados en estudios taxonómicos y nivel taxonómico cubierto.

	Categoría Taxonómica			
	≤ género	≤ familia	≤ orden	> orden*
rbcl		Brunsfeld <i>et al.</i> , 1994; Bayer <i>et al.</i> , 1996	Goremykin <i>et al.</i> , 1996	
matK		Johnson y Soltis, 1994; Steele y Vilgalys, 1994		
rps			Nadot <i>et al.</i> , 1994	
rpo		Cummings <i>et al.</i> , 1994; Downie <i>et al.</i> , 1996		
Subunidades de la NADH deshidrogenasa		Olmstead y Sweere, 1994		
Regiones intergénicas	Gielly y Taberlet, 1994			

* La ausencia de referencias en esta columna no indica que no las haya, sino que no se publicaron en las revistas ni durante los años consultados (ejemplo Chase *et al.*, 1993 citada en el texto).

car parece guiada no por argumentos filosóficos sino por los límites operativos. Las dificultades del cómputo de los valores de verosimilitud de los árboles limitan el número de taxa a comparar. Este límite metodológico es una desventaja para el uso de métodos de máxima verosimilitud. Dejando las razones filosóficas a un lado, los métodos basados en parsimonia son los únicos viables, por ahora, para el análisis filogenético de secuencias.

Agradecimientos

Agradezco a Efraín De Luna, a dos revisores anónimos y a Ken Oyama sus valiosas sugerencias y críticas al presente manuscrito. Este trabajo se realizó gracias al apoyo económico del Instituto de Ecología, A. C., mediante la cuenta 902-14 y al proyecto PACIME-CONACyT No. 1837P-N9507.

Literatura citada

- Alves-Gomes J. A., Orti G., Haygood M. y Heiligenberg W. 1995. Phylogenetic analysis of the South American electric fishes (Order Gymnotiformes) and the evolution of their electrogenic system: A synthesis based on morphology, electrophysiology, and mitochondrial sequence data. *Molecular Biology and Evolution* 12:298-318.
- Ammerman L. K. y Hillis D. M. 1992. A molecular test of bat relationships: Monophyly or diphyly? *Systematic Biology* 41:222-232.
- Aquadro C. F. y Greenberg B. D. 1983. Human mitochondrial DNA variation and evolution: Analysis of nucleotide sequences from seven individuals. *Genetics* 103:287-312.
- Archie J. W. 1989. Homoplasy excess ratios: New indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Systematic Zoology* 38:253-269.
- Arévalo E., Davis S. K. y Sites J. W., Jr. 1994. Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. *Systematic Biology* 43:387-418.
- Arias D. M. y Rieseberg L. H. 1995. Genetic relationships among domesticated and wild sunflowers (*Helianthus annuus*, Asteraceae). *Economic Botany* 49:239-248.
- Avise J. C., Arnold J., Ball R. M., Bermingham E., Lamb T., Neigel J. E., Reeb C. A. y Saunders N. C. 1987. Intraspecific phylogeography: The mitochondrial bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18:489-522.
- Babcock C. S. y Anderson W. W. 1996. Molecular evolution of the sex-ratio inversion complex in *Drosophila pseudoobscura*: Analysis of the esterase-5 gene region. *Molecular Biology and Evolution* 13:297-308.
- Barrera H. A., Ortiz R., Rojas A. y Reséndez D. 1993. Reacción en cadena de la polimerasa. *Ciencia y Desarrollo* 18:50-60.
- Bayer R. J., Hufford L. y Soltis D. E. 1996. Phylogenetic relationships in Sarraceniaceae based on *rbcL* and ITS sequences. *Systematic Botany* 21:121-134.
- Bisby F. A., Vaughn J. G. y Wright C. A., eds. 1980. *Chemosystematics: principles and practices*. Academic Press, New York.
- Boyce T. M., Zwick M. E. y Aquadro C. F. 1994. Mitochondrial DNA in the bark weevils: Phylogeny and evolution in the *Pissodes strobi* species group (Coleoptera: Curculionidae). *Molecular Biology and Evolution* 11:183-194.
- Bremer K. 1994. Branch support and tree stability. *Cladistics* 10:295-304.
- Brown J. M., Pellmyr O., Thompson J. N. y Harrison R. G. 1994. Phylogeny of *Greya* (Lepidoptera: Prodoxidae), based on nucleotide sequence variation in mitochondrial cytochrome oxidase I and II: Congruence with morphological data. *Molecular Biology and Evolution* 11:128-141.
- Brown W. M., Prager E. M., Wang A. y Wilson A. C. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *Journal of Molecular Evolution* 18:225-239.
- Bruns T. D., White T. J. y Taylor J. W. 1991. Fungal molecular systematics. *Annual Review of Ecology and Systematics* 22:525-564.
- Bruns T. D., Vilgalys R., Barns S. M., González D., Hibbet D. S., Lane D. J., Simon L., Stickel S., Szaro T. M., Weisburg W. G. y Sogin M. L. 1992. Evolutionary relationships within the fungi: Analyses of nuclear small subunit rRNA sequences. *Molecular Phylogenetics and Evolution* 1:231-241.
- Brunsfeld S. J., Soltis P. S., Soltis D. E., Gadek P. A., Quinn C. J., Strenge D. D. y Ranker T. A. 1994. Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae: Evidence from *rbcL* sequences. *Systematic Botany* 19: 253-262.
- Buth D. G. 1984. The application of electrophoretic data in systematic studies. *Annual Review of Ecology and Systematics* 15:501-522.
- Camin J. H. y Sokal R. R. 1965. A method for deducing branching sequences in phylogeny. *Evolution* 19:311-326.
- Campbell C. S., Wojciechowski M. F., Baldwin B. G., Alice L. A. y Donoghue M. J. 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Molecular Biology and Evolution* 14:81-90.
- Carpenter J. M. 1992. Random cladistics. *Cladistics* 8:147-153.
- Chase G. D. 1959. *Principles of radioisotope methodology*. Burgess Publishing Company, Minneapolis, Minnesota.
- Chase M. W., Soltis D. E., Olmstead R. G., Morgan D., Les D. H., Mishler B. D., Duvall M. R., Price R. A., Hills H. G., Qui Y. L., Kron K. A., Rettig J. H., Conti E., Palmer J. D., Manhart J. R., Sytsma K. J., Michaels H. J., Kress

- W. J., Karol K. G., Clark W. D., Hedren M., Gaut B. S., Jansen R. K., Kim K. J., Wimpee C. F., Smith J. F., Furnier G. R., Strauss S. H., Xiang Q. Y., Plunkett G. M., Soltis P. S., Swensen S. M., Williams S. E., Gadek P. A., Quinn C. J., Eguiarte L. E., Golenberg E., Learn G. H., Graham S. W., 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.
- Chaw S. M., Zharkikh A., Sung H. M., Lau T. C. y Li W. H. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molecular Biology and Evolution* **14**:56-68.
- Chen I., Dixit A., Rhoads D. D. y Roufa D. J. 1986. Homologous ribosomal proteins in bacteria, yeast, and humans. *Proceedings of the National Academy of Sciences USA* **83**:6907-6911.
- Cho S., Mitchell A., Regier J. C., Mitter C., Poole R. W., Friedlander T. P. y Zhao S. 1995. A highly conserved nuclear gene for low-level phylogenetics: Elongation factor-1-alpha recovers morphology-based tree for heliothine moths. *Molecular Biology and Evolution* **12**:650-656.
- Clark A. G., Leicht B. G. y Muse S. V. 1996. Length variation and secondary structure of introns in the Mlc 1 gene in six species of *Drosophila*. *Molecular Biology and Evolution* **13**:471-482.
- Cournoyer B., Gouy M. y Normand P. 1993. Molecular phylogeny of the symbiotic Actinomycetes of the genus *Frankia* matches host plant infection processes. *Molecular Biology and Evolution* **10**:1303-1316.
- Cracraft J. y Helm-Bychowski K. 1991. Parsimony and phylogenetic inference using DNA sequences: some methodological strategies. En: Miyamoto M. M. y Cracraft J., eds. *Phylogenetic analysis of DNA sequences*. Oxford University Press, New York.
- Crandall K. A. y Fitzpatrick J. F., Jr. 1996. Crayfish molecular systematics: Using a combination of procedures to estimate phylogeny. *Systematic Biology* **45**:1-26.
- Cummings M. P., King L. M. y Kellogg E. A. 1994. Slipped-strand mispairing in a plastid gene: *rpoC2* in grasses (Poaceae). *Molecular Biology and Evolution* **11**:1-8.
- Davis J. I. 1993. Character removal as a means for assessing stability of clades. *Cladistics* **9**:201-210.
- De Luna E. 1995. Bases filosóficas de los análisis cladísticos para la investigación taxonómica. *Acta Botánica Mexicana* **33**:63-79.
- Delgado-Salinas A., Bruneau A. y Doyle J. J. 1993. Chloroplast DNA phylogenetic studies in New World Phaseolinae (Leguminosae: Papilionoideae: Phaseoleae). *Systematic Botany* **18**:6-17.
- Donoghue M. J., Olmstead R. G., Smith J. F. y Palmer J. D. 1992. Phylogenetic relationship of Dipsacales based on *rbcl* sequences. *Annals of the Missouri Botanical Garden* **79**:333-345.
- Dorado O., Rieseberg L. H. y Arias D. M. 1992. Chloroplast DNA introgression in Southern California sunflowers. *Evolution* **46**:566-572.
- Doyle J. J. 1993. DNA, phylogeny, and the flowering of plant systematics. *BioScience* **43**:380-389.
- Downie S. R., Llanas E. y Katz-Downie D. S. 1996. Multiple independent losses of the *rpoC1* intron in angiosperm chloroplast DNA's. *Systematic Botany* **21**:135-151.
- 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. S. 1970. Methods for computing Wagner trees. *Systematic Zoology* **34**:21-34.
- Farris J. S. 1977. Phylogenetic analysis under Dollo's law. *Systematic Zoology* **26**:77-88.
- Farris J. S. 1989. The retention index and the rescaled consistency index. *Cladistics* **5**:417-419.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* **17**:368-376.
- Felsenstein J. 1984. The statistical approach to inferring evolutionary trees and what it tells us about parsimony and compatibility. En: Duncan T. y Stuessy T. F., eds. *Cladistics: Perspectives on the reconstruction of evolutionary history*. Columbia University Press, New York.
- Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791.
- Felsenstein J. 1988. Phylogenies from molecular sequences: Inference and reliability. *Annual Review of Genetics* **22**:521-565.
- Felsenstein J. y Sober E. 1986. Parsimony and likelihood: An exchange. *Systematic Zoology* **35**:617-626.
- Ferguson A. 1980. *Biochemical systematics and evolution*. J. Wiley & Sons, New York.
- Ferl R. J., Nair C. J., She J.-X., Wakeland E. y Almira E. 1991. The application of automated DNA sequence analysis to phylogenetic studies. En: Miyamoto M. M. y Cracraft J., eds. *Phylogenetic analysis of DNA sequences*. Oxford University Press, New York.
- Field K. G., Olsen G. J., Lane D. J., Giovannoni S. J., Ghiselin M. T., Raff E. C., Pace N. R. y Raff R. A. 1988. Molecular phylogeny of the animal kingdom. *Science* **239**:748-753.
- Fitch W. M. 1971. Toward defining the course of evolution: minimal change for a specific tree topology. *Systematic Zoology* **20**:406-416.
- Ford V. S., Thomas B. R. y Gottlieb L. D. 1995. The same duplication accounts for the *PgiC* genes in *Clarkia xantiana* and *C. lewisii* (Onagraceae). *Systematic Botany* **20**:147-160.
- Friedlander T. P., Regier J. C. y Mitter C. 1992. Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. *Systematic Biology* **41**: 483-490.

- Friedlander T. P., Regier J. C. y Mitter C. 1994. Phylogenetic information content of five nuclear gene sequences in animals: Initial assessment of character sets from concordance and divergence studies. *Systematic Biology* **43**:511-525.
- Friedlander T. P., Regier J. C., Mitter C. y Wagner D. L. 1996. A nuclear gene for higher level phylogenetics: phosphoenolpyruvate carboxykinase tracks mesozoic-age divergences within Lepidoptera (Insecta). *Molecular Biology and Evolution* **13**:594-604.
- Friesen V. L., Baker A. J. y Piatt J. F. 1996. Phylogenetic relationships within the Alcidae (Charadriiformes: Aves) inferred from total molecular evidence. *Molecular Biology and Evolution* **13**:359-367.
- Frye M. S. y Hedges S. B. 1995. Monophyly of the order rodentia inferred from mitochondrial DNA sequences of the genes for 12S rRNA, 16S rRNA, and tRNA-Valine. *Molecular Biology and Evolution* **12**:168-176.
- Funk D. J., Futuyma D. J., Ortí G. y Meyer A. 1995. Mitochondrial DNA sequences and multiple data sets: A phylogenetic study of phytophagous beetles (Chrysomelidae: Ophraella). *Molecular Biology and Evolution* **12**: 627-640.
- Gatesy J., Hayashi C., Cronin M. A. y Arctander P. 1996. Evidence from milk casein genes that cetaceans are close relatives of hippopotamid Artiodactyls. *Molecular Biology and Evolution* **13**:954-963.
- Giannasi D. E., Zurawski G., Learn G. y Clegg M. T. 1992. Evolutionary relationships of the Caryophyllidae based on comparative *rbL* sequences. *Systematic Botany* **17**:1-15.
- Gielly L. y Taberlet P. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbL* sequences. *Molecular Biology and Evolution* **11**:769-777.
- González D. 1992. Classification of the plant pathogenic fungus *Rhizoctonia solani* (Basidiomycotina: Tullasnelales) using ribosomal DNA sequence data. MS Thesis. Duke University. Durham, North Carolina.
- Goodman M., Czelusniak J. y Beeber J. E. 1985. Phylogeny of primates and other eutherian orders: A cladistic analysis using amino acid and nucleotide sequence data. *Cladistics* **1**:171-185.
- Goremykin V., Bobrova V., Pahnke J., Troitsky A., Antonov A. y Martin W. 1996. Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbL* data do not support gnetalean affinities of angiosperms. *Molecular Biology and Evolution* **13**:383-396.
- Gottlieb L. D. y Ford V. S. 1996. Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequences of PgiC. *Systematic Botany* **21**:45-62.
- Grant W. F., edr. 1984. *Plant biosystematics*. Academic Press, London.
- Graybeal A. 1994. Evaluating the phylogenetic utility of genes: A search for genes informative about deep divergences among vertebrates. *Systematic Biology* **43**: 174-193.
- Hamby R. K. y Zimmer E. A. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. En: Soltis P. S., Soltis D. E. y Doyle J. J., eds. *Molecular Systematics of Plants*. Chapman and Hall, New York.
- Harper C. W. 1979. A Bayesian probability view of phylogenetic systematics. *Systematic Zoology* **28**:547-553.
- Harrison R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution* **4**:6-11.
- Hasegawa M. y Adachi J. 1996. Phylogenetic position of cetacean relative to artiodactyls: Reanalysis of mitochondrial and nuclear sequences. *Molecular Biology and Evolution* **13**:710-717.
- Hay J. M., Ruvinsky I., Hedges S. B. y Maxon L. R. 1995. Phylogenetic relationships of amphibian families inferred from DNA sequences of mitochondrial 12S and 16S ribosomal RNA genes. *Molecular Biology and Evolution* **12**:928-937.
- Heise P. J., Maxson L. R., Dowling H. G. y Hedges S. B. 1995. Higher-level snake phylogeny inferred from mitochondrial DNA sequences of 12S rRNA and 16S rRNA genes. *Molecular Biology and Evolution* **12**:259-265.
- Helm-Bychowski K. y Cracraft J. 1993. Recovering phylogenetic signal from DNA sequences: Relationships within the corvine assemblage (Class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome-b gene. *Molecular Biology and Evolution* **10**:1196-1214.
- Higuchi R. G. y Ochman H. 1989. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. *Nucleic Acids Research* **17**:5865.
- Hillis D. M. y Dixon M. T. 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. *The Quarterly Review of Biology* **66**:411-453.
- Hillis D. M., Larson A., Davis S. K y Zimmer E. A. 1990. Nucleic acids III: Sequencing. En: Hillis D. M. y Moritz C., eds. *Molecular Systematics*. Sinauer, Sunderland, Ma.
- Hillis D. M. y Moritz C., eds. 1990. *Molecular Systematics*. Sinauer, Sunderland, MA.
- Hixson J. E. y Brown W. M. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: Sequence, structure, evolution and phylogenetic implications. *Molecular Biology and Evolution* **3**:1-18.
- Johnson L. A. y Soltis D. E. 1994. matK DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany* **19**:143-156.
- Keeling P. J. y Doolittle W. F. 1996. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Molecular Biology and Evolution* **13**:1297-1305.
- Knight A. y Mindell D. P. 1993. Substitution bias, weighting of DNA, sequence evolution, and the phylogenetic position of fea's viper. *Systematic Biology* **42**:18-31.

- Lafay B., Smith A. B. y Christen R. 1995. A combined morphological and molecular approach to the phylogeny of asteroids (Asteroidea: Echinodermata). *Systematic Biology* **44**:190-208.
- Lake J. A. 1987. A rate-independent technique for analysis of nucleic acid sequences: Evolutionary parsimony. *Molecular Biology and Evolution* **4**:167-191.
- Lamb T., Lydeard C., Walker R. B. y Gibbons J. W. 1994. Molecular systematics of map turtles (*Graptemys*): A comparison of mitochondrial restriction site versus sequence data. *Systematic Biology* **43**:543-559.
- Lawson F. S., Charlebois R. L. y Dillon J. A. R. 1996. Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *Molecular Biology and Evolution* **13**:970-977.
- Li W.-H. y Gouy M. 1991. Statistical methods for testing molecular phylogenies. En: Miyamoto M. M y Cracraft J., eds. *Phylogenetic analysis of DNA sequences*. Oxford University Press, New York.
- Li W.-H., Wu C.-I. y Luo C.-C. 1984. Nonrandomness of point mutations as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *Journal of Molecular Evolution* **21**:58-71.
- Lyrholm T., Leimar O. y Gyllenstein U. 1996. Low diversity and biased substitution patterns in the mitochondrial DNA control region of sperm whales: implications for estimates of time since common ancestry. *Molecular Biology and Evolution* **13**:1318-1326.
- Martin A. P. y Palumbi S. R. 1993. Protein evolution in different cellular environments: cytochrome b in sharks and mammals. *Molecular Biology and Evolution* **10**: 873-891.
- Matthee C. A. y Robinson T. J. 1997. Molecular phylogeny of the springhare, *Pedetes capensis*, based on mitochondrial DNA sequences. *Molecular Biology and Evolution* **14**:20-29.
- Mathews S. y Sharrock R. A. 1996. The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. *Molecular Biology and Evolution* **13**:1141-1150.
- Maxam A. M. y Gilbert W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences USA* **74**:560-564.
- Maxam A. M. y Gilbert W. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods in Enzymology* **65**:499-559.
- McCabe P. C. 1990. Production of single-stranded DNA by asymmetric PCR. En: Innis M. A, Gelfand D. H., Sninsky J. J. y White T. J., eds. *PCR Protocols: A guide to methods and applications*. Academic Press, New York.
- Milinkovitch M. C., Orti G. y Meyer A. 1993. Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences. *Nature* **361**:346-348.
- Milinkovitch M. C., Meyer A. y Powell J. R. 1994. Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. *Molecular Biology and Evolution* **11**:939-948.
- Mitchell L. G. y Merril C. R. 1989. Affinity generation of single-stranded DNA for dideoxy sequencing following the polymerase chain reaction. *Analytical Biochemistry* **178**:239-242.
- Miyamoto M. M., Kraus F. y Ryder O. A. 1990. Phylogeny and evolution of antlered deer determined from mitochondrial DNA sequences. *Proceedings of the National Academy of Sciences USA* **87**:6127-6131.
- Moritz C., Dowling T. E. y Brown W. M. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18**:269-292.
- Mullis K. B. y Faloona F. A. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* **155**:335-350.
- Murphy W. J. y Collier G. E. 1996. Phylogenetic relationships within the aplocheiloid fish genus *Rivulus* (Cyprinodontiformes, Rivulidae): Implications for Caribbean and Central American biogeography. *Molecular Biology and Evolution* **13**:642-649.
- Nadot S., Bajon R. y Lejeune B. 1994. The chloroplast gene rps 4 as a tool for the study of Poaceae phylogeny. *Plant Systematics and Evolution* **191**:27-38.
- Needleman S. B. y Wunsch C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**:443-453.
- Nei M. 1991. Relative efficiencies of different tree-making methods for molecular data. En: Miyamoto M. M. y Cracraft J., eds. *Phylogenetic analysis of DNA sequences*. Oxford University Press, New York.
- Nei M., Stephens J. C. y Saitou N. 1985. Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. *Molecular Biology and Evolution* **2**:66-85.
- Nurminsky D. I., Moriyama E. N., Lozovskaya E. R. y Hartl D. L. 1996. Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: Duplications of the alcohol dehydrogenase gene. *Molecular Biology and Evolution* **13**:132-149.
- Okuyama E., Shibata H., Tachida H. y Yamazaki T. 1996. Molecular evolution of the 5'-flanking regions of the duplicated *Amy* genes in *Drosophila melanogaster* species subgroup. *Molecular Biology and Evolution* **13**:574-583.
- Olmstead R. G., Michaels H. J., Scott K. M. y Palmer J. D. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcL*. *Annals of the Missouri Botanical Garden* **79**:249-265.
- Olmstead R. G. y Sweere J. A. 1994. Combining data in phylogenetic systematics: An empirical approach using

- three molecular data sets in the Solanaceae. *Systematic Biology* **43**:467-481.
- Olsen G. J., Lane D. J., Giovannoni S. J. y Pace N. R. 1986. Microbial ecology and evolution: A ribosomal RNA approach. *Annual Review of Microbiology* **40**:337-365.
- Ortí G. y Meyer A. 1996. Molecular evolution of ependymin and the phylogenetic resolution of early divergences among euteleost fishes. *Molecular Biology and Evolution* **13**:556-573.
- Pace N. R., Olsen G. J. y Woese C. R. 1986. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**:325-326.
- Palmer J. D. 1992. Mitochondrial DNA in plant systematics: applications and limitations. En: Soltis PS, Soltis D. E. y Doyle J. J., eds. *Molecular systematics of plants*. Chapman & Hall, New York.
- Palmer J. D. y Herbon L. A. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* **28**:87-97.
- Patterson C. 1987. *Molecules and morphology in evolution: Conflict or compromise?* Cambridge University Press, Cambridge.
- Penny D. y Hendy M. D. 1986. Estimating the reliability of evolutionary trees. *Molecular Biology and Evolution* **3**:403-417.
- Ritland K. y Clegg M. T. 1987. Evolutionary analysis of plant DNA sequences. *The American Naturalist* **130**:S74-S100.
- Russo C. A. M., Takezaki N. y Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. *Molecular Biology and Evolution* **12**:391-404.
- Ruvolo M., Zehr S., vonDornum M., Pan D., Chang B. y Lin J. 1993. Mitochondrial COII sequences and modern human origins. *Molecular Biology and Evolution* **10**:1115-1135.
- Saiki R. K., Gelfand D. H., Stoffel S., Scharf S. J., Higuchi R., Horn G. T., Mullis K. B. y Erlich H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**:487-491.
- Sanderson M. J. 1989. Confidence limits on phylogenies: The bootstrap revisited. *Cladistics* **5**:113-129.
- Sang T.-K., Chang H.-Y., Chen C.-T. y Hui C.-F. 1994. Population structure of the Japanese Eel, *Anguilla japonica*. *Molecular Biology and Evolution* **11**:250-260.
- Sanger F., Nicklen S. y Coulson A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**:5463-5467.
- Sites J. W. J., Davis S. K., Guerra T., Iverson J. B y Snell H. L. 1996. Character congruence and phylogenetic signal in molecular and morphological data set: a case study in the living iguanas (Squamata, Iguanidae). *Molecular Biology and Evolution* **13**:1087-1105.
- Slade R. W., Moritz C. y Heideman A. 1994. Multiple nuclear-gene phylogenies: Application to Pinnipeds and comparison with a mitochondrial DNA gene phylogeny. *Molecular Biology and Evolution* **11**:341-356.
- Smith M. R., Shivji M. S., Waddell V. G. y Stanhope M. J. 1996. Phylogenetic evidence from the IRBP gene for the paraphyly of toothed whales, with mixed support for Cetacea as a suborder of Artiodactyla. *Molecular Biology and Evolution* **13**:918-922.
- Soltis D. E., Soltis P. S., Clegg M. T. y Durbin M. 1990. *rbL* sequence divergence and phylogenetic relationships in Saxifragaceae *sensu lato*. *Proceedings of the National Academy of Sciences USA* **87**:4640-4644.
- Soltis P. S., Soltis D. E. y Doyle J., eds. 1992. *Molecular Systematics of plants*. Chapman & Hall, New York.
- Soltis D. E., Johnson L. A. y Looney C. 1996. Discordance between ITS and chloroplast topologies in the *Boykinia* group (Saxifragaceae). *Systematic Botany* **21**:169-185.
- Steele K. P. y Vilgalys R. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene matK. *Systematic Botany* **19**:126-142.
- Stevens P. F. 1984. Methaphors and typology in the development of botanical systematics 1690-1960, or the art of putting new wine in old bottles. *Taxon* **33**:169-211.
- Stevens P. F. 1986. Evolutionary classifications in botany, 1960-1985. *Journal of the Arnold Arboretum* **67**:313-339.
- Stewart D. T. y Baker A. J. 1994. Patterns of sequence variation in the mitochondrial D-loop region of shrews. *Molecular Biology and Evolution* **11**:9-21.
- Sturmbauer C. y Meyer A. 1993. Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fishes from lake Tanganyika in Eastern Africa. *Molecular Biology and Evolution* **10**:751-768.
- Sunnucks P. y Hales D. F. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution* **13**:510-524.
- Swofford D. L. y Olsen G. J. 1990. Phylogeny reconstruction. En: Hillis D. M. y Moritz C., eds. *Molecular Systematics*. Sinauer, Sunderland, MA.
- Thorpe R. S., Black H. y Malhotra A. 1996. Matrix correspondence tests on the DNA phylogeny of the Tenerife lacertid elucidate both historical causes and morphological adaptation. *Systematic Biology* **45**:335-343.
- Trueman J. W. H. 1993. Randomization confounded: A response to Carpenter. *Cladistics* **9**:101-109.
- Van Den Bussche R. A., Baker R. J., Wichman H. A. y Hamilton M. J. 1993. Molecular phylogenetics of Stenodermatini Bat genera: congruence of data from nuclear and mitochondrial DNA. *Molecular Biology and Evolution* **10**:944-959.
- Vawter L. y Brown W. M. 1993. Rates and patterns of base change in the small subunit ribosomal RNA gene. *Genetics* **134**:597-608.
- Vogler A. P., Welsh A y Hancock J. M. 1997. Phylogenetic analysis of slippage-like sequence variation in the V4 rRNA expansion segment in tiger beetles (Cicindelidae). *Molecular Biology and Evolution* **14**:6-19.

- Weller S. J., Pashley D. P., Martin J. A. y Constable J. L. 1994. Phylogeny of noctuid moths and the utility of combining independent nuclear and mitochondrial genes. *Systematic Biology* **43**:194-211.
- Wheeler W. C. 1990. Combinatorial weights in phylogenetic analysis: A statistical parsimony procedure. *Cladistics* **6**:269-275.
- Williams D. M. 1992. DNA analysis: methods. En: Forey P. L., Humphries C. J., Kitching I. L., Scotland R. W., Siebert D. J. y Williams D. M., eds. *Cladistics. A practical course in systematics*. Oxford University Press, New York.
- Williams P. L. y Fitch W. M. 1990. Phylogeny determination using dynamically weighted parsimony method. *Methods in Enzymology* **183**:615-626.
- Winnepenninckx B., Backeljau T. y DeWachter R. 1996. Investigation of molluscan phylogeny on the basis of 18S rRNA sequences. *Molecular Biology and Evolution* **13**:1306-1317.
- Winter C. E., Penha C. y Blumenthal T. 1996. Comparison of a vitellogenin gene between two distantly related Rhabditid nematode species. *Molecular Biology and Evolution* **13**:674-684.
- Xu X., Janke A. y Arnason U. 1996. The complete mitochondrial DNA sequence of the greater indian rhinoceros, *Rhinoceros unicornis*, and the phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+ Cetacea). *Molecular Biology and Evolution* **13**:1167-1173.
- Yoder A. D., Vilgalys R. y Ruvolo M. 1996. Molecular evolutionary dynamics of cytochrome b in Strepsirrhine primates: the phylogenetic significance of third-position transversions. *Molecular Biology and Evolution* **13**:1339-1350.
- Yokoyama S. y Harry D. E. 1993. Molecular phylogeny and evolutionary rates of alcohol dehydrogenases in vertebrates and plants. *Molecular Biology and Evolution* **10**:1214-1226.
- Zardoya R. y Meyer A. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution* **13**:933-942.
- Zhang X. H. y Chinnappa C. C. 1994. Molecular cloning of a cDNA encoding cytochrome c of *Stellaria longipes* (Caryophyllaceae)-and the evolutionary implications. *Molecular Biology and Evolution* **11**:365-375.
- Zurawski G. y Clegg M. T. 1987. Evolution of higher-plant chloroplast DNA-encoded genes: Implications for structure-function and phylogenetic studies. *Annual Review of Plant Physiology* **38**:391-418.