

5

INFERENCIA FILOGENÉTICA MEDIANTE SECUENCIAS DE DNA: UN EJEMPLO CON LOS PINOS PIÑONEROS

David S. Gernandt, Omar Zerón Flores e Irene Goyenechea

Introducción

La sistemática molecular pretende investigar mediante datos moleculares tanto las relaciones filogenéticas entre los organismos como los procesos evolutivos responsables de estas relaciones. Los primeros marcadores moleculares incluyeron lípidos, metabolitos secundarios, cariotipos, reacciones inmunogénicas, electroforesis de proteínas (isoenzimas) y electroforesis de fragmentos de DNA (análisis de RFLP). La invención de la técnica de reacción en cadena de la polimerasa (PCR) en la década de 1980 (Saiki *et al.*, 1985; Mullis, 1990) colocó al DNA como la fuente más importante de datos moleculares en estudios superiores al nivel de la especie (la reconstrucción de filogenias y no genética de poblaciones).

Lo que distingue a la sistemática molecular de la sistemática tradicional es, en esencia, la fuente de caracteres y no las bases filosóficas. Sin embargo, el uso de modelos explícitos de evolución para nucleótidos y aminoácidos, así como la adopción de métodos de inferencia que no se basan en la parsimonia, como la máxima verosimilitud y la inferencia bayesiana, han producido polémicas entre los investigadores que usan caracteres moleculares (secuencias de aminoácidos y de DNA) para inferencia filogenética y los que usan caracteres tradicionales (p. ej., morfológicos y ultraestructurales). Debido a que los caracteres tradicionales son determinados por los genes, el uso de secuencias de DNA se puede considerar una manera más directa de observar diferencias heredables entre especies. Sin embargo, los caracteres morfológicos son los relevantes en la descripción formal y el reconocimiento de taxones en la vida diaria. Es por eso que los taxónomos han reconocido el valor de tomar en cuenta datos tradicionales y moleculares juntos (Hillis y Wiens, 2000).

Los pinos piñoneros de América del Norte comprenden aproximadamente 12 especies de árboles de baja estatura o a veces arbustos distribuidas entre el oeste de los Estados Unidos hasta el centro y sur de México. Presentan semillas grandes, los piñones, que carecen de alas funcionales por estar adaptadas para la dispersión por vertebrados, principalmente córvidos y roedores, y no por el viento. Los pinos piñoneros han sido sujetos de varias clasificaciones (Cuadro 1). En este trabajo se usará como hipótesis inicial una clasificación basada en un análisis cladístico de caracteres tradicionales que fue presentado en una revisión reciente de los pinos neotropicales (Farjon y Styles, 1997).

De las 12 especies de pinos piñoneros, ocho son muy parecidas entre sí, particularmente con respecto al cono ovulado que no presenta una longitud de más de cinco centímetros, mientras las otras cuatro presentan conos más largos y morfologías divergentes, tanto en los conos como en la semilla, follaje, fascículos de las hojas y anatomía de la madera. Tradicionalmente se ha considerado que las ocho especies con conos pequeños conforman la subsección *Cembroides*, mientras que propuestas recientes han separado algunas de las especies con conos grandes. Malusa (1992), en un análisis cladístico con 20 caracteres morfológicos y ecológicos, dividió al grupo en dos, las ocho especies con conos pequeños y las cuatro con conos grandes. La filogenia por Farjon y Styles (1997) fue similar, pero ellos separaron dos de las especies con conos grandes, *Pinus nelsonii* y *P. pinceana*, en su propia subsección, *Nelsoniae*, y dejaron a las otras dos especies como inciertas. Otra clasificación del género, por Price *et al.* (1998), no se hizo con base en un análisis formal, sino que tomó en cuenta estudios preliminares y recientes de sistemática molecular (Liston *et al.*, 1999). Esta clasificación separó a

P. rzedowskii, el único pino piñonero que no presenta semillas grandes sin alas, en la subsección *Rzedowskianae* y agrupó a los otros pinos piñoneros en la subsección *Cembroides*. Estas dos subsecciones fueron unidas con la subsección *Balfourianae* en la sección *Parrya* del subgénero *Strobus* (Price *et al.*, 1998).

En años recientes, la filogenia de los pinos piñoneros ha sido inferida en una serie de estudios basados en secuencias de DNA. Gernandt *et al.* (2001) presentaron un análisis cladístico de la región del espaciador interno transcrito (ITS), el cual se localiza entre genes nucleares de rRNA. En este trabajo se encontró que la región del ITS en especies de pinos es tan heterogénea, que muchas veces, copias divergentes en individuos de la misma especie no forman un grupo monofilético, sino que se intercalan con las de otras especies de la

misma subsección. A pesar de esta limitación importante, fue posible inferir las relaciones entre los grupos principales, relaciones que no fueron congruentes con las clasificaciones previas de los pinos piñoneros. Las especies se dividieron en dos linajes diferentes: uno incluía las especies parecidas entre sí con conos pequeños, más tres de las especies divergentes, *P. pinceana*, *P. maximartinezii* y *P. rzedowskii*. El otro linaje sólo incluyó a individuos de *P. nelsonii*, con base en sus secuencias de ITS.

Dos estudios posteriores con secuencias de genes del cloroplasto (Gernandt *et al.*, 2003, 2005) corroboraron la separación de *P. nelsonii* de los otros pinos piñoneros. Con esta hipótesis filogenética, se propuso una clasificación que coloca a *P. nelsonii* en su propia subsección, *Nelsoniae*, y a los demás pinos piñoneros en la subsección *Cembroides*

Cuadro 1
Dos clasificaciones de los pinos piñoneros

| Farjon y Styles (1997) | Gernandt <i>et al.</i> (2005) |
|--|--|
| Subsección <i>Cembroides</i> | Subsección <i>Cembroides</i> |
| 1. <i>P. cembroides</i> | 1. <i>P. cembroides</i> |
| ssp. <i>cembroides</i> | ssp. <i>cembroides</i> |
| ssp. <i>cembroides</i> var. <i>bicolor</i> | ssp. <i>lagunae</i> |
| ssp. <i>lagunae</i> | ssp. <i>orizabensis</i> |
| ssp. <i>orizabensis</i> | 2. <i>P. culminicola</i> |
| 2. <i>P. culminicola</i> | 3. <i>P. discolor</i> |
| 3. <i>P. edulis</i> ¹ | 4. <i>P. edulis</i> |
| 4. <i>P. monophylla</i> | 5. <i>P. johannis</i> |
| 5. <i>P. quadrifolia</i> | 6. <i>P. maximartinezii</i> |
| 6. <i>P. remota</i> | 7. <i>P. monophylla</i> |
| | ssp. <i>californiarum</i> ² |
| Subsección <i>Nelsoniae</i> | ssp. <i>fallax</i> ² |
| 7. <i>P. nelsonii</i> | ssp. <i>monophylla</i> ² |
| 8. <i>P. pinceana</i> | 8. <i>P. pinceana</i> |
| | 9. <i>P. quadrifolia</i> |
| clasificación incierta | 10. <i>P. remota</i> |
| 9. <i>P. maximartinezii</i> | 11. <i>P. rzedowskii</i> |
| 10. <i>P. rzedowskii</i> | |
| | Subsección <i>Nelsoniae</i> |
| | 12. <i>P. nelsonii</i> |

¹ No incluida originalmente en Farjon y Styles (1997).

² Clasificación de Price *et al.* (1997).

(Cuadro 1). Además, fue posible inferir las relaciones entre *P. culminicola*, *P. johannis* y *P. cembroides*, lo cual sugiere que *P. johannis* no es una subespecie de *P. cembroides* como argumentaron Farjon y Styles (1997), sino el grupo hermano de *P. culminicola*.

Este capítulo toma como ejemplo a los pinos piñoneros para ilustrar los conceptos básicos de inferencia filogenética mediante secuencias de DNA y discutir los resultados con referencia a caracteres morfológicos. Para ello se presenta una matriz expandida de las secuencias del cloroplasto, con el fin de hacer comparaciones con clasificaciones tradicionales previas. Mediante una combinación de caracteres morfológicos y de DNA, se intenta investigar cómo surgieron los caracteres del cono y follaje del grupo, además de conocer qué caracteres son homoplásicos y cuáles son evidencia de ancestría común.

Métodos

Diseño del muestreo

Para probar las clasificaciones previas de pinos piñoneros, el estudio se limita a una sola secuencia por cada especie y subespecie reconocida por Malusa (1992) y Farjon y Styles (1997). El único taxón que no se incluyó fue *P. monophylla* ssp. *fallax*, subespecie reconocida por Price *et al.* (1998), pero considerado como *P. californiarum* ssp. *fallax* por Malusa (1992) y como un sinónimo de *P. monophylla* por Farjon y Styles. De acuerdo con Nixon y Carpenter (1993), se incluyeron como grupos externos representantes de grupos cercanamente emparentados con los pinos piñoneros pero que no forman un linaje adentro del grupo. Estas especies se escogieron de la subsección *Balfourianae*, también de la sección *Parrya*, así como de otras secciones de *Pinus* (sección *Quinquefoliae* del subgénero *Strobos* y la distante sección *Pinus* del subgénero *Pinus*).

Para el marcador molecular, se escogió el genoma del cloroplasto (recuérdese que además existen los genomas del núcleo y de mitocondria en las plantas). Las regiones del DNA del cloroplasto escogidas fueron los genes *rbcl* y *matK*, el intron de *rpl16*, el espaciador *atpB-rbcl* y el espaciador *rpl20-rps18*. Un estudio previo (Wang *et al.*, 1999) demostró la utilidad de *rbcl*, *matK* y la región *rpl20-rps18* en pinos debido a que presentan una variación baja pero

adecuada para resolver diferencias entre especies.

El gen *rbcl* codifica para la subunidad grande de la ribulosa, 1-5-bifosfatasa carboxilasa/oxigenasa (RUBISCO), que es la proteína principal en la fijación de dióxido de carbono en la fotosíntesis. Secuencias de este gen fueron las primeras en usarse para inferir una filogenia molecular de los linajes principales de angiospermas (Chase *et al.*, 1993). El gen *matK* es una maturasa involucrada en el procesamiento (*splicing*) de intrones grupo II que codifican para tRNALys, tiene una tasa de sustitución mayor que *rbcl* y ha sido usado recientemente para generar una filogenia molecular de angiospermas (Hilu *et al.*, 2003). La región *rpl20-rps18* fue usada con *rbcl*, *matK* y un intron, en un estudio previo de pinos que no incluyó a los pinos piñoneros (Wang *et al.*, 1999). Esta región contiene parte de los genes *rpl20* y *rps18*, además de una copia completa del gen *rps12* que está ubicada entre los dos genes anteriores (Wakasugi *et al.*, 1994). El intron de *rpl16* ha sido más usada en estudios de angiospermas (Kelchner y Wendel, 1996). La región analizada en el estudio incluye una parte del gen *rpl16*, un intrón de grupo III y otra copia del gen *rps12* (Wakasugi *et al.*, 1994). La mayoría de la variación se presenta en el intrón del grupo III (Gernandt *et al.*, 2003).

Recolecta de muestras

Por falta de acceso a una población natural, la subespecie *P. cembroides* ssp. *lagunae* fue recolectada, después de verificar los caracteres diagnósticos, del Pinetum Maximino Martínez, Universidad Autónoma de Chapingo, México. En lugar de confiar en muestras de jardines botánicos, las cuales pueden no contar con datos de recolecta o carecer de las estructuras necesarias para una identificación correcta, se recolectó material de los demás taxones en el campo. Se tomó nota del sitio de recolecta, fecha, taxón, localidad (latitud, longitud y altitud), el tipo de hábitat y nombres de los recolectores. En el caso de *P. nelsonii*, *P. maximartinezii*, *P. johannis* y *P. culminicola*, los ejemplares fueron recolectados cerca de la localidad tipo (localidad donde se recolectó el holotipo de la especie). Así, si resulta que algún taxón es polifilético o altamente variable, no habrá duda a cuál población corresponde el nombre del taxón. Las especies fueron identificadas mediante las claves de la revisión del género *Pinus* por Farjon y Styles (1997) y la reciente flora de los Estados

Unidos y Canadá por Thieret (1993). Para permitir la verificación de las identificaciones taxonómicas por investigadores posteriores, se depositaron ejemplares *voucher* de cada taxón con la excepción de *P. rzedowskii*, en los herbarios del Instituto de Biología de la Universidad Nacional Autónoma de México, Universidad Estatal de Oregon, Estados Unidos y de la Universidad Autónoma del Estado Hidalgo, México.

Extracción de DNA

En el campo se colocaron algunas hojas en pequeños sobres de papel y fueron llevadas al laboratorio para ser congeladas a -80°C . Así, algunas muestras todavía pueden ser útiles para extracciones de DNA por muchos años después de haber sido recolectadas. El tejido fue macerado en un mortero para romper las paredes celulares. La mayoría de las muestras fueron maceradas en la presencia de nitrógeno líquido, algunas solamente con buffer y otros con arena. Se siguieron dos protocolos de extracción diferentes, uno con CTAB y cloroformo (Doyle y Doyle, 1987) y otro con un *kit* comercial que usa filtros específicos de sílica gel que forman ligas con el DNA en la presencia de etanol, pero no de agua (*DNeasy Plant Mini Kit*, Qiagen). Una vez aislado el DNA, se diluye en una solución de TRIS-EDTA (o en agua, pero en este caso hay que mantener las extracciones congeladas pues si no se degrada el DNA).

Determinación de cantidad y calidad de DNA

Para mejores resultados en pasos posteriores, se midió la calidad del DNA para poder estandarizar la concentración. Para cada extracción, $2\ \mu\text{L}$ fueron cargados en un gel de 0.8% agarosa en TBE junto con un estándar (*DNA mass ladder*, Invitrogen). Se llevó a cabo la electroforesis del gel en la presencia de bromuro de etidio (substancia mutagénica que requiere manejo cuidadoso) a una concentración de $0.5\ \mu\text{g}$ por mL de gel. El DNA en el gel fue observado sobre una caja de luz UV (312 nm). El DNA de alta calidad debe verse como una banda bien definida. Si hay impurezas (carbohidratos o metabolitos secundarios) o la presencia de DNA degradado, se observan manchas antes y después de la banda. Para las extracciones que presentaron impurezas, el DNA fue reprecipitado con etanol y acetato de sodio (Anexo

A.4) y diluido de nuevo en TE. Si la cantidad de DNA resultó baja, se repitió la extracción con modificaciones de volúmenes, cantidades de tejido o tiempo de incubación.

Amplificación de una región de DNA

Se usó la técnica de PCR para amplificar DNA (incrementar la cantidad de un producto específico). Se agregó entre 0.5 y $1.0\ \mu\text{L}$ (aproximadamente $1\ \text{ng}$) de DNA genómico a un tubo que contiene una solución de ingredientes necesarios para la polimerización del DNA. La concentración final de los demás ingredientes fue: 1X buffer de PCR, $1.5\ \text{mM}$ MgCl_2 , $200\ \mu\text{M}$ de cada desoxinucleótido (dNTP: dATP, dCTP, dGTP y dTTP), $1\ \mu\text{M}$ de cada uno de los dos oligonucleótidos iniciadores (*primers* de aproximadamente 20 bases de secuencia complementaria de las regiones que flanquean a la secuencia de interés) y $1.5\ \text{U}$ de polimerasa termoestable de DNA (p. ej., *Taq*). Las reacciones se llevaron a cabo en un termociclador, primero aplicando una temperatura de 94°C por tres minutos para romper los puentes de hidrógeno entre las dos cadenas del molde de DNA y así separarlas, seguido por 30 ciclos de: 1) $3\ \text{min}$ a 94°C , 2) $50\ \text{s}$ a 55°C para que se asocien los *primers* a los sitios complementarios del molde y 3) $80\ \text{s}$ a 72°C para que la polimerasa lleve a cabo la síntesis de una hebra nueva que corresponde a la región de interés. Después de terminar los ciclos, se mantuvieron los tubos a 72°C para asegurar que la amplificación de las secuencias en los extremos sean terminadas. Después, una muestra de $2\ \mu\text{L}$ se sometió a electroforesis en un gel de agarosa para verificar cuáles reacciones fueron exitosas.

Para remover los dNTPs, *primers*, magnesio y polimerasa que sobraron, los productos de PCR fueron purificados con un kit comercial (*QIAquick PCR purification kit*, Qiagen) y diluidos en agua bidestilada. Finalmente, se midió la concentración en un gel de agarosa para determinar y estandarizar la cantidad que se ocupó para la secuenciación.

Secuenciación

El método enzimático de Sanger tiene muchos elementos en común con la PCR, la polimerización del DNA mediante una polimerasa termoestable (y su cofactor), junto con un *primer* y dNTPs. En contraste con la PCR, se usa solamente un

primer por reacción y además incluye, en concentraciones bajas, didesoxinucleótidos (ddNTPs: ddATP, ddCTP, ddGTP y ddTTP) que se incorporan en la cadena en una frecuencia menor que los dNTPs. Debido a que los ddNTPs carecen de un grupo hidroxilo en el extremo 3' con respecto a los dNTPs, una vez que uno se incorpora a la hebra que se está formando la reacción de polimerización no puede continuar. Los productos de una reacción enzimática de secuenciación conforman una solución con una población de fragmentos de DNA de una sola hebra, que varían en longitud desde aproximadamente 21 bases (el tamaño del *primer*) hasta más de 800 bases. Todos los fragmentos de un tamaño específico son idénticos y terminan con un ddNTP que complementa al nucleótido correspondiente del molde que fue copiado. Hay un tinte fluorescente unido al ddNTP (*dye terminator*) y cada uno de los cuatro ddNTPs tiene un tinte diferente que permite distinguirlos entre sí.

Para la secuenciación cíclica enzimática (una variante del método de Sanger), el producto de PCR es agregado a un tubo con *primer*, buffer y *dye terminator* (Anexo A.3). La mezcla se colocó en el termociclador por 40 ciclos de: 1) 95 °C por 20 s para separar las cadenas del DNA molde, 2) 15 s a 50 °C para que se una el *primer* y 3) 1 min a 60 °C para llevar a cabo la síntesis del DNA. Después de terminar los ciclos, se reduce la temperatura a 10 °C por 4 min. Mientras los fabricantes recomiendan limitar el número de ciclos entre 25 y 30 para evitar que errores cometidos por la polimerasa en ciclos tempranos se amplifiquen en ciclos posteriores, en el laboratorio se ha podido usar un cuarto del volumen de reactivo con 40 ciclos y así se redujeron los costos de secuenciación significativamente.

Después de llevar a cabo la reacción en un termociclador, los productos fueron precipitados con etanol y acetato de sodio (Anexo A.4) y analizados en una máquina de secuenciación automatizada (ABI 310, *Applied Biosystems, Inc.*). Los fragmentos resultantes de la secuenciación fueron separados mediante electroforesis en un capilar con un polímero especial (POP6). El tinte de cada fragmento fue excitado por un láser de argón y un sensor en la máquina recibió los fotones en la longitud de onda que corresponde a su color. Estas señales fueron enviadas a una computadora que las tradujo en una secuencia de bases.

Alineación

Las lecturas contiguas y complementarias para las dos hebras de DNA fueron editadas y alineadas con el programa gratuito *BioEdit* (Hall, 1999). Las diferencias entre la hebra *sentido* y la *antisentido* fueron resueltas comparando los *trace-files*, los cuales presentan la serie de picos de cuatro colores que representan la lectura de las bases. En este caso fue posible determinar con confianza cuál nucleótido apareció en una posición y no fue necesario usar los códigos de ambigüedad (por ejemplo: N = cualquier nucleótido, ACGT; Y = T o C; R = A o G). Se guardó una sola secuencia de consenso para cada taxón en el estudio. Las regiones del cloroplasto presentaron muy pocas instancias de inserciones o deleciones (*indels*, ya que difícilmente se distingue entre cualquiera de las dos antes del análisis). Por tanto, la alineación de las secuencias, la hipótesis de homología posicional entre sitios, se llevó a cabo manualmente, sin la necesidad de un programa como *CLUS-TALW* (Thompson *et al.*, 1994) que está incluido con *BioEdit*. Tampoco fue necesario excluir partes de la alineación donde ésta fuera incierta para algunos taxones. En este ejemplo, se decidió no incluir los *indels* como caracteres en el análisis, tal como lo recomendaron Simmons y Ochoterena (2000).

Análisis filogenético

Dos análisis fueron efectuados con *PAUP** (Swofford, 2003). El primero fue mediante el método cladístico. Este método está sustentado en el criterio de parsimonia, el cual busca minimizar el número de cambios evolutivos requeridos para explicar la distribución de estados de carácter en una matriz de datos, y encuentra grupos naturales que comparten uno o más caracteres derivados (sinapomorfías). Se realizó una búsqueda *branch-and-bound*, que es exacta porque garantiza encontrar todos los árboles más parsimoniosos (AMP). También se estimaron medidas de apoyo de ramas mediante búsquedas heurísticas (que no son exactas) de 1,000 réplicas de *bootstrap* (Felsenstein, 1985) y 1,000 réplicas de *jackknife* (Miller, 1974), con 50% de deleción de caracteres. Los dos métodos utilizaron una agregación de taxones secuencial simple (*simple sequence addition*) y un intercambio de ramas (*tree bisection reconnection o TBR*) como algoritmos de búsqueda. Además, se calculó el índice de descomposición (*decay*

| | |
|-------------|---|
| cembroides | GTCAACTTTTTCAACAATAGACTAAAGATCTGCATCTTTATCGAACAGATATC |
| californiar | |
| monophylla | ...G..... |
| quadrifolia | ...G..... |
| discolor | |
| johannis | |
| culminicola | |
| remota | |
| cembroides | |
| lagunae | |
| orizabensis | |
| pinceana | ...G..... |
| maximartin | ...G..... |
| rzedowskii | ...G.....A..... |
| nelsonii | ...G.....C..... |
| aristata | ...G.....C..... |
| longaeva | ...G.....C..... |
| balfouriana | ...G.....C..... |
| bungeana | ...G.....C..... |
| gerardiana | ...G.....C..... |
| squamata | ...G.....C..... |
| krempfii | ...G.....C..... |
| strobis | ...G.....A.....C.....AA..... |
| chiapensis | ...G.....A.....C.....AA..... |
| ayacahuite | ...G.....A.....C.....AA..... |
| thunbergii | ..TG.....C.....T.....T |

Figura 1. Alineación de parte del gen *matK* en 14 taxones de *Pinus* subsección *Cembroides* y grupos externos. Son visibles las bases en el primer taxón. En los demás taxones solamente aparecen las bases cuando difieren del primero.

index; Bremer, 1994) mediante búsquedas heurísticas (con un límite superior *MAXTREES* = 5000), con el comando *KEEP* para árboles cinco pasos más largos que los AMP y mediante los comandos *CONSTRAINTS* y *CONVERSE* para cladogramas con índices mayores a cinco.

Algunos sistemáticos moleculares han criticado a la parsimonia porque requiere muchos caracteres para encontrar una solución robusta (baja eficiencia). También, porque existe un fenómeno, la atracción entre ramas largas (*long branch attraction*), que consiste en una tendencia del método de unir ramas con altas tasas de evolución independientemente si los taxones son o no grupos hermanos (Felsenstein, 1978). Curiosamente, esta tendencia resulta más pronunciada conforme se agregan más caracteres, por lo cual se ha descrito como *positively misleading*.

El segundo método utilizado fue el de máxima verosimilitud (Lewis, 1998), que es, según sus seguidores, menos susceptible a la atracción entre ramas largas.

Además le permite hacer un uso más eficiente de los datos. Es decir, la máxima verosimilitud es capaz de encontrar una hipótesis razonable de relaciones filogenéticas con menos datos que la parsimonia (Swofford *et al.*, 1996). Con el fin de escoger un modelo evolutivo de manera explícita, la matriz del cloroplasto fue analizada con la prueba de log-verosimilitud jerárquica (*hierarchical likelihood ratio test*) y el criterio de información de Akaike por medio del programa *ModelTest* (Posada y Crandall, 1998). No obstante, no se intentó determinar si las regiones u otras particiones (por ejemplo, sitios silenciosos y no silenciosos) requerían de modelos diferentes. La búsqueda de máxima verosimilitud se llevó a cabo con *PAUP** bajo las condiciones determinadas por el programa.

Resultados

La matriz de secuencias tuvo un longitud de 5,045 pb (*atpB-rbcL* espaciador parcial: 322, *rbcL*: 1525, *matK*: 1595, región

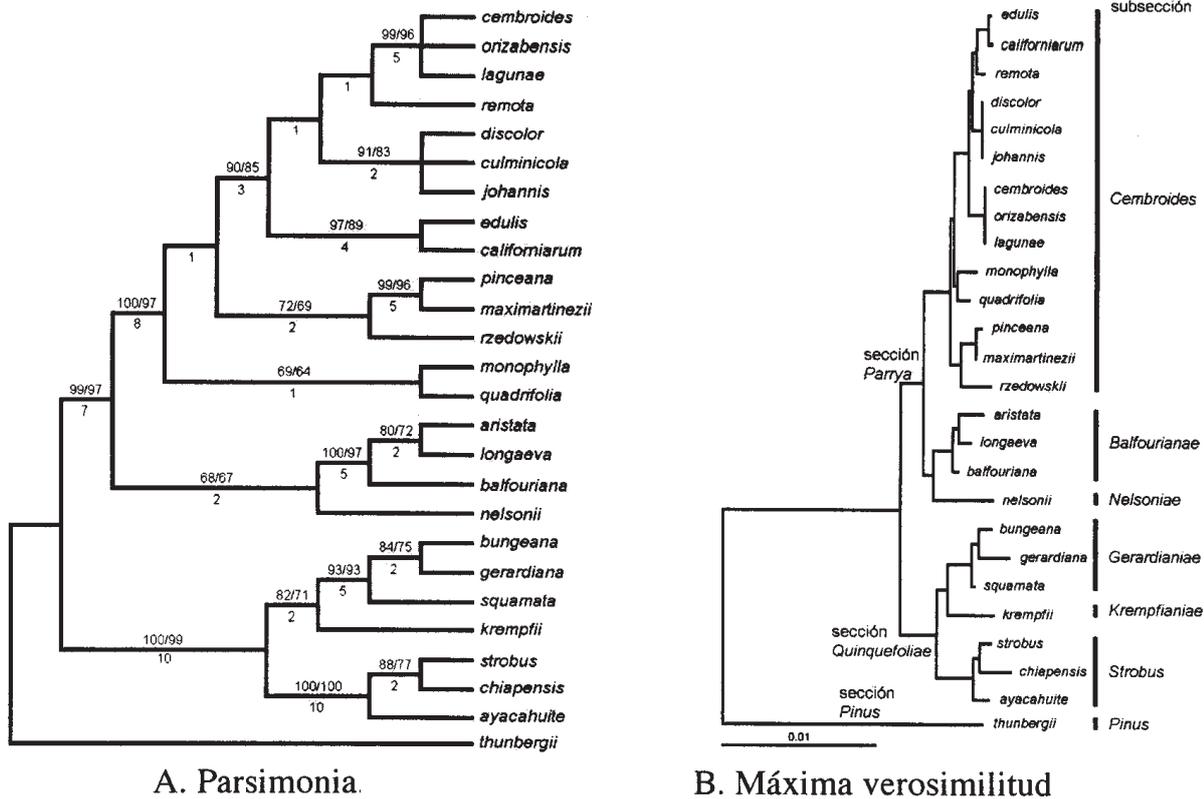


Figura 2. Filogenias de los pinos piñoneros. A, El árbol de parsimonia mostrado es un consenso estricto de dos AMP ($L = 320$, $IC = 0.844$, $ICexc = 0.648$, $IR = 0.87$). Valores encima de las ramas son porcentajes de bootstrap y jackknife, por debajo son los índices de descomposición. B, El árbol de máxima verosimilitud está basado en el modelo $TVM + G + I$ ($-\ln$ verosimilitud = 9285.98214).

de *rpl20-rps18*: 549, *rpl16*: 1131). La figura 1 muestra una parte de la alineación. La matriz tuvo 265 caracteres variables, de los cuales 89 fueron informativos. De los 265 caracteres variables, solo 47 fueron variables entre las 14 taxa de la subsección *Cembroides*, dando 3.4 caracteres variables por taxón. Las regiones que aportaron más caracteres variables dentro de la subsección *Cembroides* fueron las más largas, *rbcl* (7 sitios) y *matK* (7 sitios), sin embargo, la tasa más alta de sustitución se encontró en el espaciador *atpB-rbcl* (0.025 sustituciones por sitio) y la región *rpl20-rps18* (0.013 sustituciones por sitio).

Una búsqueda de *branch-and-bound* encontró dos árboles más parsimoniosos (AMP) con una longitud (L) de 320 pasos, de los cuales se hizo un consenso estricto (Fig. 2A). La sección *Parrya* incluye tres clados principales que corresponden a las subsecciones *Cembroides*, *Nelsoniae* (México) y *Balfourianae* (Estados Unidos). La sección *Quinquetfoliae* está

compuesta de las subsecciones *Krempfianae* (Vietnam), *Gerardianae* (este de Asia) y *Strobos* (Eurasia y América del Norte). El clado de la subsección *Cembroides* está compuesto de tres linajes principales: un linaje con una divergencia temprano de *P. quadrifolia* y *P. monophylla* ssp. *monophylla* y dos linajes más derivados. Uno de los dos incluyó a *P. maximartinezii* monofilético con *P. pinceana*, ambos como grupo hermano de *P. rzedowskii*. El segundo linaje está compuesto a su vez de cuatro linajes: 1) las tres subespecies de *P. cembroides*, 2) *P. remota*, 3) *P. johannis*, *P. discolor* y *P. culminicola* y 4) *P. edulis* y *P. monophylla* ssp. *californiarum* (*sensu* Price *et al.*, 1998 y Gernandt *et al.*, 2005).

Los análisis de bootstrap y jackknife, respectivamente, indicaron que la monofilia de la sección *Parrya* está bien apoyada (100% y 97%) igual que la sección *Quinquetfoliae* (100% y 99%). Las subsecciones *Cembroides*, *Balfourianae*,

Gerardianae y *Strobis* también recibieron apoyos superiores al 90% en los análisis de parsimonia. La monofilia de la subsección *Nelsoniae* y *Balfourianae* fue apoyada por valores menores (68% y 67%). La relación de grupos hermanos entre las subsecciones *Gerardianae* y *Krempfianae* fue apoyada por valores de 82% y 71%. Los valores de apoyo para los tres linajes principales de la subsección *Cembroides* fueron generalmente bajos; por ejemplo, el apoyo de la posición de *P. monophylla* y *P. quadrifolia* es menor de 50% (por eso no aparecen valores sobre esta rama). Los valores para las relaciones entre los cuatro linajes (p. ej., *P. cembroides*, *P. remota*, *P. johannis*, *P. edulis*) también son menores de 50%.

Los índices de descomposición fueron bajos, entre 1 y 2 para 11 de las 21 ramas internas. Los clados de los pinos piñoneros que recibieron valores mayores fueron los de las tres subespecies de *P. cembroides* ($d = 5$), el de *P. edulis* y *P. californiarum* ($d = 4$) y el de *P. pinceana* y *P. maximartinezii* ($d = 5$). El apoyo para subsecciones fue $d = 8$ para *Cembroides*, $d = 7$ para *Balfourianae*, $d = 5$ para *Gerardianae* y $d = 10$ para *Strobis*. Los índices fueron altos para las dos secciones, $d = 7$ para *Parrya* y $d = 10$ para *Quinquefoliae*.

El árbol de máxima verosimilitud tuvo un valor de $-\ln$ verosimilitud = 9285.98214 (Fig. 2B). Este árbol presentó diferencias importantes con respecto a los AMP. El linaje hermano a los demás pinos piñoneros ya no es *P. quadrifolia* y *P. monophylla*, sino que este linaje intercambió lugar con el clado de *P. maximartinezii*, *P. pinceana* y *P. rzedowskii*. Otra diferencia es que *P. remota* es el grupo hermano de *P. edulis* y *P. californiarum* en el árbol de verosimilitud. Por último, el clado de *P. johannis*, *P. discolor* y *P. culminicola* es el grupo hermano de *P. edulis*, *P. californiarum* y *P. remota* y no el de *P. cembroides*, como en los árboles de parsimonia. Las ramas que varían entre el análisis de parsimonia y de verosimilitud son las mismas que tuvieron valores bajos de apoyo (*bootstrap*, *jackknife* y descomposición).

Discusión

La filogenia del DNA del cloroplasto para los pinos piñoneros no ha variado en sus linajes principales, a pesar de las diferencias de muestreo taxonómico y un aumento en el número de caracteres a lo largo de varios años (Gernandt *et al.*,

2003, 2005, este estudio). La separación de *P. nelsonii* de los demás pinos piñoneros también fue apoyada con secuencias de la región del ITS de DNA ribosomal (Gernandt *et al.*, 2001). A pesar de que nunca ha recibido buen soporte (p. ej., *bootstrap*), *P. nelsonii* ha formado un clado con la subsección *Balfourianae* de una manera consistente en los análisis de cpDNA y del ITS. La parafilia de los pinos piñoneros podría apoyar la hipótesis de que las semillas agrandadas sin alas de los pinos piñoneros evolucionaron en dos o tres eventos independientes. La posición de *P. rzedowskii* es relevante porque es el único "pino piñonero" que no tiene una semilla modificada como un piñón. La hipótesis más parsimoniosa para explicar porque *P. rzedowskii* siempre forma un clado con *P. maximartinezii* + *P. pinceana*, en lugar de ser el grupo hermano de todas las demás especies de la subsección *Cembroides*, es que el ancestro común de la sección *Parrya* tuvo una semilla pequeña y alada (lo mismo que se encuentra en los grupos externos) y ocurrió una modificación, a una semilla tipo piñón, en tres linajes independientes; en *P. maximartinezii* + *P. pinceana*, en los demás pinos piñoneros y en *P. nelsonii*. Esta hipótesis sugiere que la evolución de un piñón se revierte fácilmente. ¿Será posible que una sola mutación cause la transformación entre una semilla alada y un piñón?

La posición de *P. rzedowskii* + *P. pinceana* + *P. maximartinezii* dentro de *P. cembroides* ha variado en diferentes análisis, por ejemplo el análisis de máxima verosimilitud lo ubica como el grupo hermano a los demás pinos piñoneros, mientras que el análisis de parsimonia ubica el clado en una rama más derivada. Ninguna de las dos topologías afecta la interpretación de la derivación del piñón mencionada arriba. Sin embargo, estas tres especies tienen conos más grandes como las de las subsecciones *Nelsoniae* y *Balfourianae*. La posición de las tres especies en el árbol de máxima verosimilitud apoya que el ancestro de la subsección *Cembroides* tuvo un cono mediano y que este tamaño fue reducido una sola vez en el linaje que dio origen al resto de la subsección *Cembroides*. Si estudios posteriores con más datos corroboran esta hipótesis, habrá que reconocer que el método de máxima verosimilitud logró resolver esta parte del árbol antes que la parsimonia.

La inferencia filogenética puede ser útil para investigar la

delimitación de especies. Brower (1999) propuso una manera de delimitarlas mediante análisis cladísticos de DNA del cloroplasto o mitocondria; si un grupo propuesto con base en una hipótesis *a priori* resulta ser contiguo en árboles no-enraizados, entonces pueden ser reconocidos como especies válidas. Existen desacuerdos en el número de especies de pinos piñoneros, así como en la clasificación de algunas especies dudosas que presentan combinaciones de caracteres que sugieren relaciones con especies diferentes.

Un ejemplo del uso de los haplotipos para delimitar especies es entre los taxones del suroeste de los Estados Unidos y los de Baja California. El número de hojas arregladas en un fascículo es un carácter fácil de observar en los pinos; muchos taxones han sido clasificados con base en eso. Por ejemplo, los pinos con una sola acícula por fascículo comúnmente son clasificados como *P. monophylla*. El taxón *P. californiarum* (Fig. 2) tiene una sola acícula por fascículo igual que *P. monophylla*. Sin embargo, tiene conos y semillas de tamaño intermedio entre *P. monophylla* y *P. edulis*, tiene un elevado contenido de grasa en su megagametofito y más canales resiníferas en las hojas (Zavarín *et al.* 1990). Este taxón misterioso ha sido reconocido bajo varios nombres que incluyen: *Pinus monophylla*, *Pinus californiarum*, *Pinus edulis* var. *californiarum* y *P. monophylla* ssp. *californiarum* (Farjon y Styles, 1997; Price *et al.*, 1998). El DNA del cloroplasto soporta fuertemente una relación entre este taxón y *P. edulis*. Este resultado parece apoyar la idea de que *P. californiarum* sea la especie hermana de *P. edulis* y que la reducción a una sola hoja por fascículo sucedió en dos linajes independientes. Sin embargo, hay que reconocer una limitación importante del resultado del DNA del cloroplasto. Las especies *P. edulis*, *P. monophylla* y *P. quadrifolia* forman híbridos, entonces puede esperarse encontrar evidencia de hibridación en su DNA. En este caso tenemos un ejemplo en el cual la filogenia de las especies puede ser incongruente con la filogenia de un genoma completo. Para determinar las relaciones entre las tres especies mencionadas hay que incluir datos de los otros dos genomas disponibles en las plantas, del núcleo y de la mitocondria, y muestrear más poblaciones a lo largo del intervalo de distribución para poder determinar los patrones en detalle.

Otro ejemplo de la delimitación de especies se encuentra

en un pino piñonero del sur de los estados de Arizona, Nuevo México y Texas, así como de la Sierra Madre Oriental y la Sierra Madre Occidental. Debido a incertidumbre en su clasificación los taxónomos lo han llamado *Pinus cembroides* var. *bicolor*, *P. discolor* y *P. johannis* (Farjon y Styles, 1997). Mientras tiene muchas similitudes con *P. cembroides*, sus estomas están limitados a la superficie adaxial de las hojas y se parece más en este respecto a *P. culminicola*, una especie que se encuentra en las cimas más altas de la Sierra Madre Oriental. Los resultados del cloroplasto (Fig. 2) sugieren que *P. johannis* y *P. discolor* son el grupo hermano de *P. culminicola* y no de *P. cembroides*. A pesar de las dudas sobre la hibridación, este resultado sugiere que el taxón *P. cembroides* var. *bicolor* no es válido sino que *P. johannis* debe ser reconocido como una especie válida (*P. discolor* puede ser un sinónimo posterior de *P. johannis*, pero este taxón requiere más investigación).

En este capítulo se expusieron varias maneras de someter a prueba hipótesis morfológicas mediante datos moleculares. Una manera más avanzada ocurre al combinar los datos moleculares con los datos tradicionales. Esto puede generar hipótesis aún más robustas y útiles. Por otra parte, los caracteres tradicionales pueden ser combinados con los moleculares para estudiar los orígenes de la endosimbiosis, la adaptación de estructuras morfológicas o anatómicas (incluso la evolución paralela o convergente) y la incorporación de taxones fósiles. Finalmente, algunas hipótesis son sometidas a prueba principalmente con caracteres moleculares. Algunos ejemplos incluyen la duplicación de genes, la evolución paralela de enzimas, la existencia de mutaciones compensatorias para mantener los tallos en el RNA estructural, el sesgo de codones para regular la traducción de proteínas y la transferencia lateral de genes.

Agradecimientos

Agradecemos a Ulises Razo Mendivil, Delia Edelmira González Garibaldi, Carmen López Ramírez, Esmeralda Salgado Hernández y Denisse Téllez Mazzocco por su ayuda en el campo y en el laboratorio. Este proyecto fue apoyado por DGAPA (UNAM) y PROMEP de la Secretaría de Educación Pública.

Literatura citada

- Bremer, K. 1994. Branch support and tree stability. *Cladistics* 10: 295-304.
- Brower, A. V. Z. 1999. Delimitation of phylogenetic species with DNA sequences: a critique of Davis and Nixon's population aggregation analysis. *Systematic Biology* 48: 199-213.
- Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duball, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sysma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. Barrett, S. Dayanandan y V. A. Albert. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528-580.
- Doyle, J. J. y J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11-15.
- Farjon, A. y B. T. Styles. 1997. *Pinus* (Pinaceae). *Flora Neotropica Monograph* 75. The New York Botanical Garden, New York.
- Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology* 27: 401-410.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.
- Gernandt, D. S., A. Liston y D. Piñero. 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implications for molecular systematic studies of pine species complexes. *Molecular Phylogenetics and Evolution* 21: 449-467.
- Gernandt, D. S., A. Liston y D. Piñero. 2003. Phylogenetics of *Pinus* subsections *Cembroides* and *Nelsoniae* inferred from cpDNA sequences. *Systematic Botany* 28: 657-673.
- Gernandt, D. S., G. Geada López, S. Ortiz García y A. Liston. 2005. Phylogeny and classification of *Pinus*. *Taxon* 54: 29-42.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
- Hillis, D. M. y J. J. Wiens. 2000. Molecules versus morphology in systematics: conflicts, artifacts, and misconceptions, pp. 1-19. En: J. J. Wiens (ed.). *Phylogenetic analysis of morphological data*. Smithsonian Institution Press, Washington.
- Hilu, K. W., T. Borsch, K. Müller, D. E. Soltis, P. S. Soltis, V. Savolainen, M. W. Chase, M. P. Powell, L. A. Alice, R. Evans, H. Sauquet, C. Neinhuis, T. Slotta, J. G. Rohwer, C. S. Campbell y L. W. Chatrou. 2003. Angiosperm phylogeny based on matK sequence information. *American Journal of Botany* 90: 1758-1776.
- Kelchner, S. A. y J. F. Wendel. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30: 259-262.
- Lewis, P. O. 1998. Maximum likelihood as an alternative to parsimony for inferring phylogeny using nucleotide sequence data, pp. 132-163. En: D. E. Soltis, P. S. Soltis y J. J. Doyle (eds.). *Molecular systematics of plants II: DNA sequencing*. Kluwer Academic Publishers, Norwell.
- Liston, A., W. A. Robinson, D. Piñero y E. R. Alvarez-Buylla. 1999. Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Molecular Phylogenetics and Evolution* 11: 95-109.
- Malusa, J. 1992. Phylogeny and biogeography of the pinyon pines (*Pinus* subsect. *Cembroides*). *Systematic Botany* 17: 42-66.
- Miller, R. G. 1974. The jackknife—a review. *Biometrika* 61: 1-15.
- Mullis, K. 1990. The unusual origin of the polymerase chain reaction. *Scientific American* 4: 56-65.
- Nixon, K. C. y J. M. Carpenter. 1993. On outgroups. *Cladistics* 9: 413-426.
- Posada, D. y K. A. Crandall. 1998. Modeltest: using the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Price, R. A., A. Liston y S. H. Strauss. 1998. Phylogeny and systematics of *Pinus*, pp. 49-68. En: D. M. Richardson (ed.). *Ecology and biogeography of Pinus*. Cambridge University Press, Cambridge.
- Saiki, R., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich y N. Arnheim. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-1354.
- Simmons, M. y H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369-381.
- Swofford, D. L. 2003. PAUP*. *Phylogenetic analysis using parsimony (*and other methods)*. Version 4.0b10. Sinauer Associates, Sunderland.
- Swofford, D. L., G. J. Olsen, P. J. Waddell y D. M. Hillis. 1996. Phylogenetic inference, pp. 407-514. En: D. M. Hillis, C. Moritz y B. K. Mable (eds.). *Molecular systematics*, 2da ed. Sinauer Associates, Sunderland.
- Thieret, J. W. 1993. Pinaceae Lindley: pine family, pp. 352-398. En: *Flora of North America Editorial Committee* (ed.). *Flora of North America North of Mexico*, vol. 2. Oxford University Press, New York.
- Thompson, J. D., D. G. Higgins y T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.
- Wakasugi, T., J. T. Tsudzuki, S. Ito, K. Nakashima, T. Tsudzuki y M. Sugiura. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences* 91: 9794-9798.
- Wang, X.-R., Y. Tsumura, H. Yoshimaru, K. Nagasaka y A. E. Szmidt. 1999. Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer, and *trnV* intron sequences. *American Journal of Botany* 86: 1742-1753.
- Zavarin, E., K. Snajberk y L. Cool. 1990. Chemical differentiation in relation to the morphology of the single-needle pinyons. *Biochemical Systematics and Ecology* 18: 125-137.

ANEXO

Protocolos útiles para la secuenciación de DNA

A. 1 Dilución de *Primers* a 100 μM .

El *primer stock* se puede almacenar a 100 μM (= 100 $\mu\text{moles/L}$). Para calcular cuántos μL de TE se debe agregar a primers liofilizados, hay que convertir las unidades del primer a μmoles :

$$(48.3 \text{ nmoles primer})(1 \mu\text{mol} / 1000 \text{ nmol}) = 0.0483 \mu\text{moles}$$

Y aplicar la fórmula:

No. μmol soluto = (concentración en $\mu\text{mol/L}$)(volumen en L)

$$0.0483 \mu\text{moles} = (100 \mu\text{moles/L})(V L)$$

$$V L = 0.0483 \mu\text{moles} / 100 \mu\text{moles/L} = 0.000483 \text{ L}$$

$$0.000483 \text{ L} (1 \times 10^6 \mu\text{L} / \text{L}) = 483 \mu\text{L}$$

Agrega 483 μL TE a 48.3 nmoles del primer para preparar una solución de 100 μM

Entonces, para preparar un *stock* de 100 μM de primer mediante un tubo de primer comercial, averiguar la cantidad de nmoles en el tubo y multiplicar por 10. Esto te da el número de μL TE que se debe agregar.

Ejemplo: para diluir 48.3 nmol de primer a 100 μM , agrega $48.3 \times 10 = 483 \mu\text{L}$ TE.

A. 2 Condiciones típicas para una PCR.

| reactivo | volumen (μL) | concentración final |
|--|---------------------------|-------------------------------|
| ddH ₂ O | 80.9 | |
| 10X buffer | 10 | 1 X ¹ |
| 50 mM MgCl ₂ | 3 | 1.5 mM |
| 25mM dNTP mix | 0.8 | 0.2 mM de cada dNTP |
| 50 μM primer 1 | 2 | 1 μM |
| 50 μM primer 2 | 2 | 1 μM |
| 5 U/ μL <i>Taq</i> polimerasa | 0.3 | 1.5 U / reacción ² |
| 10 ng/ μL DNA genómico | 1 | 0.1 ng / μL |
| | 100 | |

¹El buffer 10X incluye 200 mM Tris-HCl (pH 8.4) y 500 mM KCl

²La cantidad de enzima no se expresa por volumen sino por tubo.

A. 3 Condiciones típicas para una reacción de secuenciación.

| reactivo | volumen (μL) |
|-----------------------|---------------------------|
| producto purificado | 6 |
| 5X buffer SEQ | 1 |
| 2 uM primer | 1 |
| BigDye Terminator 3.1 | 2 |

A. 4 Precipitación con etanol y acetato de sodio.

1) Por cada 10 μL de reacción de secuenciación, agregar 1 μL de 3M acetato de sodio, (pH = 4.6) y 25 μL de 95% etanol. *Revisa que el volumen de la reacción de secuenciación sea de 10 μL : la concentración final de etanol debe ser aproximadamente 65%.*

2) Agitar los tubos y dejarlos a temperatura ambiente por 15 min para precipitar los productos de secuenciación. *Si el tiempo es menor de 15 min ocasionará la pérdida de los productos de extensión más cortos.*

3) Centrifugar los tubos por 20 min a 12000 rpm.

4) Cuidadosamente quitar el sobrenadante con una pipeta y desecharlo.

5) Enjuagar la bolita (no tiene que ser visible) con 150 μL de 70% etanol. *No es necesario mezclar.*

6) Centrifugar por 5 min a 12000 rpm. Vuelva a quitar el sobrenadante y desecharlo.

7) Secar la bolita en una centrifuga al vacío (SpeedVac) o en el aire.