

MARCADORES MOLECULARES, VARIABILIDAD GENÉTICA Y EVOLUCIÓN

Marcelino Pérez de la Vega

Area de Genética

Universidad de León

La generalización de las técnicas de Biología molecular ha tenido, entre otras, dos consecuencias trascendentales. Por una parte nos han dotado de una poderosa herramienta de uso general para el análisis genético tanto en estudios básicos como aplicados, los marcadores moleculares. Por otra nos han dado una visión mucho más completa de la variabilidad genética y su distribución a distintos niveles: individual, población, especie. El uso de tales herramientas y los nuevos datos adquiridos con ellas están teniendo profundas repercusiones teóricas y prácticas.

En esta conferencia describiré algunas técnicas moleculares para generar marcadores, la visión que gracias a ellos tenemos actualmente sobre la distribución de la variabilidad genética y la evolución, y algunas consecuencias teóricas y prácticas de todo ello.

1. MARCADORES GENÉTICOS

Existen varias definiciones de marcador genético. Rieger et al. (1982) lo definen como cualquier diferencia fenotípica controlada genéticamente utilizada en el análisis genético. Gale (1994) lo define como cualquier medio para identificar cualquier locus específico en un cromosoma. En definitiva podemos usar como marcador el efecto de un gen fácilmente observable en los individuos (genéricamente conocidos como caracteres morfológicos), metabolitos característicos de bajo peso molecular, proteínas que puedan extraerse y observarse con facilidad (isoenzimas, proteínas de reserva, proteínas del suero) generalmente tras un fraccionamiento mediante electroforesis, o segmentos de DNA que puede obtenerse e identificarse por toda una serie de técnicas moleculares (Pérez de la Vega, 1993). Es obvio que para ser informativo un marcador debe estar presente en formas alélicas alternativas en los individuos en estudio (parentales de un cruzamiento, individuos de una población, etc.), y para

una mayor utilidad es necesario saber donde se localiza en un cromosoma específico.

Un aspecto discutible es qué se entiende por marcador molecular. Para algunos autores sólo los ácidos nucleicos deben ser incluidos en esa categoría, otros incluyen también sus productos primarios, las proteínas, y distinguen entre marcadores proteicos (isoenzimas, por ejemplo) y marcadores DNA. Lo cierto es que el término de marcador molecular se empezó a generalizar cuando las técnicas permitieron estudiar fácilmente los polimorfismos a nivel de DNA y como distinción de los polimorfismos a nivel isoenzimático que venían estudiándose durante dos décadas. En esta conferencia utilizaré el sentido restrictivo (molecular = DNA) aunque las referencias y comparaciones con los marcadores de tipo proteico, y a las isoenzimas en particular, serán frecuentes. En la actualidad los marcadores moleculares han sobrepasado y suplantado el uso de otros marcadores genéticos en numerosísimas aplicaciones; sin embargo, en algunas, las isoenzimas siguen siendo preferidas, por ejemplo en la estimación de parámetros de selección, o en la evaluación de tasas de autogamia-alogamia en plantas.

2. LOS MARCADORES EN ESTUDIOS EVOLUTIVOS Y POBLACIONALES

El estudio de la evolución comprende dos áreas: la historia evolutiva de la vida y los mecanismos de evolución. Las fronteras entre ambos empezaron a eliminarse cuando los marcadores isoenzimáticos y la secuenciación de proteínas se empezaron a usar en los 60 como fuente de información evolutiva. Pero en la práctica la mayoría de los evolucionistas sólo se preocupaban de uno de estos problemas, incluso después de estos años. Los evolucionistas 'bioquímicos' se interesaban principalmente en construir árboles filogenéticos entre organismos evolutivamente alejados mientras que los genéticos de poblaciones se preocupaban de medir los niveles de polimorfismo en y entre poblaciones. La erosión real de estas fronteras comenzó cuando las técnicas de secuenciación de DNA y los marcadores moleculares se introdujeron en los estudios evolutivos (Nei, 1987).

Los marcadores bioquímicos y moleculares se han usado extensivamente en el estudio de la estructura genética de poblaciones y de la evolución (Pérez de la Vega, 1993) y se han usado como herramientas en mejora, fundamentalmente la de plantas (Arús y Moreno-González 1993; Lee, 1995). En ambos casos su uso representaba grandes ventajas sobre los marcadores 'morfológicos', que en el caso de las isoenzimas se concretaban en: 1) Normalmente la expresión alélica es codominante, libre de efectos epistáticos o del medio ambiente. 2) La especificidad enzimática permite atribuir los alelos a loci concretos, y comparar loci entre poblaciones y especies distintas. 3) Las diferencias alélicas se detectan como diferencias en movilidad, independientemente del papel funcional o del grado de variación de la enzima de que se trate. 4) Los loci que se estudian están determinados por su expresión en el tejido elegido, por su posible extracción y por la disponibilidad de tinción, más que por si el gen es o no variable. 5) Se

pueden estudiar simultáneamente varios marcadores, que pueden acumularse en un individuo sin causarle un efecto deletéreo (a diferencia de lo que ocurre frecuentemente con la acumulación de variantes morfológicas). El uso de isoenzimas permitió que por primera vez se pudiese elegir un grupo de genes sin saber previamente si eran variables o no, o su grado de variabilidad, en el nivel taxonómico a estudiar. Los genes para isoenzimas representaban por tanto un grupo de genes que podía considerarse representativo del acervo genético de una población o especie, es decir la información genética total codificada en la suma total de sus genes en un momento dado, que permitía una estima correcta de la variabilidad total, o lo más próximo a ello (Stebbins y Pérez de la Vega, 1989).

Los polimorfismos basados en el DNA ofrecen algunas ventajas sobre los anteriores: 1) Muestran niveles de variación más altos y pueden observarse mutaciones no detectables a otros niveles. 2) Mucha de la variación a nivel de nucleótidos es selectivamente neutra (desde luego más que la de las proteínas). 3) Se pueden analizar tres genomas distintos, nuclear, mitocondrial, y de cloroplasto, que evolucionan según modos y ritmos distintos. 4) Los polimorfismos a nivel de nucleótidos no se ven afectados por modificaciones post-transcripcionales (a nivel de cDNA es posible) o post-traduccionales. 5) El DNA de todas las células y tejidos es el mismo y su extracción no depende de una expresión previa. El segundo punto debe considerarse sólo como una ventaja relativa. La naturaleza selectivamente neutra de genes y marcadores es importante en Genética de poblaciones y evolutiva porque muchos de los modelos matemáticos para el estudio de la evolución se basan en este tipo de caracteres. Parece probado sin embargo que las distintas alternativas isoenzimática son, al menos parcialmente, adaptativas (Pérez de la Vega, 1996), y este papel adaptativo es, por el contrario, una ventaja cuando se trata de usar los marcadores en mejora y en la conservación de recursos genéticos. En este caso las isoenzimas tienen un doble interés: como genes adaptativos y como marcadores genéticos. El mayor inconveniente de algunos marcadores moleculares es su naturaleza dominante.

¿Por qué es tan importante disponer de marcadores que puedan ser escogidos sin saber a priori su grado de variabilidad e independientemente de su papel funcional?. Ante la imposibilidad de analizar el acervo genético de una población o especie, el estudio debe realizarse en muestras representativas de individuos y genes. Conseguir una muestra representativa de individuos se soluciona mediante un muestreo correcto de un número suficiente de individuos. El problema era más difícil de resolver para los genes. En el caso de los marcadores 'morfológicos' sólo sabemos que existen si hay variabilidad; sólo cuando hay al menos dos alternativas alélicas sabemos que un gen controla un carácter del tipo 'forma o color'. Sin embargo, en el caso de las proteínas, la presencia de una proteína diferenciable de cualquier otra (por movilidad electroforética, por ejemplo) implica que hay un gen que la codifica, independientemente de que la proteína sea absolutamente invariante en la población o especie de que se trate. Pero proteínas y marcadores 'morfológicos' representan sólo una parte del genoma, la que se expresa. Cada marcador DNA representa la existencia de una secuencia diferente variable o no, pero que

pueden expresarse o no (DNA repetitivo, pseudogenes, etc.), por lo que representan mejor al conjunto del genoma.

Otro aspecto importante es la naturaleza selectiva o neutra de los marcadores. ¿Por qué se insiste tanto en la ventaja de ser neutros respecto a la selección natural?. La razón es tanto técnica como metodológica. Es un hecho que es prácticamente imposible estimar la intensidad de la selección natural sobre caracteres concretos en la propia naturaleza, fuera de los experimentos controlados en el laboratorio, y mucho menos deducir cuál fue en el pasado la intensidad de ésta (Endler, 1986). Todo ello lleva frecuentemente a un razonamiento tautológico: de acuerdo con los principios de la selección natural aquellas alternativas favorecidas aumentarán su frecuencia, por ello todo gen frecuente tiende a suponerse como selectivamente ventajoso; la tautología proviene de olvidar que los cambios en las frecuencias génicas pueden producirse por causas distintas a la selección. La evolución puede producirse por causas diferentes a la selección natural. Por tanto si es difícil saber en algunos casos si un gen es adaptativo o no y siempre es extremadamente difícil estimar la selección natural, es obvio que cualquier modelo que incluya este parámetro es difícilmente comprobable experimentalmente. Sin embargo, para alternativas selectivamente neutras (total o prácticamente neutras) los modelos estadísticos que predicen su cambio no incluyen este parámetro, son de más fácil aplicación y, fundamentalmente, permiten probar la hipótesis de su neutralidad. De hecho, aunque a mediados de los 60 se habían desarrollado ya muchos modelos teóricos sobre la dinámica de los genes en las poblaciones, estas teorías eran usadas raramente para interpretar los datos experimentales excepto en contadas ocasiones. La situación cambió abruptamente cuando se dispuso de datos moleculares sobre el cambio evolutivo de los genes. Desde entonces las teorías se usan profusamente para probar hipótesis alternativas sobre los mecanismos de evolución. La interacción entre la teoría y los datos ha estimulado el trabajo en nuevas teorías matemáticas sobre la evolución molecular y la genética de poblaciones, que a su vez pueden usarse en la comprobación de hipótesis. De particular importancia ha sido el desarrollo de teorías para probar la hipótesis nula sobre las mutaciones neutras (Nei, 1987). Los marcadores moleculares incluyen todo tipo de secuencias: aquellas que son claramente adaptativas, alternativas alélicas neutras que no afectan la secuencia proteica codificada, secuencias sin función (DNA repetitivo, pseudogenes, por ejemplo) que pueden cambiar libremente, etc.

3. MARCADORES MOLECULARES: CARACTERÍSTICAS Y LIMITACIONES

Los marcadores moleculares de uso generalizado pueden clasificarse en función de la técnica empleada para la obtención de los segmentos discretos de DNA: aquellos que se obtienen tras la fragmentación del genoma correspondiente con endonucleasas de restricción (restringidas), y aquellos que se obtienen por amplificación selectiva de secuencias mediante la reacción en cadena de la polimerasa (PCR, polymerase chain reaction). En algunas técnicas

se combinan estos dos procedimientos básicos. Los polimorfismos obtenidos por el primer procedimiento son conocidos como polimorfismos de longitud de fragmentos de restricción (RFLP, restriction fragment length polymorphisms). En este caso los fragmentos discretos de DNA generados por una restricción son separados por electroforesis en gel, visualizándose selectivamente determinados fragmentos mediante la hibridación con sondas marcadas. En el segundo procedimiento se utilizan cebadores particulares en cada caso para amplificar por replicación selectiva segmentos discretos de DNA. En ambos casos podemos visualizar polimorfismos de segmentos anónimos, cuya secuencia, función y naturaleza nos son desconocidas, o segmentos conocidos, genes conocidos o parte de ellos, secuencias repetidas, etc. Ello depende de la utilización de sondas o cebadores anónimos, elegidos al azar entre miles posibles simplemente porque generan un buen polimorfismo, o que corresponden a secuencias cuya naturaleza conocemos.

Tanto en estudios teóricos de tipo evolutivo como en la utilización práctica de los marcadores moleculares es importante conocer dos de sus características: el carácter dominante o codominante de cada tipo de marcador, y el nivel de polimorfismo que genera. El carácter dominante o codominante determina el nivel de facilidad con que pueden llevarse a cabo estimaciones de parámetros genéticos y evolutivos tan básicos y de uso generalizado como, por ejemplo, frecuencias de recombinación, distancia de mapa, heterocigosidad, distancia evolutiva, etc. El nivel de polimorfismo, y también la naturaleza del propio marcador, determina su utilidad en función del conjunto de organismos a estudiar. Así, por ejemplo, un marcador que genere un nivel bajo de polimorfismo puede ser útil para estudiar especies o géneros dentro de una familia, pero ineficaz para estudiar diferencias entre individuos de una misma población o entre descendientes de un cruzamiento. Esto último es fácil de comprender, si un marcador es poco polimórfico la mayoría o todos los individuos de una población o especie tendrán el mismo fenotipo, lo que lo hace poco útil para diferenciar poblaciones o individuos.

Según Lewontin (1974), inferir la historia de las poblaciones o razas y obtener información sobre los procesos genéticos de la especiación requiere la estimación de frecuencia génicas. La teoría de la Genética de poblaciones es un ejercicio abstracto a no ser que se puedan determinar las frecuencias de alelos alternativos en varios loci, en diferentes poblaciones y en momentos diferentes de la historia de una población. Hoy día disponemos de las técnicas adecuadas para poder hacerlo, pero es obvio que la estimación de las frecuencias génicas es más precisa en genes codominantes que dominantes. Con los primeros los genotipos se observan directamente y la conversión de las frecuencias genotípicas en génicas es inmediata. Con los segundos sólo podemos calcular las frecuencias génicas si se cumplen otros requisitos en la población, por ejemplo que se ajuste a panmixia, lo que no siempre ocurre en particular en especies vegetales, o que los individuos hayan alcanzado una homocigosidad total, lo que por ejemplo es frecuente en especies vegetales autógamas. Pero en cualquier caso, para una muestra del mismo número de individuos, el error estadístico de la estima es mayor con los dominantes que con los codominantes.

Veamos ahora los marcadores moleculares más usados, o más prometedores, en los estudios poblacionales o evolutivos y los de mayor uso aplicado. La Tabla 1 recoge algunas de las características de estos marcadores. De los cuatro, los RFLP son los únicos que no implican alguna reacción de PCR, y fueron los primeros en usarse, generalizándose su uso durante los 80.

Tabla I. Comparación de las características de los sistemas de marcadores moleculares

	RFLP	RAPD	AFLP	SSR
Principio técnico	Restricción Transferencia Hibridación	Amplificación con cebadores al azar	Amplificación limitada de fragmentos	Amplificación por PCR de microsatélites.
Polimorfismos detectados	Camb. puntuales Delecciones Inserciones	Camb. puntuales Delecciones Inserciones	Camb. puntuales Delecciones Inserciones	Nº de repeticiones
Naturaleza	Codominantes	Dominantes	Dominantes	Codominantes
Abundancia	Alta/Muy alta	Muy alta	Alta	
Nivel de polimorfismo	Medio	Medio	Bajo	Alto
Cantidad de DNA requerida	µg	ng	ng	ng
Conocimiento de secuencias	No requerido	No requerido	No requerido	Requerido
Detección con radiactivos	Si/no	No	Si/no	No
Costo	Medio	Bajo	Medio	Alto
Costo puesta a punto	Medio/alto	Bajo	Medio/alto	Alto
Heterocigosidad*	0.41	0.41	0.32	0.60
Multiplicidad*	1-2	5-20	20-100	1

* Estimadas para soja. La multiplicidad hace referencia al número de loci que se pueden estudiar en una sola reacción. Modificado de Rafalski y Tingey, 1993, y Mazur y Tingey, 1995.

Esta técnica detecta las diferencias en longitud de segmentos de DNA y la pérdida o ganancia de dianas para una restrictasa. Los cambios en la movilidad

electroforética de un fragmento de DNA generados por una restrictasa indican diferencias en la longitud del fragmento debidas a inserciones o deleciones. Los cambios en el número de fragmentos junto con la aparición de otros indican la pérdida o ganancia de dianas de restricción para tal enzima causadas generalmente por la sustitución de bases en la diana (Figura 1). En líneas generales la técnica consiste en cortar DNA genómico extraído de una muestra de tejido con una endonucleasa de restricción. Muestras del DNA de cada individuo pueden ser cortadas con diferentes enzimas y analizadas separadamente. Los fragmentos obtenidos se separan por electroforesis en gel, normalmente de agarosa. Puesto que el genoma de un eucarionte superior puede oscilar de 10^9 a 10^{11} pares de bases, una restrictasa genera un enorme número de fragmentos discretos de muy diferentes tamaños. El resultado de la electroforesis es por tanto un rastro continuo de fragmentos, lo que implica la necesidad de usar un método que muestre sólo un conjunto de fragmentos. Para ello se requiere de sondas específicas. Después de la electroforesis los fragmentos son desnaturizados y transferidos a una membrana (blotting) manteniendo el orden y posición relativa del gel. Los fragmentos así fijados se visualizan mediante la hibridación con una sonda marcada. Las sondas pueden haberse generado a partir de cDNA, que representa DNA codificante o de DNA genómico nuclear, codificante o no codificante. En ambos casos las sondas pueden ser de un gen conocido o de genes o segmentos anónimos. Los filtros pueden ser lavados para desprender la sonda y utilizados de nuevo con otra sonda diferente. El resultado es que cada restrictasa y cada sonda genera un patrón discreto de bandas en cada individuo de la población.

La información obtenida con esta técnica depende del número de sondas y del número de restrictasas usadas. Cada sonda hibrida con un conjunto diferente de fragmentos de DNA genómico y cada enzima corta el DNA genómico en puntos diferentes. Los fragmentos de restricción pueden asignarse a loci genéticos, y los datos interpretados en términos genéticos fácilmente, estimando directamente los niveles de variación genética en poblaciones y especies. Muchos de los fragmentos no codificantes pueden ser selectivamente neutros y representar secuencias que divergen más rápidamente que el cDNA. Algunos son muy polimórficos y por tanto útiles para distinguir entre individuos, razas, variedades, etc. Por otro lado las sondas cDNA representan secuencias más conservadas, lo que permite usarlas como marcadores a través de grupos de especies relacionadas. De hecho, restringiendo las condiciones para la hibridación se puede asegurar que una sonda obtenida en una especie hibride con fragmentos homólogos u homeólogos del genoma de otra especie, como ocurre entre maíz, arroz, trigo y otras gramíneas. Para la estimación de datos poblacionales y evolutivos existen numerosos modelos matemáticos (Nei y Miller, 1990; Weir, 1990).

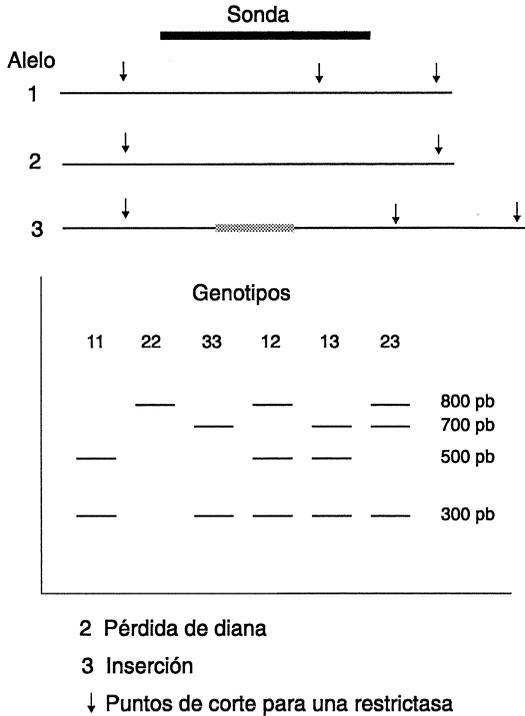


Figura 1. Patrones de bandas RFLP producidos a partir de los genotipos indicados. El alelo 2 difiere del 1 por la pérdida de una diana, y el 3 por la inserción de una secuencia de 100 pb en el fragmento de 700 pb.

Esta propiedad ha sido transcendental para demostrar una característica de los mapas genéticos entre especies vegetales, la sintenia: la conservación de la ordenación relativa entre grupos de genes en diferentes especies incluso lejanamente relacionadas. La construcción de mapas genéticos usando RFLP ha demostrado que, dentro de grandes grupos de especies, como gramíneas o leguminosas, el contenido y la ordenación de los genes se ha conservado, aunque ciertamente se han producido reordenaciones entre bloques de genes a lo largo de la evolución de estos grupos. En las gramíneas, dentro de un brazo o de un segmento cromosómico, el orden de los genes se ha conservado enormemente a pesar de los 60 millones de años de evolución de estas especies (Bennetzen, 1996; Devos et al., 1995). Es cada vez más evidente que dentro de las familias de plantas la sintenia es la norma. Este hecho representa un avance clave en el uso de los mapas genéticos, en el conocimiento de la organización genómica y su evolución, así como en su aplicación a la mejora (Bennetzen, 1996). Un ejemplo: si la sintenia es la norma, conocida la posición de un gen en un mapa genético de una especie con bajo contenido en DNA (garbanzo por ejemplo) es mucho más fácil localizar y clonar el gen homólogo en otra especie de la misma

familia con un mayor contenido en DNA (lenteja), porque marcadores moleculares homólogos flanquearán a dicho gen en ambas especies.

El resto de los marcadores requieren el uso de las técnicas de PCR, por ello haremos un esquema de la técnica (Figura 2). La técnica de reacción en cadena de la polimerasa (PCR) es relativamente simple y consta de tres pasos principales: 1) El DNA de doble hélice se desnaturaliza incrementando la temperatura hasta 94°C durante 1-2 minutos. 2) El segundo paso implica la hibridación específica de los cebadores al DNA molde, por lo que la temperatura se baja hasta unos 50°C (este paso es bastante crítico y la temperatura óptima debe buscarse en cada caso) durante 1 a 1,5 minutos. 3) Por último se sintetizan cadenas complementarias a partir del extremo 3' del cebador, para ello se utiliza una polimerasa termoestable que trabaja mejor a temperaturas próximas a 70-72°C, por lo que se eleva la temperatura y se mantiene durante 2-3 minutos (el tiempo de elongación depende de la longitud de fragmentos que deseamos amplificar). El proceso se repite tantas veces como se desee para obtener el nivel de amplificación requerido (normalmente 30 ciclos son suficientes, en teoría 2^{30} copias por cada fragmento inicial). Los fragmentos amplificados se separan por electroforesis en gel y pueden observarse fácilmente mediante bromuro de etidio y luz UV. La técnica es muy versátil y puede usarse con DNA o con RNA para generar cDNA, incluso se pueden amplificar secuencias de DNA "fósil", como por ejemplo las secuencias del gen *rbcl* de cloroplasto a partir de hojas fósiles de magnolias (Golenberg et al., 1990).

Los cebadores oscilan entre 5 y 25 nucleótidos, se sintetizan químicamente a gusto del usuario o se utilizan juegos disponibles comercialmente. La clase de cebadores usados determina dos técnicas alternativas: la amplificación de secuencias conocidas, o la amplificación de secuencias de DNA anónimas al azar, RAPD (random amplified polymorphic DNA). En la estimación de la variabilidad genética en poblaciones la segunda técnica es, a priori, más versátil puesto que genera un nivel más alto de polimorfismo, aunque requiere un enorme cuidado en su ejecución para evitar resultados poco fiables. En ambos casos se pueden combinar la amplificación mediante PCR y la digestión con restrictasas para aumentar el nivel de polimorfismo. Sobre estas y otras técnicas derivadas se puede consultar el trabajo de Rafalski y Tingey (1993).

La utilización de polimorfismos DNA amplificados al azar es técnicamente sencilla. Se requieren sólo nanogramos de DNA para amplificar segmentos situados entre secuencias repetidas invertidas distribuidas por el genoma. En cada reacción se usa un único cebador sintético corto (normalmente de 10 nucleótidos) de secuencia aleatoria, y disponibles comercialmente. Como los cebadores son cortos es muy probable que encuentren varios lugares complementarios en un genoma de eucariontes. Esencialmente amplifican secuencias de DNA de longitud variable entre repeticiones invertidas cortas. Los productos de amplificación sólo dependen de las combinaciones del cebador y DNA molde (del individuo, especie, etc.). Los polimorfismos generados se comportan como dominantes y se heredan como caracteres mendelianos.

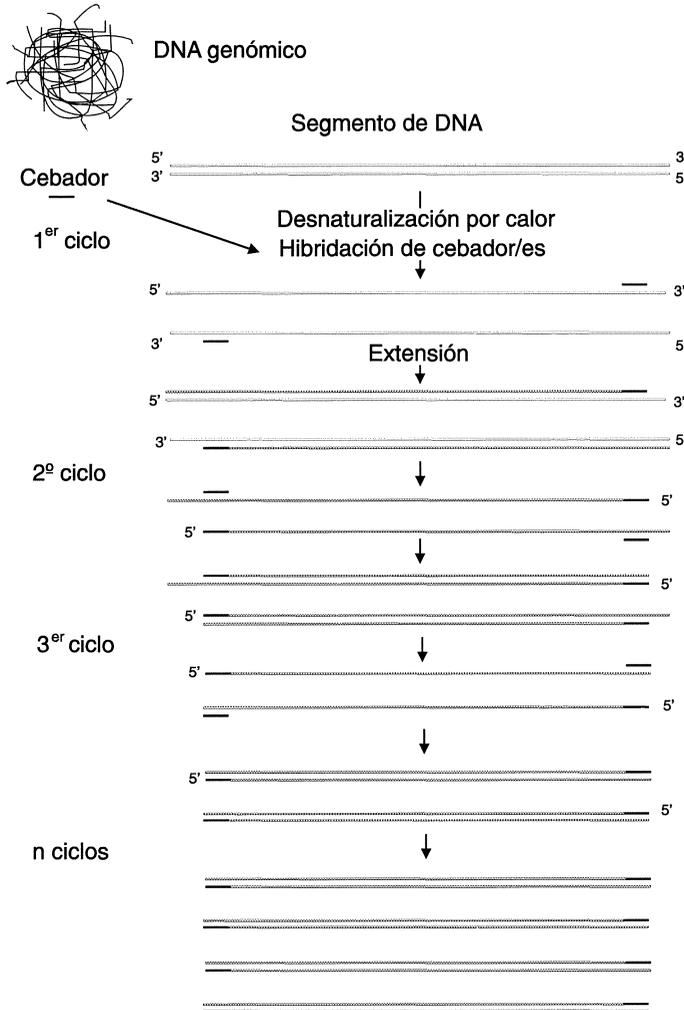


Figura 2. Esquema de la amplificación de secuencias mediante la reacción en cadena de la polimerasa (PCR).

Las ventajas de esta técnica son: 1) Puede usarse en numerosos tipos de estudios como determinación de la variabilidad intraespecífica, flujo genético, etc. 2) No se requiere un conocimiento previo del genoma y puede aplicarse a cualquier DNA no degradado. 3) El número de polimorfismos que, al menos teóricamente, pueden ser estudiados con esta técnica es enorme. 4) Al igual que los RFLP, los loci RAPD se distribuyen aleatoriamente por el genoma, generando una imagen representativa de la distribución de la variación. 5) Es

una técnica rápida, fácil y relativamente barata. Como ya se ha indicado, el mayor inconveniente técnico es que requiere una gran precisión en el mantenimiento de las variables técnicas para evitar errores y resultados no interpretables, puesto que ligeros cambios en las condiciones pueden producir resultados irrepetibles. Otro inconveniente es que secuencias de tamaño semejante migran en los geles a la misma velocidad, por tanto, secuencias no relacionadas pero de igual longitud pueden interpretarse como idénticas. La probabilidad de que un mismo cebador amplifique dos secuencias no relacionadas pero que por azar tengan una longitud similar es mayor a medida que la relación filogenética de los individuos disminuye. Por eso los marcadores RAPDs son útiles para comparar poblaciones y especies congénicas, pero es dudoso que lo sean para comparar entre niveles taxonómicos superiores.

El tercer tipo de marcador incluido son los polimorfismos de longitud amplificados (AFLP, amplified fragment length polymorphism) y se basan en la generación de fragmentos de restricción mediante una restrictasa y la amplificación selectiva mediante PCR de un subconjunto de éstos (Figura 3). La técnica requiere de la unión de unos adaptadores a los fragmentos generados por la restrictasa y la amplificación por PCR mediante cebadores que incluyen la secuencia del adaptador, de la diana de la restrictasa y algunos otros nucleótidos en el extremo 3', lo que determina que cada cebador sólo amplifique un subconjunto de todos los fragmentos. Como consecuencia se puede generar una gran cantidad de polimorfismos combinando distintas restrictasas y diferentes secuencias 3' de los cebadores.

El último tipo de marcadores incluido se basa en la amplificación mediante PCR de mini- o microsatélites, es decir, secuencias cortas que se repiten en tandem (por ejemplo, $(CAC)_n$ o $(GT)_n$) y que se encuentran distribuidas por el genoma. Este tipo de secuencias cortas son denominadas también secuencias simples lo que ha generado su designación de SSR (simple sequence repeats) aunque también se han denominado como VNTR. El polimorfismo se basa en este caso en el diferente número de repeticiones de los distintos alelos, lo que se detecta por la distinta movilidad en gel de los fragmentos amplificados usando como cebadores complementarios a secuencias únicas a ambos lados de las repeticiones. Esta técnica es fácil de ejecutar pero requiere pasos previos laboriosos para identificar los DNA satélites y los cebadores a usar.

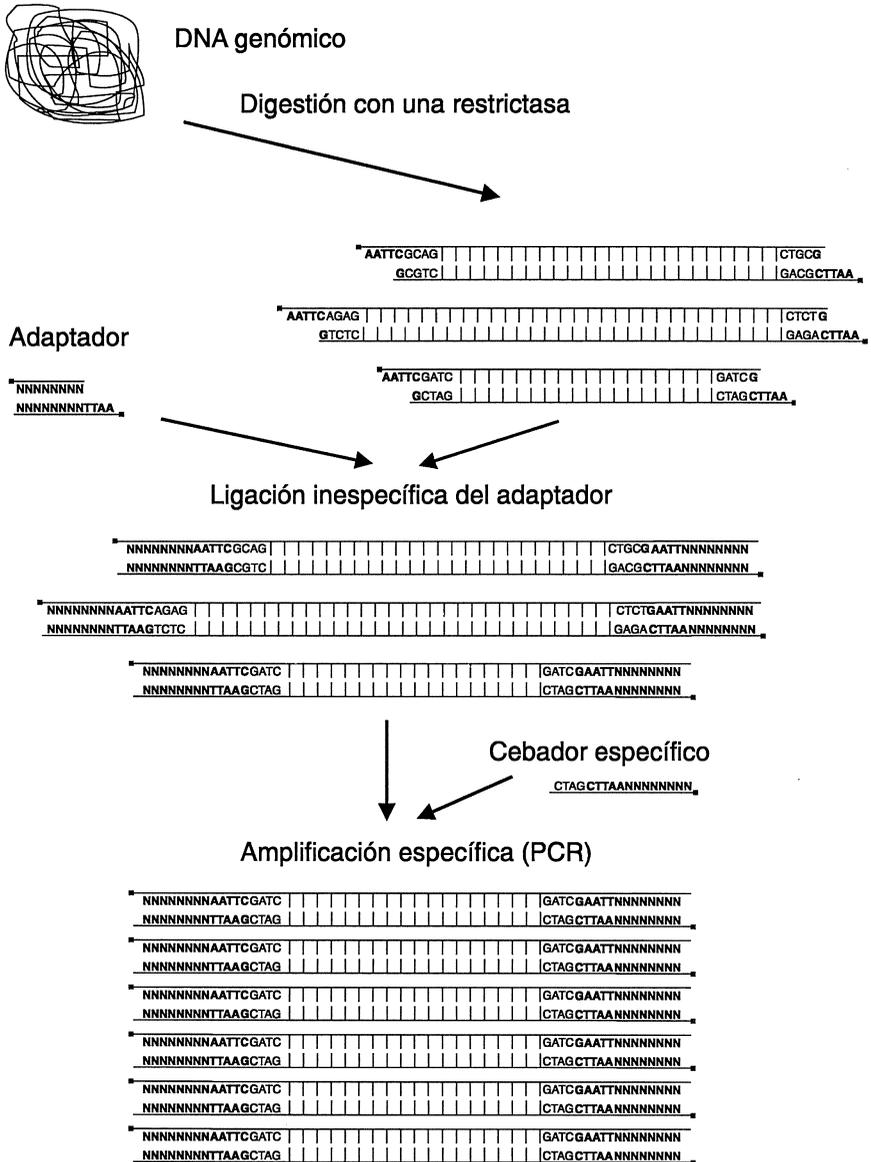


Figura 3. Esquema de la generación de fragmentos AFLP. En este esquema el cebador específico amplifica sólo el fragmento más corto. Los polimorfismos se generan por razones similares a los de RFLP.

Un aspecto a tener en cuenta es la utilidad de los distintos tipos de marcadores moleculares en estudios evolutivos. Comentaremos la características de los más ampliamente usados en la actualidad. Los RFLP son apropiados para establecer relaciones filogenéticas entre niveles taxonómicos dentro de familias, y en algunos casos incluso entre familias. Ello se debe a que restringiendo las condiciones de hibridación de las sondas, éstas sólo hibridan con secuencias cuya homología asegura un origen evolutivo común. Los RAPD, como se ha mencionado antes, son dudosamente útiles en comparaciones entre taxones por encima del género, pero lo son a niveles intraespecíficos e intragenéricos. Por ejemplo, recientes trabajos en *Brassica napus* y *B. oleracea* indican que los RAPD dan un nivel de resolución equivalente al de RFLP en la determinación de relaciones genéticas intraespecíficas (Halldén et al., 1994; dos Santos et al., 1994), o son capaces de identificar distintas especies dentro del género *Centrosema* (Penteado et al., 1996).

El mejor nivel de estudio y la más completa información evolutiva se obtiene de la comparación de secuencias de nucleótidos. Ahora que las técnicas de secuenciación se han generalizado y automatizado nos podemos preguntar si los marcadores siguen teniendo sentido en estudios evolutivos, puesto que aportan una información menos completa. Citaré a Avise (1994) para justificar su utilidad: "...los datos genéticos definitivos son las secuencias mismas. Sin embargo, sería una equivocación concluir que la información de secuenciación suministra invariablemente el conjunto preferible y más accesible de marcadores genéticos para todas las aplicaciones biológicas. Varios métodos alternativos son aún de un enorme poder y utilidad y debido a su facilidad, coste, la cantidad de información genética accesible, y facilidad de interpretación de los datos, continúan siendo las técnicas a elegir para muchos problemas evolutivos".

4. INFORMACIÓN OBTENIDA CON LOS MARCADORES

En los estudios evolutivos son importantes dos aspectos metodológicos: el muestreo de material biológico y la elección del marcador. Hay dos criterios para un muestreo adecuado. El primer criterio es el ecogeográfico. En el estudio deben estar presentes muestras representativas de todo el rango ecológico y de distribución del taxón de que se trate. El segundo criterio de muestreo es el genómico (Gepts, 1993): genoma nuclear, o mitocondrial, o de cloroplasto en su caso. Tanto el DNA mitocondrial (mtDNA) como el de cloroplasto (cpDNA) se han usado extensivamente en estudios filogenéticos y evolutivos, aunque los genomas nucleares son preferidos con diferencia, porque en general evolucionan más rápidamente que los de orgánulos y porque su mucho mayor tamaño genómico proporciona un número mucho más alto de polimorfismos. Por otra parte, la información evolutiva que ambos tipos de genomas, orgánulos y nuclear, aportan es distinta y complementaria en aquellas especies (la mayoría) en que los orgánulos citoplásmicos son heredados uniparentalmente. En esta situación la recombinación no es un mecanismo generador de variación y todos los descendientes del parental transmisor (generalmente el materno) son

idénticos entre sí y con su genitor en cuanto a los genomas citoplásmicos. El DNA mitocondrial ha sido una fuente muy útil de información evolutiva en animales, incluida nuestra propia especie como veremos posteriormente.

Tomando como ejemplo las plantas, podemos ver que la dinámica evolutiva peculiar del su mtDNA, caracterizado por niveles altos de reordenaciones, baja tasa de mutaciones puntuales y la incorporación de secuencias externas (por ejemplo secuencias del genoma de cloroplasto), hacen que sea difícil establecer filogenias en base a la utilización de RFLP; aunque los rápidos cambios debidos a reordenaciones hacen que el mtDNA sea útil en la búsqueda intraespecífica de tipos genómicos o de variación somaclonal generada por cultivo *in vitro*. Por ejemplo, ha sido posible detectar variación citoplásmica en un único cultivar de *Lolium perenne* mediante RFLP del mtDNA (Sato et al., 1995). Por el contrario, el tamaño y el orden de los genes en el cpDNA es muy conservativo, pero la tasa de sustitución de nucleótidos es mayor que en el mtDNA, pero lo suficientemente baja como para que los polimorfismos a nivel de cpDNA hayan sido muy útiles en los análisis filogenéticos a nivel de especies y taxones superiores en muchas familias de plantas. Como ejemplos se pueden citar revisiones filogenéticas en gramíneas, leguminosas o coníferas (Doyle et al., 1992; Doyle y Doyle, 1993, Tsumura et al., 1995), o estudios sobre introgresión genética y expansión post-glacial en Europa de especies del género *Quercus* (Ferris et al., 1993).

Otros ejemplos de la utilización de marcadores moleculares pueden encontrarse en las revisiones incluidas en la bibliografía. Pero he escogido la polémica sobre la 'Eva mitocondrial' como un ejemplo significativo de unos datos evolutivos que a la vez han aportado nueva luz sobre la evolución y promovido nuevos estudios para refutar o confirmar las conclusiones iniciales. El DNA mitocondrial de primates tiene unas 16.500 pb y 37 genes, es de herencia materna y en él se producen frecuentemente mutaciones de tipo neutro. Por tanto la evolución del DNA mitocondrial puede estudiarse según un modelo sencillo en el que sólo depende de una tasa constante de mutación (Wilson y Cann, 1992). Los estudios mediante RFLP, y posteriormente mediante la secuenciación de algunos segmentos característicos, de mitocondrias de distintos grupos étnicos humanos llevaron a la conclusión de que los seres humanos actuales son todos descendientes de una mujer africana que vivió hace no más de 200.000 años (Cann et al., 1987, Wilson y Cann, 1992). Un origen 'unimaterno' reciente parece indicar además la sustitución rápida y total de los humanos antiguos por otros más modernos. Todas estas conclusiones no fueron aceptadas por otros antropólogos y evolucionistas y desde 1987 la polémica sobre la realidad de la 'Eva mitocondrial' sigue abierta. Posteriormente se ha incluido en este estudio la contrapartida masculina: secuencias específicas del cromosoma Y que, obviamente, se heredan vía paterna y tampoco están sujetas a fenómenos de recombinación al carecer de secuencia homóloga en el cromosoma X. Los datos con las secuencias del cromosoma Y indican unos 270.000 años como la fecha más probable para un progenitor común. La 'hipótesis de Eva' sin embargo es resultado de la confusión entre genealogías gónicas y genealogías individuales. Ayala en un reciente artículo (1995), basado en secuencias de genes para el sistema inmunológico, concluye que la población ancestral humana nunca ha

sufrido una reducción tan drástica (una o unas pocas parejas) y que el tamaño efectivo ha sido de unos 100.000 individuos o mayor durante varios millones de años. En sus conclusiones Ayala (1995) indica que los datos moleculares favorecen un origen africano reciente de los humanos modernos; que las diferenciaciones étnicas entre los humanos modernos son evolutivamente recientes (entre unos 50.000 a 100.000 años). Sin embargo indica que la sustitución de los *Homo sapiens* arcaicos por los anatómicamente modernos pudo no ser completa en todos los lugares, y que el mestizaje entre los humanos modernos y las poblaciones locales puede explicar la aparente continuidad morfológicas en algunas regiones.

Como resumen de los conocimientos que hemos adquirido sobre el nivel de variación genética en los seres vivos gracias a los marcadores moleculares, y también a otros estudios moleculares, podemos citar estas tres conclusiones:

- 1) La cantidad de variación a nivel de DNA en las especies es enorme.
- 2) La tasa de cambio varía considerablemente según las distintas secuencias del genoma.
- 3) Cuanto más importante es la función de una secuencia menor es su tasa de cambio.

5. APLICACIONES DE LOS MARCADORES.

Los marcadores moleculares no sólo son útiles en estudios teóricos de tipo poblacional o evolutivo. Los conocimientos que con ellos hemos obtenido, y los propios marcadores como herramientas, son útiles en trabajos aplicados. Son una herramienta cada vez más útil y utilizada en mejora genética, particularmente en plantas. Por otra parte la información obtenida de ellos nos permite diseñar mejores formas de conservar la propia variabilidad genética, la biodiversidad.

Respecto a la utilidad de los marcadores en mejora sólo haré una enumeración de sus distintos usos en mejora vegetal, puesto que algunas de estas aplicaciones son descritas con más detalle en otro capítulo. 1) Identificación de plantas transgénicas, 2) identificación de mutantes somaclonales generados por cultivo *in vitro* de tejidos, 3) como 'etiquetas' marcadores de genes cualitativos y cuantitativos, 4) como ayuda para aislar y clonar genes, 5) para construir mapas genéticos saturados, 6) en la selección asistida por marcadores, 7) para caracterizar poblaciones y razas de patógenos, 8) en el estudio del flujo genético desde individuos transgénicos y su dispersión de en la naturaleza.

Un aspecto en el que los marcadores moleculares están aumentando espectacularmente su importancia es en el de la Biología y la Genética conservacionista. Es cada vez mayor la preocupación sobre la conservación de la biodiversidad, lo que en definitiva representa la conservación de diversidad genética. Avise (1994) dedica un capítulo de su libro a describir el uso de marcadores en aspectos tales como estructura de las poblaciones y partición de la variabilidad en especies fragmentadas y amenazadas de extinción, identificación de estas especies, identificación de poblaciones, de reservas pesqueras, diseño

racional de reservas naturales, etc. Sin embargo, desde mi punto de vista, Avise comete un error bastante común, iguala heterocigosidad con diversidad, probablemente porque sólo piensa en animales. Lo realmente importante es la diversidad, las numerosísimas especies vegetales autóгамas, debido al sistema natural de autofecundación, mantienen un alto grado de diversidad entre individuos que son homocigóticos. Mantienen variabilidad en ausencia de heterocigosidad. Es claro que no siempre la presencia de dos alelos distintos es la mejor solución adaptativa (García et al., 1991).

Un ejemplo del uso de los marcadores moleculares en estudios filogenéticos y taxonómicos, y de las importantes repercusiones que están teniendo en la percepción de la diversidad biológica y su conservación, puede verse en la reciente reseña de G. Martin (1996) sobre la diversidad de especies de aves. Así el uso de marcadores moleculares ha llevado a algunos taxónomos a sugerir la posible duplicación del número reconocido de especies de aves. Una lista mundial de 20.000 especies, en oposición a la actual de 10.000, se ha convertido en una perspectiva muy real a medida que las técnicas de DNA amenazan con desplazar al libro de notas de campo como la herramienta más importante en la taxonomía aviar.

Hasta ahora para los ornitólogos las especies parecían estar bien definidas y, aparte de algunos problemas de menor cuantía, existía un acuerdo general sobre la lista de especies, aunque las relaciones evolutivas entre ellas estaban sometidas a debate. Esta certeza sobre las especies ha facilitado un rápido intercambio internacional de información. Las aves han sido adoptadas como un paradigma del conservacionismo, y mucha de la legislación nacional e internacional utiliza a las especies de aves como ejemplo de los temas ecológicos. Esta aparente certeza sobre lo que es una especie de aves se ha cimentado, sin embargo, en un suelo poco firme, y han surgido opiniones contrapuestas que amenazan con polarizarse, y el hacerlo tiene implicaciones más allá de los límites de la ornitología científica.

La utilización de los polimorfismos basados en el mtDNA representa un ejemplo de marcadores que han revolucionado la visión sobre el número de especies. Como ejemplo Martin (1996) se pregunta cuántos pájaros carpinteros de tres dedos hay. *Picoides tridactylus* tiene una amplia distribución desde los Alpes, por Escandinavia y el norte de Rusia hasta el Pacífico y las latitudes septentrionales de Norteamérica. Actualmente se la considera como una única especie dividida en unas 12 subespecies. Pero el análisis del mtDNA de individuos de cuatro de estas subespecies, a ambos lados del estrecho de Bering, ha demostrado diferencias suficientemente grandes en el mtDNA como para sugerir que estas subespecies deberían ser reconocidas como especies distintas.

La implicación que tienen estos resultados no es baladí. Porque si entendemos la taxonomía como una herramienta para comprender mejor la diversidad biológica, no como un fin en sí misma, entonces las consecuencias de aplicar este tipo de resultados junto con el controvertido concepto de especie (Stebbins, 1987; Stebbins y Pérez de la Vega, 1989) son trascendentales. La adopción del concepto filogenético de especie lleva inequívocamente a la pérdida del concepto de subespecie, bien al agrupar las subespecies en nuevos

grupos de especies o bien elevando las subespecies al nivel de especies. En el caso de muchas de las especies actualmente admitidas que tienen un rango de distribución geográfica y mucha diversidad de formas, la agrupación de subespecies llevaría a una pérdida del reconocimiento de tal diversidad. Por otra parte, el elevar muchas subespecies al nivel de especie resultaría en una plétora de nuevas especies, cada una con una distribución geográfica más pequeña, y quizás no distinguibles en el campo. Esto podría llevar, por ejemplo, a duplicar la lista mundial de especies de aves hasta unas 20.000.

Duplicar el número de especies de aves o de otras especies afectaría el estatus de éstas en la conservación internacional de la vida salvaje. La introducción de un número mayor de especies definidas por características difíciles o imposibles de detectar en el campo, y que pueden tener una distribución geográfica limitada, haría imposible aplicar mucha de la legislación existente en la protección y conservación nacional e internacional de especies. Ello cambiaría dramáticamente la unidad de biodiversidad, y haría más difícil la aplicación y seguimiento de las medidas de conservación. Pero claramente, ésta es una controversia que aún está empezando.

Otro aspecto de observar la biodiversidad es entendiéndola como recurso genético, que de acuerdo con la F.A.O. se define como "el material hereditario con valor económico, científico o social contenido en las especies". Es claro que este concepto engloba a todas las especies que real o potencialmente son útiles a la humanidad, lo que bien entendido engloba a la inmensa mayoría de la biodiversidad. De nuevo tomando como ejemplo las plantas, los marcadores moleculares nos están ayudando a conocer mejor la variabilidad genética de las especies cultivadas y las silvestres afines, a estudiar las relaciones filogenéticas entre unas y otras, a conservar la variabilidad que se está viendo sometida a un proceso progresivamente acelerado de erosión genética, en fin, a una mejor y más racional utilización de la variabilidad en la mejora genética (Bretting y Widrechner, 1995; Lavi et al., 1994; Lee, 1995). Es éste un campo en el que los conocimientos poblacional-evolutivos teóricos convergen con los técnicos aplicados, y cuya expansión es una de las más rápidas en el campo de la fitogenética.

6. REFERENCIAS

- Arús, P. & J. Moreno-González (1993). Marker-assisted selection. En *'Plant Breeding: Principles and Prospects'*. M. D. Hayward, N. O. Bosemark & I. Romagosa Eds. Chapman & Hall. London. pp 314-331.
- Avise, J. C. (1994). *Molecular Markers, Natural History, and Evolution*. Chapman & Hall, New York.
- Ayala, F. J. (1995). The myth of Eve: Molecular Biology and human origins. *Science*, 270:1930-1936.
- Bennetzen. J. L. (1996). The use of comparative genome mapping in the identification, cloning and manipulation of important plant genes. En *'The*

- Impact of Plant Molecular Genetics'*. B. W. S. Sobral (Ed.). Birhäuser, Boston. pp. 71-85.
- Bretting, P. K. & M. P. Widrlechner (1995). Genetic markers and plant genetic resources management. *Plant Breed. Rev.*, 13: 11-86.
- Cann, R. L., M. Stoneking & A.C. Wilson (1987). Mitochondrial DNA and human evolution. *Nature*, 325:31-36.
- Devos, K. M., G. Moore & M. D. Gale (1995). Conservation of marker synteny during evolution. *Euphytica*, 85:367-372.
- Doyle, J. F., J. I. Davis, R. J. Soreng, D. Garvin & M. J. Anderson (1992). Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proc. Natl. Acad. Sci. USA*, 89:7722-7726.
- Doyle, J. F. & J. L. Doyle (1993). Chloroplast DNA phylogeny of the Papilionoid legume tribe Phaseoleae. *Syst. Bot.*, 18:309-327.
- Endler, J. A. (1986). Natural Selection in the Wild. *Monographs in Population Biology* 21. Princeton Univ. Press. Princeton, New Jersey.
- Ferris, C., R. P. Oliver, A. J. Davy & G. M. Hewitt (1993). Native oak chloroplast reveal an ancient divide across Europe. *Mol. Ecol.*, 2:337-344.
- Gale, M. D. (1994). Genetic markers, maps and wheat breeding. *J. R. Agric. Soc. Engl.*, 155:162-176.
- García, P., M. I. Morris, L. E. Sáenz de Miera, R. W. Allard & M. Pérez de la Vega (1991). Genetic diversity and adaptedness in tetraploid *Avena barbata* and its diploid ancestors *Avena hirtula* and *Avena weistii*. *Proc. Natl. Acad. Sci. USA* 88:1207-1211.
- Gepts, P. (1993). The use of molecular and biochemical markers in crop evolution studies. *Evol. Biol.*, 27:51-94.
- Golenberg, E. M., D. E. Giannasi, M. T. Clegg, C. J. Smiley, M. Durbin, D. Henderson & G. Zurawski (1990). Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature*, 334:656-658.
- Halldén, C., N.-O. Nilsson, L. M. Radding & T. Säll (1994). Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. *Theor. Appl. Genet.*; 88:123-128.
- Lavi, U., P. Cregan, T. Schaap & J. Hillel (1994). Application of DNA markers for identification and breeding of perennial fruit crops. *Plant Breed. Rev.*, 12:195-226.
- Lee, M. (1995). DNA markers and plant breeding programs. *Adv. Agron.*, 55:265-344.

- Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. Columbia Univ. Press, New York.
- Martin, G. (1996). Birds in double trouble. *Nature*, 380:666-667.
- Mazur, B. J. & S. V. Tingey (1995). Genetic mapping and introgression of genes of agronomic importance. *Curr. Opinion Biotechnol.*, 6:175-182.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia Univ. Press. New York.
- Nei, M. & J. C. Miller (1990). A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics*, 125:873-879.
- Penteado, M. I. de O., P. García & M. Pérez de la Vega (1996). Genetic variability and mating system in three species of the genus *Centrosema*. *J. Hered.*, 87:124-130.
- Pérez de la Vega, M. (1993). Biochemical characterization of populations. En '*Plant Breeding: Principles and Prospects*'. M. D. Hayward, N. O. Bosermark & I. Romagosa Eds. Chapman & Hall. London. pp 184-200.
- Pérez de la Vega, M. (1996). Plant genetic adaptedness to climatic and edaphic environment. *Euphytica*, (en prensa).
- Rafalski, J. & S. V. Tingey (1993). Genetic diagnosis in plant breeding: RAPDs, microsatellites and machines. *Trends Genet.*, 9:275-280.
- Rieger, R., A. Michealis & M. M. Green (1982). *Diccionario de Genética y Citogenética Clásica y Molecular*. Alhambra, Madrid.
- Santos, J. B. dos, J. Nienhuis, P. Skroch, J. Tivang & M. K. Slocum (1994). Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. *Theor. Appl. Genet.*, 87:909-915.
- Sato, M., M. Yamashita, S. Kato, T. Mikami & Y. Shimamoto (1995). Mitochondrial DNA analysis reveals cytoplasmic variation within a single cultivar of perennial ryegrass (*Lolium perenne* L.). *Euphytica*, 83:205-208.
- Stebbins, G. L. (1987). Species concepts: Semantics and actual situations. *Biol. Philos.*, 1:76-82.
- Stebbins, G. L. & M. Pérez de la Vega. (1989). Estimación de la variabilidad genética. En '*Evolución hacia una Nueva Síntesis. Contribuciones desde el Reino Vegetal*'. G. L. Stebbins & M. Pérez de la Vega. Sec. Publ. Univ. de León, León. pp. 63-81.

- Tsumura, Y., K. Yoshimura, N. Tomura & Y. Ohba (1995). Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. *Theor. Appl. Genet.*, 91:1222-1236.
- Weir, B. S. (1990). *Genetic Data Analysis*. Sinauer Assoc., Sunderland, Massachusetts.
- Wilson, A. C. & R. L. Cann (1992). Origen africano reciente de los humanos. *Investigación y Ciencia*, N° 189:8-13.