

Tercera parte

El conocimiento de la variabilidad genética

14 La variabilidad genética de las especies: aspectos conceptuales y sus aplicaciones y perspectivas en México

AUTOR RESPONSABLE: Daniel Piñero

COAUTORES: Ana Barahona • Luis Eguiarte • Axayácatl Rocha Olivares • Rodolfo Salas Lizana

REVISORES: Píndaro Díaz • Eduardo Morales Guillaumin • Daniel Zizumbo-Villarreal

CONTENIDO

- 14.1 Las "escuelas" de genética en México y su influencia en el estudio de la variabilidad / 416
 - 14.1.1 La introducción del mendelismo en México / 417
 - 14.1.2 Genética y mejoramiento vegetal / 417
 - 14.1.3 Institucionalización de la genética / 418
- 14.2 Estimados de la variación con caracteres moleculares / 419
- 14.3 Estimados de la variación de caracteres cuantitativos / 419
- 14.4 Estimados de la estructura genética / 420
 - 14.4.1 Introducción / 420
 - 14.4.2 Estadísticos F de Wright y estimadores análogos / 420
 - 14.4.3 Flujo génico / 421
 - 14.4.4 Métodos directos para estimar el flujo génico / 422
 - 14.4.5 Métodos indirectos para estimar el flujo génico / 422
 - 14.4.6 Distancias genéticas / 423
 - 14.4.7 Aislamiento por distancia / 423
 - 14.4.8 Inferencia de estructura y proporción ancestral / 424
- 14.5 La teoría de coalescencia / 424
 - 14.5.1 Aplicaciones de la teoría de coalescencia / 424
 - 14.5.2 Programas más comúnmente usados para hacer inferencias usando la teoría de coalescencia / 425
 - 14.5.3 Perspectivas de la teoría de coalescencia en México / 425
- 14.6 Filogeografía / 425
 - 14.6.1 Origen y desarrollo / 425
 - 14.6.2 Principios y teoría / 426
 - 14.6.3 Concordancia genealógica / 427
 - 14.6.4 Perspectivas / 430
- 14.7 Infraestructura y grupos de investigación / 431
- 14.8 Conclusiones / 431
- Referencias / 432

Piñero, D., et al. 2008. La variabilidad genética de las especies: aspectos conceptuales y sus aplicaciones y perspectivas en México, en *Capital natural de México*, vol. I: *Conocimiento actual de la biodiversidad*. CONABIO, México, pp. 415-435.

Resumen

Un aspecto fundamental de la biodiversidad es la variación genética de las poblaciones y las especies. Esta variación puede ser clasificada en cuatro tipos en función de sus consecuencias en la adecuación: variación de alelos letales, deletéreos o neutros, y la asociada con alelos que aumentan la adecuación de sus portadores. Asimismo se pueden dividir los tipos de variación genética en aquellos que estudian rasgos cualitativos (determinados por un *locus*) y los que son cuantitativos.

Este capítulo está dividido en siete partes. La primera se refiere a los diferentes enfoques o escuelas con que se aborda el estudio de la genética en México y su influencia en el estudio de la variabilidad, incluyendo la agricultura, la medicina y la biología de la conservación. La segunda parte trata los estimados de variación en caracteres moleculares; en esta parte se incluyen las preguntas más importantes que se pueden contestar usando marcadores moleculares mediante los estimados de variación como (θ) theta y (π) pi. La tercera tiene que ver con los estimados de variación en caracteres cuantitativos. La cuarta con

los estimados sobre la estructura genética. Estos enfoques incluyen estimados como F_{ST} , y pruebas de hipótesis del modelo de aislamiento por distancia. La quinta se ocupa de los enfoques asociados a la teoría de coalescencia y sus aplicaciones tanto en la biología de la conservación como en la teoría evolutiva por medio de estadísticos de resumen y simulaciones de procesos coalescentes bajo supuestos de deriva, selección, escenarios demográficos alternos y distintos modelos de recombinación. La sexta se refiere a los enfoques y métodos de la filogeografía, con los que se pueden hacer inferencias acerca de los escenarios pasados en la historia de estos linajes utilizando la llamada filogeografía estadística. Finalmente se presenta la infraestructura y los grupos de investigación existentes en México para generar información necesaria sobre políticas de apoyo. Este capítulo se complementa con el capítulo 15 de este volumen, que incluye los resultados obtenidos en diversos grupos de especies mexicanas y que muestran también un resumen de los descubrimientos que se han hecho en este campo en especies mexicanas, tanto útiles como claves en su ecosistema y raras o en peligro de extinción.

En este capítulo se proporciona una visión de los enfoques que se han utilizado para estudiar la variación genética en especies de México y sobre las instituciones, acervos, metodologías y enfoques conceptuales que han contribuido a este conocimiento. Durante los últimos años se ha hecho un enorme esfuerzo para ampliar el conocimiento de la estructura genética en poblaciones naturales en México. Estas contribuciones abarcan muy diversos grupos taxonómicos y están asociadas a las diferentes ramas de la genética en México, así como al desarrollo de la infraestructura necesaria para generar la información acerca de la diversidad genética en especies mexicanas. Las preguntas más importantes que se han planteado incluyen la cantidad de variación que hay en poblaciones naturales, la estructuración de esa variación en diversas poblaciones y las consecuencias para la adaptación, la especiación y la extinción que la rareza de especies tiene. Estos estudios han arrojado en muchos casos respuestas que complementan y amplían las conclusiones obtenidas por estudios sistemáticos y taxonómicos de la biodiversidad. Por ejemplo, algunos muestran divergencias genéticas profundas en grupos morfológicamente homogéneos, mientras que otros muestran homogeneidad genética en grupos que a priori se han mantenido separados. En particular, en este capítulo se abordan:

1] las “escuelas” de genética en México y su influencia en el estudio de la variabilidad; 2] estimados o parámetros usados para conocer la variación basados en caracteres moleculares; 3] estimados o parámetros de variación en caracteres cuantitativos; 4] estimados de los niveles de estructura genética; 5] filogeografía; 6] coalescencia y modelos de mutación, y 7] infraestructura y grupos de investigación.

14.1 LAS “ESCUELAS” DE GENÉTICA EN MÉXICO Y SU INFLUENCIA EN EL ESTUDIO DE LA VARIABILIDAD

Las escuelas de genética en México se pueden agrupar y distinguir por su relación con la medicina, la agricultura, la conservación o la biotecnología y la genómica. En ellas se han establecido líneas de investigación basadas en la información generada en torno a variación genética, que tratan problemas particulares. La integración de estos grupos de investigación ha sido fundamental para el entendimiento de las especies mexicanas, y su inclusión dentro de grupos que estudian aspectos de interés para la sociedad mexicana ha sido su motivación fundamental.

14.1.1 La introducción del mendelismo en México

Las primeras investigaciones acerca de la herencia biológica en México fueron motivadas por el interés que los naturalistas y los médicos tenían en el papel de la herencia en la determinación de algunos rasgos humanos y, sobre todo, en las enfermedades. Estas investigaciones incluyeron, hacia finales del siglo XIX, el efecto del ambiente en la herencia de las malformaciones, como el consumo de algún tipo de drogas, y la influencia de matrimonios entre parientes.

Durante las primeras décadas del siglo XX, al menos en las publicaciones médicas y en los primeros libros de texto de biología, ya se habla de “leyes de la herencia” o “hipótesis sobre la herencia”. Alfonso L. Herrera, considerado el fundador de la biología en México, estableció en 1902 la primera cátedra de biología en México, en la Escuela Normal de Maestros, y en 1904 publicó *Nociones de biología* (Herrera 1904), obra que significó la primera introducción científica a la biología moderna y al darwinismo en México (Ledesma y Barahona 2003).

Sin embargo, este conocimiento no dio lugar a que se iniciara programa alguno de investigación sistemática en genética. No sería sino hasta la década de 1930 cuando la genética realmente se introduce en la agricultura en México, y con mayor fuerza a partir de la década de 1940.

14.1.2 Genética y mejoramiento vegetal

La reforma agraria en el campo mexicano surgida de la Revolución (1910-1917) se expresó en los programas políticos de los gobiernos posrevolucionarios para la recuperación del agro e incluyó la educación agrícola como un detonador importante, fomentándose así la formación de agrónomos y técnicos con la creación de la Escuela Nacional de Agricultura (ENA, actual Universidad Autónoma Chapingo en Texcoco, Estado de México) que en 1907 pasaría a formar parte de la Secretaría de Agricultura y Fomento.¹

El ingeniero agrónomo Edmundo Taboada Ramírez (1906-1983), egresado de la ENA, fue “el primer técnico mexicano en agronomía que tuvo la oportunidad de hacer estudios de postgrado durante 1932 y 1933 en la Universidad de Cornell, N.Y., sobre genética vegetal, y en la de Minnesota con el doctor E.C. Stackman, en parasitología vegetal” (INIA 1985), específicamente en el chahuixtle del trigo. Desde 1936 se incorporó a la ENA, donde fue uno de los primeros catedráticos de genética vegetal y

aplicada y el primer autor en escribir un libro de texto sobre genética general, *Apuntes de genética*, en donde explica de manera muy clara y precisa la teoría cromosómica de la herencia y las leyes de Mendel (Taboada 1938). Taboada desarrolló un método experimental basado en las leyes de la herencia y su mayor contribución fue la creación del denominado maíz “estabilizado”. A finales de los años 1940 se contaba con dos programas de mejoramiento del maíz, uno a cargo de Taboada y el otro a cargo de Eduardo Limón en la Estación Experimental de León, Guanajuato.

El grupo encabezado por Taboada, que creó los maíces estabilizados que compitieron favorablemente con los maíces híbridos, se formó y llevó a cabo sus investigaciones en los campos experimentales de la Secretaría de Agricultura fundados en 1933. Este conjunto de campos se convirtió más tarde en el Departamento de Campos Experimentales (DCE), y hacia finales de los cuarenta, en el Instituto de Investigaciones Agrícolas (IIA) (Barahona y Gaona 2001; Barahona *et al.* 2003).

Al comenzar la década de 1940 se formó otro grupo de investigadores mexicanos en agricultura integrados en el llamado Programa Agrícola Mexicano, mediante una colaboración entre el gobierno de nuestro país y la Fundación Rockefeller de Estados Unidos.²

El Programa Agrícola Mexicano se puso en marcha en 1943 con la finalidad principal de desarrollar investigación básica sobre métodos y materiales de utilidad para incrementar los cultivos básicos, y dar mayor énfasis a la formación y adiestramiento de profesionales. Surgió así en 1944 la Oficina de Estudios Especiales (OEE) que se dedicó principalmente al mejoramiento del maíz y el trigo, así como a la introducción de un paquete tecnológico de insumos y prácticas, semillas mejoradas, fertilizantes químicos, insecticidas y herbicidas, y riego, elementos necesarios para explotar el rendimiento de nuevas variedades mejoradas genéticamente. Esta revolución en la producción agrícola fue exportada más tarde a otros países en desarrollo (Fitzgerald 1994).³

Finalmente, en 1960 ambos grupos se unieron para formar el Instituto Nacional de Investigaciones Agrícolas (INIA). En 1963 fue creado el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) en Texcoco, Estado de México, gracias al apoyo de la Fundación Rockefeller y de la Fundación Ford, y como producto de las investigaciones llevadas a cabo por la OEE y el INIA. El CIMMYT es un centro líder en innovación de maíz y trigo, y en el desarrollo de tecnología y conocimiento práctico tendientes a incrementar la productividad de los sistemas agrológicos y el desarrollo sustentable.⁴

14.1.3 Institucionalización de la genética

En 1960 Alfonso León de Garay fundó el Programa de Genética y Radiobiología como parte de la Comisión Nacional de Energía Nuclear (creada en 1956), cuyos principales objetivos eran, por un lado, el estudio de los efectos de las radiaciones en la salud humana, y por otro, el estudio de diversos aspectos específicos del proceso hereditario, desde el nivel molecular hasta la genética de poblaciones (De Garay 1960). La primera cátedra en genética fue establecida en 1961 en la Facultad de Ciencias de la UNAM, como parte de las labores docentes del programa, misma que fue rápidamente extendida a otras universidades como la Autónoma de Puebla en 1963. En 1966 fue creada la Sociedad Mexicana de Genética que impulsó la consolidación de la genética humana, la genética molecular y la genética de poblaciones. En 1972 se creó el Instituto Nacional de Energía Nuclear que en 1979 se convirtió en el Instituto Nacional de Investigaciones Nucleares (ININ), donde se desarrolló inicialmente la genética de poblaciones de *Drosophila* spp.

Uno de los proyectos más importantes del Programa de Genética y Radiobiología fue el emprendido en 1974 en colaboración con Theodosius Dobzhansky, denominado "Population genetics of Mexican *Drosophila*", que marcó el desarrollo de la genética de poblaciones dentro del programa en el ININ. Dobzhansky trataba de entender las bases ecológicas de la variación genética en las poblaciones naturales de *Drosophila pseudoobscura*, y la relación entre la cantidad de variación genética presente en las poblaciones y su tasa de evolución (Dobzhansky *et al.* 1975). Este programa fue adoptado por investigadores mexicanos como María Esther de la Rosa, Judith Guzmán Rincón, Olga Olvera y Víctor Manuel Salceda Sacanelles, quienes colaboraron con el grupo de Dobzhansky por varias décadas.

La investigación en genética de poblaciones en comunidades mestizas e indígenas usando marcadores genéticos, como isoenzimas y otras proteínas estructurales, comenzó entre los años 1960 y 1970 (Salazar-Mallén *et al.* 1952; De Garay 1963; Lisker 1981). La investigación genética médica en México estuvo ligada principalmente a los centros de salud, en donde Antonio Velázquez (Velázquez-Arellano *et al.* 1986) y Rubén Lisker (Lisker 1981) hicieron contribuciones muy importantes.

El primer centro mexicano para la investigación en genética humana fue la Unidad de Investigación de Genética Humana del IMSS, fundada en 1966, cuyas contribuciones más importantes han sido el desarrollo de nuevos métodos de bandeado cromosómico y los estudios sobre

los efectos de la malnutrición en los cromosomas (Armendares *et al.* 1971). En 1968 se creó la Asociación Mexicana de Genética Humana, y el primer programa de posgrado en genética médica se estableció en 1969 en la Unidad de Investigación en Genética Humana del Hospital Pediátrico dependiente del Centro Médico Nacional del Instituto Mexicano del Seguro Social (Salamanca y Armendares, 1995). Esta área se ha desarrollado en diferentes instituciones y uno de los resultados de ello es que, en 2003, México fue el primer país latinoamericano en crear un Instituto Nacional de Medicina Genómica (Inmegen).

En el área de biología molecular sobresalen los trabajos pioneros de Luis Herrera-Estrella, quien ha destacado por sus estudios sobre la transformación genética en plantas (Herrera-Estrella *et al.* 1983) y Francisco Bolívar, quien desarrolló métodos para producir proteínas recombinantes humanas en bacterias (Bolívar *et al.* 1977) y fue el responsable del desarrollo de nuevas herramientas moleculares para la clonación y expresión del ADN en *Escherichia coli* (Itakura *et al.* 1977). Por último, el área de la biología molecular fue desarrollada en México, principalmente en el Centro de Investigación y de Estudios Avanzados (Cinvestav) y en la Universidad Nacional Autónoma de México (UNAM). En 1973 fue fundado el primer Departamento de Biología Molecular en el Instituto de Investigaciones Biomédicas de la UNAM, y posteriormente, debido al rápido desarrollo de la genética molecular y sus aplicaciones en la biotecnología, en 1982 se fundó el Centro de Investigaciones en Genética y Biotecnología en la UNAM, mismo que fue transformado en 1991 en el Instituto de Biotecnología, en donde se desarrollan diversas disciplinas relacionadas con la ingeniería celular, la biología del desarrollo, la biología molecular de plantas, así como la microbiología y la medicina moleculares, entre otras. Por otro lado, en 1975 fue fundado el Departamento de Genética y Biología Molecular del Cinvestav, donde se desarrollan líneas de investigación sobre la regulación y la expresión genética en modelos bacterianos y sistemas eucariontes, biología molecular de la respuesta inmune, de las enfermedades hereditarias y del cáncer, neurobiología, y genética de poblaciones. Posteriormente se fundaron unidades de investigación en varias de estas líneas en algunos de los centros públicos de investigación, como el Centro de Investigaciones Científicas de Yucatán (CICY), el Instituto de Ecología, A.C., en Xalapa (IE-X) y el Centro de Investigaciones Biológicas del Noroeste (Cibnor), que junto con varios grupos creados en algunas universidades estatales (de Guanajuato, de Guadalajara y Autónoma de Baja California Sur) han

hecho posible un crecimiento vigoroso de la genética en México.

La genética de la conservación, que en otras partes del mundo había comenzado en la década de los años ochenta del siglo xx, se inició formalmente en México con la publicación por Coello *et al.* (1993) de un trabajo que mostraba que no había variación genética en la rarísima especie *Lacandonia schismatica*. Durante los siguientes años esta área del conocimiento se ha estado desarrollando notablemente en el país y los resultados de esos trabajos forman la parte fundamental de la revisión que se presenta en el capítulo 15 de este volumen.

14.2 ESTIMADOS DE LA VARIACIÓN CON CARACTERES MOLECULARES

Los enfoques que se han usado para entender la variación genética en especies mexicanas incluyen isoenzimas, polimorfismos de fragmentos de restricción (RFLP, por sus siglas en inglés), polimorfismos de ADN amplificados aleatoriamente en mayor o menor grado (RAPD y AFLP), repetidos de secuencias internas (ISSR, por sus siglas en inglés), microsatélites y secuencias de ADN.

La elección de los marcadores ha dependido de varios factores, como el tipo de pregunta y algunas de las características de marcador como tipo de herencia y, de manera secundaria, de la facilidad en su uso y costo. Sin duda la elección de los marcadores debe estar guiada por el modelo de mutación al que se ajustan y por ello a los algoritmos de inferencia de los parámetros genéticos. Por ejemplo, los datos basados en isoenzimas se ajustan a un modelo de alelos infinitos, mientras que los datos de secuencia de ADN usan un modelo de sitios infinitos.

Adicionalmente las preguntas que se han respondido con estos marcadores incluyen: ¿cuál es la magnitud, distribución y dinámica de los acervos silvestres de plantas domesticadas?, ¿cuáles han sido los centros de origen y diversificación de las especies cultivadas de plantas mexicanas?, ¿cuál es la mejor estrategia de conservación de la variación genética en especies cultivadas, especies raras o en peligro de extinción en México?, ¿cuál ha sido el tamaño histórico efectivo de la población de especies raras, claves o útiles y cómo se puede usar esa información para concebir el tamaño efectivo viable de las poblaciones?, ¿cuál es la asociación de parentesco entre animales y/o plantas de una población particular?, ¿cuál es la asociación entre los haplotipos o genotipos de especies patógenas y su grado de virulencia? El capítulo 15 de este

volumen muestra muchos de los estudios que se han llevado a cabo en México para responder estas preguntas.

Los estimados que normalmente se utilizan para evaluar la variación genética incluyen la heterocigosis esperada, el polimorfismo, π y θ (Gillespie 2004; Hedrick, 2005; Hartl y Clark 2006). Estos estimados se han desarrollado para todos los marcadores aunque algunos de ellos, como la heterocigosis observada, no pueden estimarse usando métodos tradicionales con marcadores que muestran dominancia como los RAPD, AFLP o ISSR. Por ello se han desarrollado métodos bayesianos para estimar los estadísticos F_{IS} y F_{ST} de Wright (Holsinger *et al.* 2002; Holsinger y Wallace 2004).

Los valores de π tienen la característica de ser estimados de la heterocigosis en el nivel nucleotídico (Hedrick, 2005), es decir, pueden ser interpretados como la heterocigosis esperada. Su método de estimación toma en cuenta la distancia que hay entre dos secuencias de ADN como el número de diferencias y lo corrige multiplicándola por las frecuencias de ambas secuencias. La suma de ese estimado para todas las combinaciones posibles de pares de secuencias es π . Otro estimado de la diversidad genética en el nivel nucleotídico es el número de sitios segregantes S . El número de sitios segregantes es una manera muy precisa de estimar la variación, siempre y cuando el modelo de mutación sea de sitios infinitos y como consecuencia no haya homoplasia. Estos estimados pueden extrapolarse a secuencias de aminoácidos de forma sencilla. Existen otros estimadores que están empezando a usarse y que consideran la distribución del número de sitios de nucleótidos con 1,2,3...n muestras (Wakeley 2008).

14.3 ESTIMADOS DE LA VARIACIÓN DE CARACTERES CUANTITATIVOS

Muchos de los caracteres clave en el proceso evolutivo, en la agricultura o en la virulencia de un patógeno, son caracteres cuantitativos. El estudio de esta variación está conceptual y metodológicamente separado del resto de la genética evolutiva. Existen muy pocos estudios de este tipo de caracteres tanto en especies de importancia agro-nómica como en especies silvestres en México. Estos estudios han mostrado que existen niveles significativos de heredabilidad que garantizan una respuesta a la selección tanto natural como artificial. Algunos ejemplos de estudios de genética de poblaciones recientes en México incluyen análisis sobre el comportamiento de defensa de

las abejas (Guzmán-Novoa *et al.* 2005), la identificación de genes de tolerancia al frío en maíz (Frachebaud *et al.* 2002), la determinación del número de genes que afectan el tamaño de la semilla y la concentración de minerales y taninos en frijol usando loci de rasgos cuantitativos (QTL por sus siglas en inglés) (Guzmán-Maldonado *et al.* 2003) y la variación genética en el costo y el beneficio de la herbivoría en toloache (Fornoni *et al.* 2004). La profundización en aspectos que tienen que ver con el número de genes que afectan un carácter y la asociación que la genética cuantitativa debe de tener con la genética del desarrollo y la genómica están por desarrollarse en México, y por ello se requiere apoyarlos decididamente.

14.4 ESTIMADOS DE LA ESTRUCTURA GENÉTICA

14.4.1 Introducción

Dentro de la teoría de genética de poblaciones (que estudia los patrones y causas de la diversidad genética en las poblaciones) existen conceptos que descomponen la variación genética en diferentes factores. Hay una cantidad muy grande de métodos descritos en la literatura sobre genética de poblaciones para estimar la diferenciación y la estructura genéticas. Las distintas medidas de estructura genética están relacionadas entre sí, y se basan en analizar las diferencias en las frecuencias alélicas. Las pruebas más sencillas simplemente comparan estadísticamente estas últimas. Otros métodos se basan en estimar la proporción de variación genética que se encuentra dentro y entre las poblaciones, usualmente empleando el estadístico F_{ST} o sus análogos, que permiten comparar de manera clara y cuantitativa la diferenciación existente entre poblaciones de distintas especies. Estos estimadores tienen en principio la ventaja adicional de que nos pueden dar el estimado conjunto de la importancia de la deriva génica y el flujo génico, es decir, el número efectivo de migrantes por generación, o Nm . Por último, hay diferentes posibles medidas de “distancia genética” entre pares de poblaciones. La más conocida es la distancia genética de Nei, que refleja el número de sustituciones nucleotídicas en el ADN, y que como se ha usado ampliamente en diferentes grupos de organismos, permite comparaciones a distintos niveles (entre poblaciones, entre variedades o subespecies, entre especies). Adicionalmente podemos utilizar las distancias genéticas entre pares de poblaciones para reconstruir las genealogías de poblaciones, ya sea mediante métodos de agrupamiento como

el *Unweighted Pair-Groups Method with Arithmetic Averages* (UPGMA) o *Neighbor-joining*, o usar estas distancias en análisis de aislamiento por distancia, graficando las distancias genéticas pareadas como función de la distancia geográfica que separa a las poblaciones, bajo el supuesto de que las diferencias genéticas se incrementan con la distancia que las separa, con una transformación logarítmica y usualmente analizando la confiabilidad con una prueba de Mantel. Veamos con mayor detalle algunos puntos para cada posible metodología. También veremos otros puntos relacionados con la estructura genética, como la estimación de la endogamia, el grado de clonalidad y otros métodos finos, como las autocorrelaciones. Para detectar diferencias estadísticas significativas en las frecuencias alélicas entre las poblaciones se pueden realizar diferentes pruebas estadísticas equivalentes; la más comúnmente usada es la prueba de heterogeneidad en las frecuencias alélicas de Workman y Niswander (1970) basada en una χ^2 (Hedrick 2005).

14.4.2 Estadísticos F de Wright y estimadores análogos

Wright (1951) introdujo un método para partir el coeficiente de endogamia en una población subdividida (F_{IT}) entre el componente debido a apareamientos no aleatorios dentro de poblaciones (F_{IS}) y la subdivisión entre poblaciones (F_{ST}). Así, la endogamia total tendría un componente generado por la cruce entre parientes dentro de una población (F_{IS}) y otro por el balance entre deriva génica y flujo (F_{ST}). La definición original de Wright se basaba en el coeficiente de endogamia. Así, los estadísticos F pueden ser vistos como la correlación entre genes homólogos tomados de un nivel de la subdivisión en relación con cualquier otro nivel superior: la correlación entre los genes de los individuos (I) y los de la población total (T) es representada por F_{IT} , que corresponde a la endogamia total. La correlación entre los genes de los individuos y los de la subpoblación (S) es representada por F_{IS} , mientras que la correlación entre los genes de la subpoblación y los de la población total está representada por F_{ST} (Excoffier 2001), que es igual a la probabilidad de que dos alelos idénticos por descendencia (*identical by descent*) se combinen en un cigoto, o autocigosis. Los estadísticos F se relacionan entre sí de la siguiente manera: $(1 - F_{IT}) = (1 - F_{ST})(1 - F_{IS})$ y por lo tanto

$$F_{ST} = \frac{F_{IT} - F_{IS}}{1 - F_{IS}}.$$

La estimación de la F_{ST} es más fácil de visualizar siguiendo la definición de Nei (1973), quien considera la probabilidad de heterocigosis en función de la autocigosis:

$$F_{ST} = \frac{H_T - H_S}{H_T},$$

donde H_T es el promedio de la heterocigosis esperada en la población total para todos los loci y H_S es el promedio de la heterocigosis esperada dentro de subpoblaciones para todos los loci. F_{ST} mide la reducción en la heterocigosis debida a diferenciación genética entre poblaciones. También se puede definir la F_{ST} en términos de las varianzas en las frecuencias alélicas entre las poblaciones:

$$F_{ST} = \frac{\text{var}(p)}{p(1-p)};$$

cuanto mayor haya sido la deriva génica, mayor será la varianza en las frecuencias alélicas entre poblaciones.

La F_{ST} ofrece dos atractivos: se ha calculado en gran cantidad de organismos, por lo cual se facilita su comparación; además, su interpretación es sencilla: si es de 0, indica que las frecuencias alélicas son iguales en todas las poblaciones, no ha habido diferenciación (*i.e.*, $H_t = H_s$). El máximo posible es de 1, cuando cada población está fija en alelos diferentes (*i.e.*, $H_s = 0$). La otra ventaja es que, como discutimos más adelante, a partir de la F_{ST} se puede obtener una aproximación de Nm si se ha llegado al equilibrio. Es importante tener cuidado al interpretar la F_{ST} en comparaciones entre especies, ya que los muestreos en distintos estudios pueden ser muy diferentes: algunos se realizan en una sola localidad, comparando partes de una población o manchones de individuos aislados por metros, mientras que en otros estudios se comparan poblaciones separadas por cientos o incluso miles de kilómetros. Ciertamente, aunque la diferenciación depende de los detalles de la biología e historia de cada especie, en los estudios que contemplan una menor área geográfica se puede esperar menor diferenciación medida como F_{ST} que en estudios que abarcan un área muy grande.

Existen diversas formas de calcular la F_{ST} y sus equivalentes. Wright (1951) la definió solo para un locus con dos alelos. Nei (1973) amplió la definición para poder incluir un número ilimitado de alelos por loci, en términos de la comparación de la heterocigosis como vimos antes, y la llamó G_{ST} . Posteriormente, Cockerham (1973), Weir y Cockerham (1984) y Weir (1996) desarrollaron un algo-

ritmo que hace una partición de la varianza análoga a un análisis de varianza, y esta estimación se ha llamado coeficiente de coancestría o θ (Weir y Cockerham 1984), y tiene la ventaja de estar menos sesgado que las otras dos estimaciones, de forma tal que permite manejar de manera más adecuada diferencias en el número de individuos y loci entre localidades. Asimismo se han desarrollado métodos de estimación de F_{ST} para marcadores moleculares dominantes como RAPD, AFLP e ISSR (Holsinger y Wallace 2004). Más recientemente, Excoffier *et al.* (2005) han generalizado el método anterior en el AMOVA (*Analysis of Molecular Variance*), que tiene menos requerimientos estadísticos y resulta menos sesgado que una estimación de F_{ST} normal, que se puede estimar a varios niveles y es especialmente útil para datos dominantes (como RAPD) (Excoffier 2001). Para otros modelos de mutación, como el *stepwise mutation model*, que se supone siguen los microsatélites, se han propuesto otras medidas, como la R_{ST} de Slatkin (1995), o para otros datos tipo secuencias, como la N_{ST} (Lynch y Crease 1990).

14.4.3 Flujo génico

El flujo génico o migración se refiere a todos los mecanismos que resultan en el movimiento de genes de una población a otra. Puede referirse al restablecimiento del flujo génico posterior a procesos de extinción y recolonización de poblaciones (Slatkin 1985). Durante un tiempo se sugirió que el flujo génico en general era muy restringido y de poca importancia evolutiva (Levin 1981). Sin embargo, al estimar el flujo génico en distintas especies se ha observado que los niveles pueden ser bastante altos (Rieseberg y Burke 2001) y que pueden actuar como una fuerza que mantiene la cohesión entre las poblaciones de una especie. El flujo génico es un importante componente de la estructura poblacional, ya que sus patrones y niveles determinan hasta qué grado cada población de una especie es una unidad evolutiva independiente (Slatkin 1994). La tasa de movimiento de genes de una población a otra afecta de manera importante a las especies. La cantidad de flujo génico necesaria para prevenir la evolución independiente de las poblaciones de una especie dependerá de las otras fuerzas evolutivas.

Una medida que resulta fácil de calcular y conceptualmente útil es la Nm , que es la multiplicación del tamaño efectivo (N_e) por la tasa de migración (m), y nos habla del número de "migrantes efectivos". Si Nm o el número de migrantes efectivos es mayor que 1, teóricamente el flujo génico supera los efectos de la deriva génica y previene la

diferenciación local. Si Nm es menor que 1 entonces se puede decir que la deriva actúa independientemente en cada una de las poblaciones, y si es mayor de 4 entonces las poblaciones se comportan como una gran población más o menos panmíctica. No hay que olvidar que los valores calculados de Nm dependen directamente de nuestra estimación de la F_{ST} y la estimación final de Nm reflejará los errores del sesgo que haya en la estimación de la F_{ST} (Whitlock y McCauley 1999).

14.4.4 Métodos directos para estimar el flujo génico

Se basan en observaciones o experimentos que miden el grado de dispersión de gametos o individuos para tratar de estimar de estos datos ecológicos y demográficos directamente la tasa de migración (m). Por ejemplo, la dispersión de individuos o gametos marcados y la captura y recaptura de individuos marcados. Un método alternativo que permiten los marcadores moleculares es hacer un análisis de parentesco que pueda identificar padres y después cuantificar el patrón del movimiento de genes. Los análisis de parentesco nos permiten estimar la distribución de las distancias de dispersión y examinar el movimiento de genes por polen y semilla (Eguarte *et al.* 1993) dentro de una población. Las estimaciones obtenidas con este tipo de datos no pueden ser interpretados directamente como flujo génico entre poblaciones, ya que mide la migración dentro de un área determinada, que no es necesariamente una población y no miden el flujo entre poblaciones. Tampoco pueden ser interpretados como flujo génico en una escala evolutiva, sino solo como flujo génico que ocurre en el periodo en el que se midió, lo que involucra generalmente uno o pocos eventos reproductivos. Las medidas directas de dispersión no necesariamente reflejan el movimiento de genes, porque no se sabe si el migrante se reproduce exitosamente (Whitlock y McCauley 1999). Adicionalmente, estos métodos subestiman la frecuencia de la dispersión a larga distancia, ya que puede ser difícil detectarla, y obviamente no estiman la importancia de extinciones y recolonizaciones como una fuente de flujo génico (Slatkin 1985). Por último, estas estimaciones están limitadas en el tiempo y no reflejan eventos raros que pueden ser muy importantes. Aun así, estas estimaciones pueden ayudar a tener una idea global de la situación actual de la migración para compararla con otras estimaciones genéticas que hacen promedios históricos, como F_{ST} .

14.4.5 Métodos indirectos para estimar el flujo génico

Los métodos indirectos se basan en el análisis de la distribución espacial de alelos en las poblaciones y de esta manera se hacen inferencias de los niveles o patrones de flujo génico en las poblaciones (Slatkin 1985). La mayoría de los modelos teóricos de flujo génico surgen de los conceptos desarrollados por Sewall Wright, basados en poblaciones continuas, y utilizan un enfoque de aislamiento por distancia o en poblaciones que funcionan como islas que se diferencian por mutación y deriva génica (Wright 1943). El modelo usado comúnmente para estimar flujo génico es el modelo de islas infinitas (*infinite islands model*) de Wright (1951). Este modelo considera condiciones en equilibrio entre un número infinito de islas o subpoblaciones de igual tamaño, que intercambian migrantes entre cualquiera de las islas con la misma probabilidad a una tasa constante. Las poblaciones pueden ser tratadas como réplicas y el modelo se puede caracterizar con solo dos parámetros: tamaño poblacional (N) y tasa de migración (m). La importancia de la deriva génica es proporcional a $1/N$, mientras que la importancia del flujo génico es proporcional a m (Slatkin, 1985). Por su parte, en el modelo de *stepping-stone*, introducido por Kimura (1953), las poblaciones se localizan en una especie de enrejado de una, dos o tres dimensiones y los individuos solo pueden moverse entre poblaciones adyacentes, pero los resultados del equilibrio, aunque análogos a los del modelo de Wright, son más complicados y menos generales.

Para estimar el flujo génico con un método indirecto usualmente se emplea la fórmula de Wright (1951), cuando la F_{ST} llega al equilibrio entre deriva génica y migración en el modelo de islas infinitas:

$$F_{ST} \approx \frac{1}{4Nm + 1};$$

despejando,

$$Nm \approx \frac{1 - F_{ST}}{4}.$$

De acuerdo con este modelo, Wright encontró que una tasa de migración >1 en cada generación es suficiente para contrarrestar la diferenciación genética debida a deriva génica. Crow y Aoki (1984) encontraron que, considerando la mutación y un número finito de islas (modelo

de n islas), donde el equivalente para alelos múltiples de F_{ST} es G_{ST} de Nei (1973), se puede concluir que

$$G_{ST} \approx \frac{1}{4Nm\alpha + 1},$$

donde

$$\alpha = \left(\frac{n}{n-1} \right)^2,$$

y n es el número de subpoblaciones. Slatkin y Barton (1989) señalan que esta estimación de Nm a partir de la F_{ST} es uno de los métodos más sólidos. Esta es la fórmula que generalmente se utiliza para estimar indirectamente Nm . Sin embargo, la utilidad de estos métodos ha sido severamente cuestionada en artículos como los de Whitlock y McCauley (1999) y Neigel (2002), ya que la inferencia supone un número infinito de poblaciones, que el flujo génico no es afectado por la distancia geográfica y que cada población está en un equilibrio entre migración y deriva. Adicionalmente, Pearse y Crandall (2004) han propuesto una ruta analítica cuando se trata de hacer análisis para conservación que va más allá de F_{ST} .

Otro método para estimar Nm es el de “alelos privados” (Slatkin 1981). Sin embargo, diversas simulaciones han demostrado que el método de alelos privados es más sensible a errores en la recolección de datos (Slatkin y Barton 1989).

14.4.6 Distancias genéticas

Los índices de distancia genética se calculan entre pares de poblaciones y pretende describir el grado de diferencia genética entre estas dos poblaciones. Así, se puede calcular directamente la F_{ST} entre dos poblaciones y usarse como distancia genética, o utilizar la estimación de Nm (M en la nomenclatura de Slatkin (1993, 1994)). Existen adicionalmente una infinidad de índices de distancia genética, muchos de ellos diseñados originalmente para análisis cuantitativo de datos multivariados. Nei (1987, págs. 208 a 253) hace una revisión de esta extensa literatura.

La medida de distancia genética más usada es la distancia de Nei (1973, 1977). Este índice pretende estimar el número de mutaciones que a nivel nucleotídico se han acumulado en las secuencias de dos linajes a partir del tiempo que ha transcurrido desde su divergencia original. La idea (Nei 1987) es que los diferentes alelos (o la presencia o ausencia de los loci en el caso de marcadores basados en PCR, como RAPD, AFLP u OSSR) es originada

gracias a que un codón (o una base nucleotídica en el caso de marcadores relacionados con el PCR) es distinto. Por lo tanto, a partir de datos de frecuencias alélicas debería ser posible calcular estadísticamente el número promedio de las diferencias en los codones o en las bases nucleotídicas por locus. Dado que este número es una medida directa de las diferencias genéticas entre dos poblaciones, se le considera como una medida de distancia genética (Nei 1973). Así, el promedio en el número neto de sustituciones nucleotídicas, D , estaría dado por $D = -\log_e I$, donde I (la identidad genética) es igual a

$$\frac{J_{xy}}{(J_x J_y)^{1/2}},$$

y $J_x = \sum x_i^2$ (la sumatoria de las frecuencias alélicas al cuadrado en la población 1), $J_y = \sum y_i^2$ (la sumatoria de las frecuencia alélicas al cuadrado de la población 2), y J_{xy} es igual a la sumatoria del producto, para cada alelo, de las frecuencia alélicas en la población 1 multiplicada por la frecuencia alélica de ese mismo alelo en la otra población. Si las frecuencias alélicas son las mismas, $J_x = J_y$, y la I equivale a 1 (esto es, la identidad máxima posible), mientras que si no comparten ninguno de los alelos, las J_{xy} van a ser 0, y en consecuencia la I equivale a 0 (que implica la identidad mínima posible). La D por lo tanto toma valores de 0, si son idénticas las poblaciones, a infinito, si son completamente diferentes. Si no ha pasado mucho tiempo, digamos que menos de un millón de años (Nei, 1987), $D = \alpha t$, donde α es igual a la tasa de mutación por locus por años. Nei sugiere que si la α del marcador genético es 10^{-7} , el tiempo de separación entre dos poblaciones sería $t = 5 \times 10^6 D$ (Nei 1987).

14.4.7 Aislamiento por distancia

Con el fin de saber si una especie sigue un modelo de aislamiento por distancia, es decir, si existe una correlación entre las distancias geográficas y las genéticas entre los distintos pares de poblaciones (Wright 1943), se puede hacer una comparación formal con, por ejemplo, una prueba de Mantel (1967). La distancia genética puede ser directamente la distancia genética de Nei, o con las estimaciones de Nm pareadas (M) en logaritmo que recomienda Slatkin (1993, 1994) o seguir la sugerencia de Rousset (2001) y graficar

$$\frac{F_{ST}}{1 - F_{ST}}.$$

Independientemente de la distancia genética empleada y su estandarización, la idea es la misma: si hay una correlación significativa entre las distancias geográfica y genética, existe evidencia de aislamiento por distancia (Slatkin 1994; Rousset 2001).

14.4.8 Inferencia de estructura y proporción ancestral

Recientemente se han implementado metodologías que están basadas en modelos de evolución específicos para inferir la probabilidad de que un individuo pertenezca o no a algún grupo particular (Pritchard *et al.* 2000). Estos grupos permiten inferir la estructura genética sin tomar en cuenta información previa sobre el número y distribución de las localidades analizadas originalmente. El método se propuso para genes no ligados pero recientemente se han propuesto métodos y algoritmos para genes ligados (Falush *et al.* 2003; Piry *et al.* 2004; Corander y Tang 2007). Estos algoritmos tienen que ser validados usando simulaciones y aplicados en diversas situaciones naturales para evaluar de forma adecuada su capacidad para inferir procesos de migración.

14.5 LA TEORÍA DE COALESCENCIA

Esta teoría es un modelo retrospectivo basado en la genealogía de los alelos. Usa un modelo básico (normalmente el modelo Wright-Fisher en el que solo hay estocasticidad en la reproducción) para describir la dinámica de la unión de los linajes hasta un ancestro común. Así se establecen diferentes hipótesis (incluyendo otros procesos como mutación, migración o recombinación) acerca de cómo sería la distribución de la variación genética usando diferentes escenarios de coalescencia y la estimación de otros parámetros genéticos, como el tamaño efectivo poblacional y el tiempo al ancestro común más reciente (TMRCA, por sus siglas en inglés). Esta teoría surge como una extensión natural de la genética de poblaciones clásica, tomando también algunas características e ideas seminales de la filogenia y de la teoría neutral de la evolución molecular (Hein *et al.* 2005). La genética de poblaciones describe las fuerzas responsables de la variación genética, como la mutación, la deriva genética y la selección natural. Por su parte, la filogenia converge con la teoría de coalescencia en el uso y la descripción de árboles que relacionan unidades evolutivas, aunque a veces esta semejanza puede ser solo metafórica. Finalmen-

te, la teoría neutral provee, por un lado, un marco teórico para estudiar la evolución en el nivel molecular y, por otro, un modelo nulo con el cual contrastar escenarios más complejos de evolución.

A pesar de esta relación con otras disciplinas, la teoría de coalescencia se distingue de ellas perfectamente. Al contrario de la genética de poblaciones, que generalmente es prospectiva y se aplica a poblaciones enteras, la coalescencia es retrospectiva y utiliza datos de muestras actuales como base. Asimismo, en contraposición a la filogenia, en la teoría de coalescencia la topología de un solo árbol no es tomada en cuenta en la mayoría de los casos y se pondera al azar en la generación de la muestra actual.

14.5.1 Aplicaciones de la teoría de coalescencia

La teoría de coalescencia es usada para detectar o estimar parámetros de los procesos evolutivos que dieron lugar a la variación genética observada en una muestra de organismos actuales. Los estudios que utilizan la coalescencia generalmente comprenden trabajos de genética de poblaciones y sus posibles aplicaciones, como genética de la conservación. Asimismo, tiene gran repercusión en los estudios de filogeografía y han mostrado ser muy útiles en la solución de problemas de sistemática para complejos de especies muy cercanamente relacionadas. Sin embargo, en donde ha habido más desarrollo y aplicación de la teoría es en lo relacionado con la variación genética humana; tanto en lo que se refiere al origen y evolución del hombre como en lo que toca al descubrimiento de variantes relacionadas con enfermedades de origen genético. También de importancia para el género humano ha sido el uso de la coalescencia en el estudio de enfermedades causadas por patógenos, como el sida y la malaria.

La implementación de la teoría de coalescencia tiene varias vertientes. Por un lado, están los llamados estadísticos de resumen, que sintetizan la información contenida en la muestra comparándola con lo que se esperaría obtener bajo un modelo nulo o neutro, como las poblaciones ideales de Wright-Fisher (Fisher 1932; Wright 1931, 1951). De esta manera se pueden detectar procesos que afectaron a la muestra en el pasado; por ejemplo, cambios en el tamaño de la población, estructuración poblacional, selección natural y recombinación. Esta es la aplicación más asequible y, por lo tanto, la más utilizada en la literatura. Algunos de los estadísticos de resumen más comunes son la D de Tajima (1989) y la distribución de las diferencias pareadas (Rogers y Harpending 1992).

Una revisión de estos estadísticos fue publicada por Ramos-Onsins y Rozas (2002) que muestran que el enfoque original de Rogers y Harpending (1992) es conservador.

Por otra parte, la coalescencia puede ser usada para estimar parámetros relacionados con los procesos antes descritos, en lo que podría llamarse inferencia por coalescencia completa. La coalescencia completa permite no solo detectar los procesos evolutivos que afectaron a la muestra sino que, además, infiere parámetros relacionados con estos procesos que pudieron dar lugar a la misma. Por ejemplo, la tasa de mutación poblacional, el número de migrantes por generación entre n poblaciones, la tasa de crecimiento poblacional, el tiempo desde el ancestro común más reciente de la muestra (TMRCA), el tiempo de divergencia entre dos poblaciones o la tasa de recombinación de la población.

La ventaja de la coalescencia completa sobre los estadísticos de resumen es que utiliza toda la información contenida en la muestra. Lo anterior se consigue ponderando una muestra grande de todas las genealogías que pudieron dar lugar a la muestra. Sin embargo, una desventaja es que aún son pocos los modelos de poblaciones para los que existe inferencia de coalescencia completa y por lo general son computacionalmente intensos.

Una manera más de utilizar la coalescencia es mediante el uso de simulaciones. Es relativamente sencillo generar grupos de datos o genealogías utilizando modelos de coalescencia (Hudson 2001). Existen muchas formas de utilizar las simulaciones hechas de esta manera. Por ejemplo, usar las simulaciones como hipótesis nula al contrastar los resultados con los datos reales; utilizar desviaciones del coalescente básico y observar si se recuperan los datos reales; usar las simulaciones de forma heurística para obtener parámetros de los datos simulados que converjan con los datos reales; generar datos a partir de parámetros extraídos de la muestra y calcular los límites de confianza de tales parámetros (Rozas *et al.* 2003).

14.5.2 Programas más comúnmente usados para hacer inferencias usando la teoría de coalescencia

Existen varios programas que usan explícitamente la coalescencia para hacer inferencias poblacionales. Aquí se mencionan los más frecuentemente usados. Dentro del paquete LAMARC existe la posibilidad de estudiar los patrones demográficos históricos usando el programa Fluctuate (Kuhner *et al.* 1995, 1998) o los patrones de migración históricos y recientes utilizando el programa Migrate (Beerli y Felsenstein 2001). Otros paquetes más generales

que hacen inferencias usando coalescencia pero que también estiman parámetros más tradicionales de la genética de poblaciones son ADNSp (Rozas *et al.* 2003), Arlequin (Excoffier *et al.* 2005) y SITES (Hey y Wakeley 1997). Los programas LDhat (McVean *et al.* 2002) y RDP (*Recombination Detection Program*) estiman parámetros de recombinación usando secuencias de ADN. El programa Mesquite (Maddison y Maddison 2004) usa un enfoque filogenético pero estima varios aspectos poblacionales usando simulaciones. Estos son solo algunos de los más utilizados; una amplia revisión de programas puede consultarse en Excoffier y Heckel (2006).

14.5.3 Perspectivas de la teoría de coalescencia en México

Definitivamente la teoría de coalescencia necesita conocerse más entre los científicos mexicanos. En los campos de genética de poblaciones, sistemática y filogeografía existen muchas publicaciones internacionales de mexicanos que no hacen uso de la teoría de coalescencia, pudiendo obtener más información de sus datos con ella. Muy probablemente esto se deba, por un lado, a que hace muy poco tiempo que empezaron a generarse datos con secuencias de ADN, el marcador ideal para coalescencia; y, por otro, a que apenas existen los primeros libros de texto que profundizan en las bases y aplicaciones de esta teoría.

Finalmente, no cabe duda de que la teoría de coalescencia ocupará un lugar importante en todos los estudios de genética evolutiva en México, conforme más datos estén disponibles y existan más científicos adiestrados en el tema. Sobre esto, habría que hacer énfasis en la relevancia que tiene la teoría de coalescencia en áreas clave para el futuro del país, como la medicina con bases genéticas y la conservación y aprovechamiento de recursos bióticos. Estas aplicaciones son particularmente importantes para definir criterios de restauración, para encontrar genes asociados a enfermedades o para producir programas de mejoramiento genético que incorporen genes de resistencia a condiciones agronómicas adversas como la sequía.

14.6 FILOGEOGRAFÍA

14.6.1 Origen y desarrollo

El término “filogeografía” fue acuñado a finales de los años ochenta por John Avise y colaboradores para refe-

rirse al “estudio de los principios y procesos que determinan la distribución geográfica de los linajes genealógicos, especialmente dentro de y entre taxones cercanamente emparentados” (Avice *et al.* 1987). Desde entonces, la filogeografía se ha transformado de forma extraordinaria en una disciplina integradora en la que confluyen la biología molecular y otros campos macroevolutivos, como la filogenética, la paleontología y la biogeografía, y microevolutivos, como la genética de poblaciones, la autoecología, la demografía y la etología (Avice *et al.* 2000). Los análisis filogeográficos consisten en “sobreponer” al espacio geográfico filogenias de linajes moleculares (o genealogías génicas) para buscar patrones que permitan inferir los procesos históricos y demográficos que les dieron lugar. Esta superposición puede ser literalmente gráfica o bien puede representarse mediante la codificación de linajes en un árbol filogenético de acuerdo con su origen geográfico.

Desde su inicio, el marcador molecular más utilizado en filogeografía ha sido el ADN mitocondrial (ADNmt) animal (Avice *et al.* 1987), debido a su alta tasa de evolución molecular, su estado clonal en el organismo, su herencia matrilineal asexual y a la ausencia de recombinación intermolecular (Avice 1991). Estas características evolutivas dan pie a que las genealogías mitocondriales resulten en árboles bifurcantes que representan la transmisión vertical del ADNmt de madres a hijas. En comparación con el crecimiento exponencial de estudios filogeográficos en animales, los estudios en plantas han tenido un rezago por la falta de marcadores moleculares apropiados, dado que los genomas citoplásmicos vegetales no comparten las características evolutivas del ADNmt animal (Schaal *et al.* 1998). No obstante, se han realizado esfuerzos notables (Nason *et al.* 2002; Grivet y Petit 2003; Schaal *et al.* 2003).

14.6.2 Principios y teoría

La mayoría de los principios de la filogeografía no son propios sino que provienen de las disciplinas que en ella confluyen. Por ejemplo, de la filogenética y de la cladística adopta los métodos de reconstrucción y el lenguaje para describir la topología de los árboles y las relaciones entre organismos; mientras que de la biogeografía toma escenarios hipotéticos como la dispersión y la vicariancia, sobre los que se construyen modelos nulos. Una de las piedras angulares de la filogeografía ha sido el desarrollo de una teoría matemática y estadística de los procesos de bifurcación (o coalescencia, en sentido cronoló-

gico opuesto) y del arreglo de linajes (*lineage sorting*) que suceden en las genealogías génicas de marcadores neutrales dentro de y entre especies cercanamente emparentadas, teoría conocida como “el coalescente” o “coalescencia de linajes” (Felsenstein 1971; Griffiths 1980; Hudson 1990; Nielsen y Wakeley 2001).

Otros principios, como las cinco categorías filogeográficas (Avice 2000), se han fraguado empíricamente y se muestran en la figura 14.1. La categoría de tipo I consiste en árboles en los que hay bifurcaciones profundas que resultan de grandes distancias de mutación, surgidas de la existencia de barreras al flujo génico (alopatría). Esta categoría frecuentemente se asocia a la existencia de barreras extrínsecas de largo plazo. La categoría de tipo II se caracteriza por un árbol génico con bifurcaciones profundas en algunas de sus ramas cuyos linajes principales son simpátricos a lo largo de una amplia área geográfica (Fig. 14.1). Hay diferentes escenarios hipotéticos que podrían dar lugar a este patrón filogeográfico; sin embargo, en la mayoría de los casos reales documentados, la evidencia apunta hacia una zona de contacto secundario y mezcla de linajes que divergieron históricamente en alopatría. La categoría de tipo III se manifiesta en un árbol con divergencias someras de linajes que históricamente experimentaron procesos de divergencia como resultado de alopatría (Fig. 14.1). Esta categoría permite evidenciar poblaciones que han mantenido un contacto reciente entre grupos inicialmente alopátricos que sufrieron los efectos de la divergencia genética por el efecto de la deriva génica al azar y el arreglo de linajes, o bien por selección natural. Esto implica que las poblaciones deben de haber mantenido un bajo flujo genético relativo a su tamaño efectivo poblacional para permitir la divergencia entre poblaciones. Por su parte, la categoría de tipo IV también se caracteriza por una genealogía somera pero con linajes simpátricos (Fig. 14.1). Este patrón corresponde a especies con altos niveles de flujo genético y tamaños efectivos poblacionales moderados o pequeños, cuyas poblaciones no han sido divididas por barreras filogeográficas de largo plazo. La categoría de tipo V, intermedia entre la de tipo III y la de tipo IV, se caracteriza por genealogías someras en las que existen linajes comunes ampliamente distribuidos junto con linajes cercanamente relacionados que son exclusivos (“privados”) de localidades cercanas (Fig. 14.1). Este patrón se asocia a niveles modestos de flujo genético contemporáneo entre poblaciones que han estado históricamente conectadas. En este caso, los haplotipos comunes suelen ser los ancestrales mientras que los privados se suponen derivados y apomórficos.

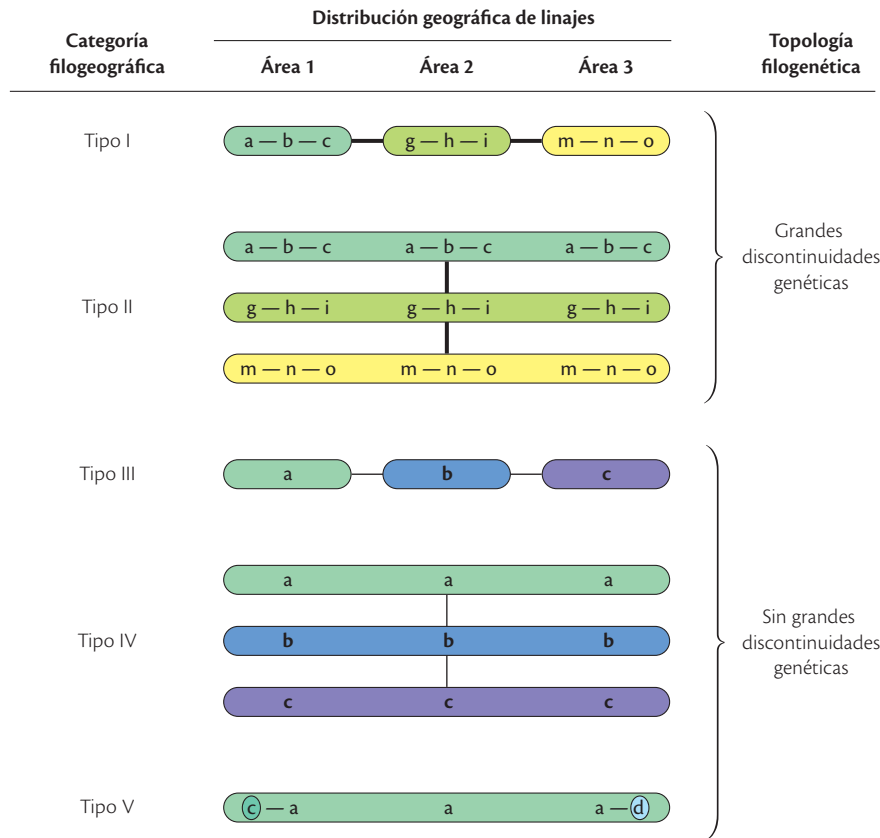


Figura 14.1 Categorías filogeográficas basadas en la distribución geográfica de linajes filogenéticos (representados por óvalos) y sus niveles de divergencia genética. Las grandes discontinuidades se encuentran representadas por líneas gruesas. El nivel de similitud genética entre linajes está codificado por la similitud en colores (modificado de Avise 2000).

14.6.3 Concordancia genealógica

Es otro principio filogeográfico fundamental. Este aspecto de la filogeografía es una extensión de los principios revisados anteriormente, aplicables a una especie y a un gen, a varios genes y a múltiples especies simpátricas en una región, haciendo uso de datos biogeográficos más tradicionales. Por lo anterior, este aspecto se conoce también como filogeografía comparativa (Bermingham y Moritz 1998). El poder de la concordancia genealógica se manifiesta en varias escalas cronológicas y espaciales, desde proveer un contexto temporal, geográfico y genético al proceso de especiación (Rocha-Olivares *et al.* 2001) hasta permitir que se realicen inferencias sobre la historia de una fauna regional (Avise 1992; Bernardi *et al.* 2003). La concordancia genealógica se da en cuatro aspectos. El aspecto I se refiere a la concordancia entre los caracteres (sitios variables) del gen en favor de la estimación de la genealogía relevante (*e. gr.*, una bifurcación profunda) (Fig. 14.2).

Es una medida de la señal filogenética de los datos moleculares y de la solidez de la reconstrucción filogenética. Generalmente se obtiene de análisis estadísticos de remuestreo como el *bootstrap* no paramétrico o el *jackknife* acoplados a los análisis filogenéticos (Felsenstein, 1985). El aspecto II se refiere a la coherencia genealógica entre genes independientes (no ligados) de un mismo taxón (Fig. 14.2). Este aspecto se diagnostica comparando directamente las topologías de ambos genes o se pueden aplicar métodos estadísticos para cuantificar la coincidencia genealógica entre matrices de caracteres (*e. gr.*, Lyons-Weiler y Milinkovitch 1997). Generalmente se busca la concordancia entre genealogías mitocondriales y nucleares, aunque debido al tamaño efectivo poblacional cuatro veces menor del μ_{ADNmt} , el proceso de coalescencia de genes nucleares se realiza más lentamente. En consecuencia, las concordancias genealógicas entre genes nucleares y mitocondriales son evidencia sustancial para realizar inferencias sobre los procesos que han determinado la

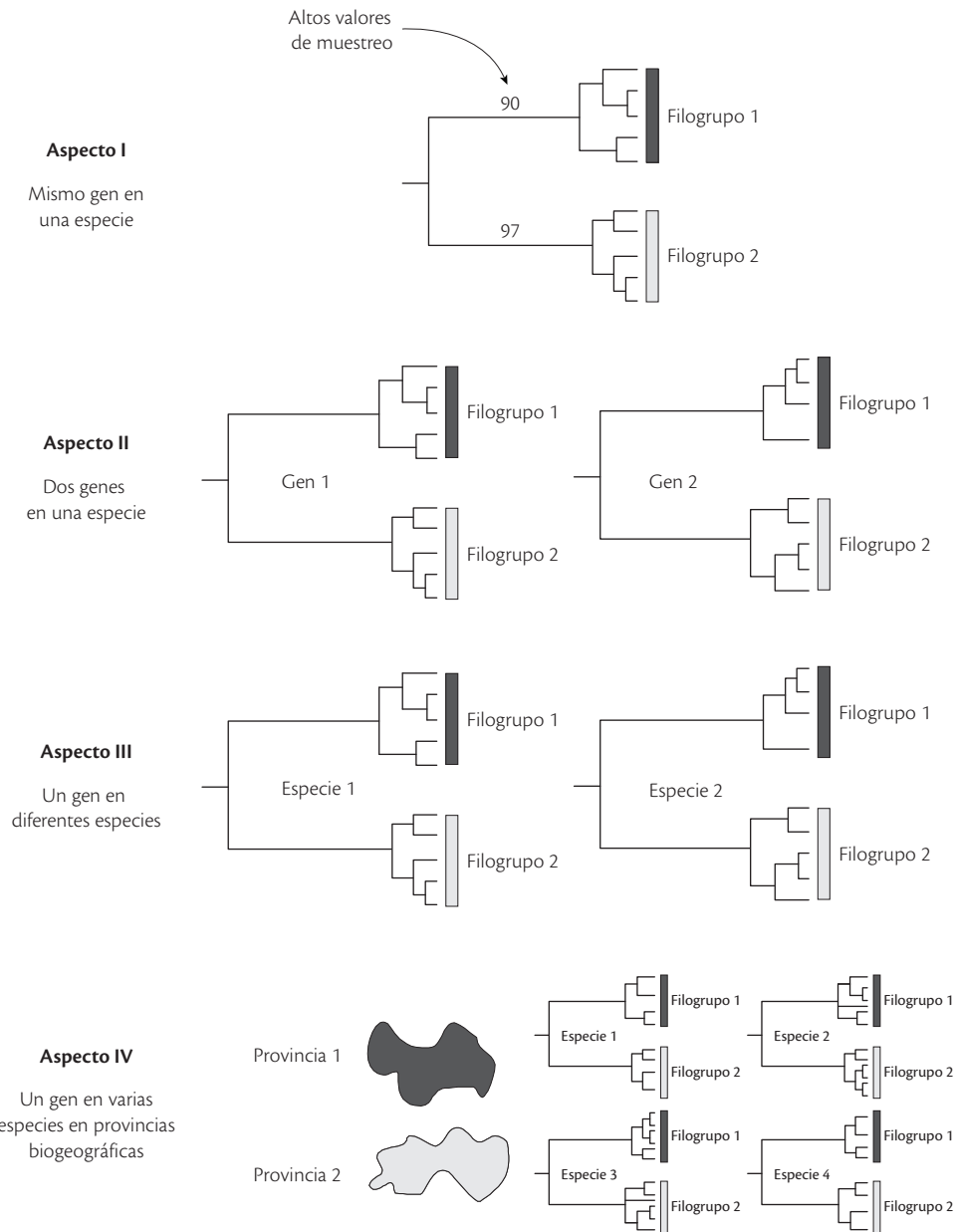


Figura 14.2 Aspectos de concordancia filogeográfica entre genealogías génicas (modificado de Avise 2000).

historia evolutiva del organismo en cuestión. El aspecto III se refiere a la concordancia entre filogenias de un mismo gen de especies distribuidas simpátricamente en una región. En este aspecto los linajes divergentes de cada árbol específico se distribuyen cada uno en áreas geográficas distintas que coinciden en cada especie (Fig. 14.2). La existencia de esta concordancia genealógica generalmente se asocia a especies con características ecológicas similares y refleja la influencia de los mismos factores ecológicos y evolutivos que han moldeado la arquitectura

filogeográfica de los organismos de la región. El aspecto IV es el más incluyente de todos y se refiere a la concordancia entre las filogenias de una biota regional con otros datos. Es el aspecto más trascendente del análisis filogeográfico, en el que se integra información acerca de la biogeografía regional, como la presencia de provincias y subprovincias biogeográficas (Fig. 14.2). La existencia de ramas filogeográficas muy divergentes implican un aislamiento de largo plazo debido a barreras históricas a la dispersión de origen geográfico o geológico, como es el

caso de los eventos históricos de vicariancia (Riddle *et al.* 2000a). La disponibilidad de un número creciente de estudios filogeográficos moleculares ha permitido detectar que la existencia de filogrupos intraespecíficos, separados por divergencias profundas, frecuentemente se encuentran distribuidos en distintas provincias o subprovincias biogeográficas, de tal suerte que los factores que han determinado que haya distintos grupos faunísticos en cada provincia también trascienden en la estructura filogeográfica de las especies que ocupan ambos espacios (Avice 1992).

Además de los principios empíricos anteriores, se han desarrollado métodos estadísticos de inferencia filogeográfica, agrupados en la filogeografía estadística (Knowles y Maddison 2002), que incorporan la naturaleza estocástica de los procesos genéticos y permiten estimar la influencia de eventos evolutivos ante las expectativas teóricas de modelos explícitos (*e. gr.* Neigel *et al.* 1991; Templeton 1998; Knowles 2004). Aquí presentaremos brevemente el análisis de clados anidados (ACA) desarrollado por Templeton y concebido originalmente para analizar el efecto fenotípico de las mutaciones (Templeton *et al.* 1987, 1988, 1992; Templeton y Sing 1993; Templeton 1995). El ACA aplicado en filogeografía consiste en la

superposición de distancias geográficas sobre un árbol genético aplicando un método estadístico riguroso diseñado para medir el grado de asociación entre la geografía y filogenia, así como los procesos evolutivos responsables de dicha asociación que permiten diferenciar entre patrones filogeográficos debidos a flujo genético recurrente pero restringido respecto de factores históricos (por ejemplo, colonización de nuevos espacios o ampliación de su rango de distribución) que han delineado la genealogía de los haplotipos. El método consiste primero en reconstruir una genealogía molecular no enraizada en forma de red mediante un algoritmo estadístico (Templeton *et al.* 1992; Templeton y Sing 1993). Posteriormente, se aplica un algoritmo de anidamiento (Templeton *et al.* 1987) para agrupar a los haplotipos individuales en clados anidados. Primero se determinan los clados que agrupan a los haplotipos, que corresponden a clados de primer nivel y representan las ramas terminales de la genealogía. Estos clados son entonces agrupados en otros más incluyentes de segundo nivel y así sucesivamente hasta que toda la genealogía se agrupa en un clado total. Conforme aumenta el nivel de anidamiento se reconstruyen clados más ancestrales (Fig. 14.3). Una vez determinados los clados, se utiliza la información geográfica

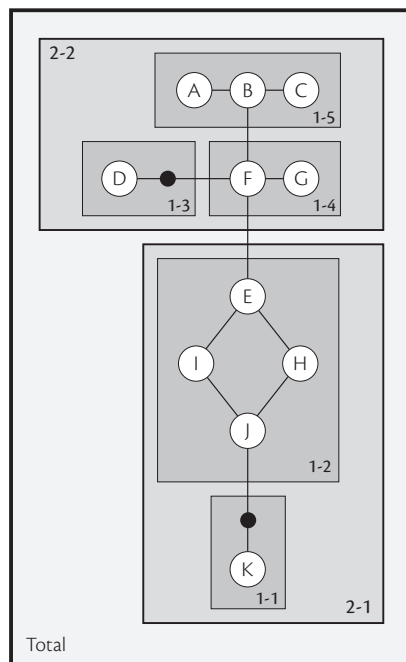


Figura 14.3 Diagrama hipotético de un cladograma mitocondrial de máxima parsimonia entre 11 haplotipos mitocondriales (A-K) indicando dos niveles de anidamiento. Los círculos representan haplotipos inferidos no observados. Los clados anidados se indican por su nivel y número de clado (*e. g.*, 2-1); la profundidad de los clados está representada por el grosor del marco que los rodea.

sobre su distribución y las distancias geográficas entre las áreas de estudio se usan para realizar el cálculo de dos estadísticos de distancia: la *distancia de clado* y la *distancia de clado-anidado*. La distancia de clado (D_c) está definida como el promedio de las distancias de los miembros de un clado al centro geográfico del clado; de manera análoga la distancia de clado-anidado (D) es la distancia promedio de las distancias de los miembros de un clado-anidado al centro geográfico del clado-anidado. Estas distancias cuantifican el grado de dispersión de los linajes y no necesitan ser rectilíneas sino que pueden ajustarse a los mecanismos y rutas de dispersión de los organismos. Los centros geográficos de los clados son análogos a centros de masa en el espacio cartesiano. Para probar la hipótesis nula de ausencia de asociación entre geografía y filogenia se realizan pruebas de permutación que permiten determinar cuáles de las D_c o D_n son estadísticamente menores o mayores que las esperadas bajo la hipótesis nula. El ACA constituye un marco de referencia estadístico objetivo para identificar la existencia de patrones filogeográficos (cuadro 14.1). Además, provee una clave para inferir las causas biológicas de las asociaciones encontradas entre la filogenia y la geografía mediante la interpretación de los estadísticos de distancia D_c , D_n y otros derivados de ellos (Templeton *et al.* 1995). Este método de inferencia ha recibido críticas recientes en la literatura y aunque ha sido usado en cientos de artículos publicados se le ha tachado de subjetivo por Panchal y Beaumont (2007), quienes usando simulaciones concluyen que el NCPA sobreestima las inferencias, es decir, infiere eventos sobre todo de flujo de genes restringido con aislamiento por distancia y expansión contiguo del rango que son inferencias frecuentes en las simulaciones. Por ello Petit (2008) ha sugerido que mientras no se sepa con más detalle qué tipo de hipótesis se podrían tratar de rechazar, el método no debiera de usarse. Estos comentarios y otros anteriores (Knowles y Madison 2002) han generado tanto modificaciones al programa (Templeton 2001, 2004) como respuestas (Templeton 2008). Templeton defiende el método basado en la idea de que precisamente hace pruebas de hipótesis específicas, entre otros aspectos. Knowles (2004) ya había mencionado que desde el punto de vista estadístico y biológico la filogeografía estadística sufre limitaciones que deben de considerarse. Sin duda, la cualidad más importante del NCPA es su capacidad para generar una serie de inferencias que deben verificarse usando otros enfoques y metodologías pero en esa cualidad está también su defecto, es decir, la generación de hipótesis, algunas de las cuales probablemente no tienen justificación estadística.

Cuadro 14.1 Algunas expectativas del análisis de clados anidados bajo diferentes patrones de estructura poblacional y eventos demográficos históricos (modificado de Templeton *et al.* 1995)

Patrón 1. Flujo genético restringido
a] Valores de D_c significativamente pequeños, en especial para los clados externos. Algunos clados internos con D_c significativamente grandes.
b] Los D_c promedio deben aumentar (y ocasionalmente nivelarse) conforme se incrementa el nivel de anidamiento. Si las distancias se nivelan, entonces la hipótesis nula de ausencia de asociación geográfica no debe ser rechazada, aunque lo haya sido en clados de niveles inferiores.
Patrón 2. Expansión de rango
a] Valores de D_c y D_n significativamente grandes en clados externos, y a veces significativamente pequeños para clados interiores, en el caso de expansión de rango contigua, aunque algunos clados externos deben mostrar D_c significativamente pequeños en el caso de colonización a gran distancia.
Patrón 3. Fragmentación alopatrica
a] D_c significativamente pequeños, sobre todo en los clados de niveles mayores. Los D_n en estos niveles pueden aumentar rápidamente mientras que los D_c permanecen restringidos, dependiendo de la configuración geográfica de las poblaciones aisladas.

14.6.4 Perspectivas

En un lapso de tan solo 20 años, el campo de la genética poblacional y el estudio de la especiación han sufrido un cambio de paradigma con la adopción de análisis de genealogías intraespecíficas. La contribución más importante de la filogeografía ha sido destacar los aspectos históricos y de desequilibrio entre fuerzas evolutivas a escala microevolutiva, así como ayudar a estudiar formalmente las relaciones entre la demografía y la genealogía (Avice 2000). El futuro de la filogeografía es sin duda promisorio. Con la disponibilidad de un mayor número de estudios de especies individuales se hace cada vez más factible el desarrollo de la filogeografía comparativa. De hecho existen regiones en México en donde se han hecho importantes análisis comparativos, como en el noroeste del país (Riddle *et al.* 2000b; Bernardi *et al.* 2003). En el capítulo 15 de este volumen se presentan más estudios hechos en México. Los avances teóricos y conceptuales en los modelos de coalescencia han contribuido al estudio de genes nucleares y su utilización en análisis filogeográficos (Hare 2001). Se puede prever un mayor uso de los conceptos genealógicos propios de la

filogeografía en la conservación y manejo de recursos, para lo que se están adoptando acercamientos filogenéticos para la identificación de unidades de manejo y protección (Moritz y Faith 1998; Newton *et al.* 1999). Sin duda, el refinamiento de los métodos de la filogeografía estadística, a la par de los avances en los modelos utilizados, ayudarán a hacer los métodos filogeográficos más poderosos, de tal manera que será posible su integración con otros modelos biológicos, como la modelación metapoblacional (Wakeley 2004). Asimismo es evidente la necesidad de que en México se consolide la teoría de coalescencia como una de las herramientas más importantes que soportarán los estudios filogeográficos en México para ampliar el conocimiento sobre las regularidades filogeográficas y fortalecer las hipótesis alternativas de la genética de poblaciones.

14.7 INFRAESTRUCTURA Y GRUPOS DE INVESTIGACIÓN

Durante las últimas décadas se han desarrollado diferentes grupos de investigación que trabajan en aspectos de diversidad genética en especies mexicanas. Asimismo hay grupos de investigación en otros países que han trabajado en aspectos de variación genética, particularmente por medio de colaboraciones. Estos grupos incluyen especies de importancia médica, agronómica y forestal y especies raras o en peligro de extinción o especies clave o muy abundantes en diversos ecosistemas. Estos grupos de investigación tanto en ecosistemas terrestres como acuáticos están dispersos en algunas instituciones del sistema de educación superior e investigación científica que incluyen al Cinvestav (Irapuato), la Escuela Nacional de Ciencias Biológicas del IPN, el Instituto Nacional de Investigaciones Nucleares, el Centro de Investigaciones Biológicas del Noroeste (Cibnor), la Universidad Autónoma de Baja California, el Instituto de Ecología, A.C., la Universidad Michoacana de San Nicolás de Hidalgo, el Centro de Investigaciones Científicas de Yucatán (CICY) y en la Universidad Nacional Autónoma de México, la Facultad de Estudios Superiores-Iztacala, la Facultad de Ciencias, la Facultad de Medicina, el Centro de Investigaciones en Ecosistemas, el Centro de Ciencias Genómicas, el Instituto de Ciencias del Mar y Limnología, el Instituto de Biotecnología, el Instituto de Investigaciones Biomédicas y el Instituto de Ecología. Existen entre 30 y 50 investigadores con sus respectivos grupos de investigación que están explorando en México la variabilidad ge-

nética de especies mexicanas. Los productos de la investigación de estos grupos están descritos en el capítulo 15 de este volumen. Aun así es importante que los diferentes grupos involucrados se asocien para generar información molecular a un paso más acelerado y con mayor eficiencia, sin olvidar que los sistemas que más necesitan esta información son las especies que por su importancia ecológica, médica, agronómica o evolutiva pueden generar datos importantes para responder las preguntas que se plantea la genética de poblaciones, la teoría de coalescencia y la filogeografía, y que aporten información sobre su uso, manejo y conservación.

14.8 CONCLUSIONES

El reto mayor al que se enfrenta el estudio de la diversidad genética en plantas y animales, desde el punto de vista de las capacidades en personal de investigación y de enfoques e infraestructura reside, por un lado, en la gran cantidad de especies por estudiar y, por otro, en generar una cantidad importante de datos en cada uno de los casos para poder hacer inferencia estadísticamente significativa acerca de los procesos poblacionales pasados, el impacto de los eventos recientes en la variación genética y la prospección hacia el futuro de estos procesos considerando diferentes escenarios climáticos. En resumen, este campo de conocimiento es uno de los más dinámicos en el ámbito internacional, por sus implicaciones y servicios que ofrece, y México cuenta con una muy limitada masa crítica que es necesario reforzar con la formación de recursos humanos.

NOTAS

- 1 Después de pasar por varias reestructuraciones y desarrollos, actualmente, y desde 1974, es la Universidad Autónoma de Chapingo.
- 2 El comité enviado por la Fundación Rockefeller estuvo constituido por E.C. Stakman, jefe de la División de Fitopatología de la Universidad de Minnesota, Paul Mangelsdorf, director del Museo Botánico de la Universidad de Harvard, y Richard Bradfield, jefe del Departamento de Agronomía de la Universidad de Cornell.
- 3 A diferencia del programa de mejoramiento del maíz, el programa de mejoramiento de trigo, a cargo de Norman Borlaug, sí tuvo éxito en México. Borlaug fue galardonado por sus trabajos en trigo con el Premio Nobel de la Paz en 1970.

4 Su programa de investigación y la Revolución Verde que se desprende de él han significado un gran avance tecnológico y de mejoramiento genético para los países en desarrollo. El CIMMYT es considerado un modelo de asistencia técnica y de apoyo en la investigación, y forma parte del consorcio Grupo Consultivo de Investigación Agrícola Internacional (CGIAR, por sus siglas en inglés), que provee apoyo y asistencia a las principales cosechas en todo el mundo.

REFERENCIAS

- Armendares, S., F. Salamanca y S. Frenk. 1971. Chromosome abnormalities in severe protein calorie malnutrition. *Nature* **232**: 271-273.
- Avise, J.C. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual Reviews in Genetics* **25**: 45-69.
- Avise, J.C. 1992. Molecular population structure and the biogeographic history of a regional fauna: A case history with lessons for conservation biology. *Oikos* **63**: 62-76.
- Avise, J.C. 2000. *Phylogeography: The history and formation of species*. Harvard University Press, Harvard.
- Avise, J.C., W.S. Nelson, B.W. Bowen y D. Walker. 2000. Phylogeography of colonially nesting seabirds, with special reference to global matrilineal patterns in the sooty tern (*Sterna fuscata*). *Molecular Ecology* **9**: 1783-1792.
- Avise, J.C., J. Arnold, R.M. Ball, E. Bermingham, T. Lamb *et al.* 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Reviews in Ecology and Systematics* **18**: 489-522.
- Barahona, A., y A.L. Gaona. 2001. The history of science and the introduction of plant genetics in Mexico. *History and Philosophy of the Life Sciences* **23**: 157-168.
- Barahona, A., S. Pinar y F.J. Ayala. 2003. *La genética en México. Institucionalización de una disciplina*. UNAM, México.
- Berli, P., y J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 4563-4568.
- Bermingham, E., y C. Moritz. 1998. Comparative phylogeography: Concepts and applications. *Molecular Ecology* **7**: 367-369.
- Bernardi, G., L. Findley y A. Rocha-Olivares. 2003. Vicariance and dispersal across Baja California in disjunct marine fish populations. *Evolution* **57**: 1599-1609.
- Bolívar, F., R.L. Rodríguez, P.J. Greene, M.C. Betlach, H.L. Heynker y H.W. Boyer. 1977. **Construction and characterization of new cloning vehicles. II. A multipurpose cloning system.** *Gene* **2**: 95-113.
- Cockerham, C.C. 1973. Analyses of gene frequencies. *Genetics* **74**: 679-700.
- Coello, G., A. Escalante y J. Soberón. 1993. Lack of genetic variation in *Lacandonia schismatica* (Lacandoniaceae: Triuridales) in its only known locality. *Annals of the Missouri Botanical Garden* **80**: 898-901.
- Corander, J., y J. Tang. 2007. Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* **205**: 19-31.
- Crow, J.F., y K. Aoki. 1984. Group selection for a polygenic behavioural trait: Estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences of the United States of America* **81**: 6073-6077.
- Dobzhansky, T., R. Félix, J. Guzmán, L. Levine, O. Olvera *et al.* 1975. Population genetics of Mexican *Drosophila* I: Chromosomal variation in natural populations of *Drosophila pseudoobscura* from central Mexico. *Journal of Heredity* **66**: 203-206.
- De Garay, A.L. 1960. *Programa de genética y radiobiología. Informe de labores 1960*. Comisión Nacional de Energía Nuclear, Archivo de Información, Biblioteca del ININ, México.
- De Garay, A.L. 1963. *Programa de genética y radiobiología. Informe de labores 1963*. Comisión Nacional de Energía Nuclear, Archivo de Información, Biblioteca del ININ, México.
- Eguiarte, L.E., A. Búrquez, J. Rodríguez, M. Martínez-Ramos, J. Sarukhán y D. Piñero. 1993. Direct and indirect estimates of neighborhood and effective population size in a tropical palm, *Astrocaryum mexicanum*. *Evolution* **47**: 75-87.
- Excoffier, L. 2001. Analysis of population subdivision, en D.J. Baldwin, C. Cannings y M. Bishop (eds.), *Handbook of statistical genetics*. John Wiley, Nueva York, pp. 271-308.
- Excoffier, L., G. Laval y J.A. Schneider. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.
- Excoffier, L., y G. Heckel. 2006. Computer programs for population genetics data analysis: A survival guide. *Nature Reviews Genetics* **7**: 745-758.
- Falush, D., M. Stephens y J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.
- Felsenstein, J. 1971. The rate of loss of multiple alleles in finite haploid populations. *Theoretical Population Biology* **2**: 391-403.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-781.
- Fisher, R.A. 1932. *The genetical theory of natural selection*. Oxford University Press, Londres.
- Fitzgerald, D. 1994. Exporting American agriculture: The Rockefeller Foundation in Mexico, en M. Cueto (ed.),

- Missionaries of science. The Rockefeller Foundation and Latin America.* Indiana University Press, Bloomington, pp. 72-96.
- Fracheboud, Y., J.M. Ribaut, M. Vargas, R. Messer y P. Stamp. 2002. Identification of quantitative traits loci for cold tolerance of photosynthesis maize (*Zea mays* L.). *Journal of Experimental Botany* **53**: 1967-1977.
- Fornoni, J., P.L. Valverde y J. Núñez-Farfán. 2004. Population variation in the cost and benefit of tolerance and resistance against herbivory in *Datura stramonium*. *Evolution* **58**: 1696-1704.
- Gillespie, J.H. 2004. *Population genetics: A concise guide.* The John Hopkins University Press, Baltimore.
- Griffiths, R.C. 1980. Lines of descent in the diffusion approximation of neutral Wright-Fisher models. *Theoretical Population Biology* **17**: 40-50.
- Grivet, D., y R.E.J. Petit. 2003. Chloroplast DNA phylogeography of the hornbeam in Europe: Evidence for a bottleneck at the outset of postglacial colonization. *Conservation Genetics* **4**: 47-56.
- Guzmán-Maldonado, S.H., O. Martínez, J.A. Acosta-Gallegos, F. Guevara-Lara y O. Paredes-López. 2003. Putative quantitative trait loci for physical and chemical components of common bean. *Crop Science* **43**: 1029-1035.
- Guzmán-Novoa, E., G.J. Hunt, R.E. Page, Jr., J.L. Uribe-Rubio, D. Prieto-Merlos y F. Becerra-Guzmán. 2005. Paternal effects on the defensive behavior of honeybees. *Journal of Heredity* **96**: 376-380.
- Hare, M.P. 2001. Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* **16**: 700-706.
- Hartl, D.L., y A.G. Clark, 2006. *Principles of population genetics.* 4 ed. Sinauer Associates, Sunderland.
- Hedrick, P.W. 2005. *Genetics of populations.* Jones and Bartlett Publishers, Inc., Londres.
- Hein, J., M.H. Schierup y C. Wiuf. 2005. *Gene genealogies, variation and evolution: A primer in coalescent theory.* Oxford University Press, Londres.
- Herrera, A.L. 1904. *Nociones de biología.* Imprenta de la Secretaría de Fomento, México. Edición facsimilar, Universidad Autónoma de Puebla, México, 1992.
- Herrera-Estrella, L., A. Depicker, M. Van Montagu y J. Schell. 1983. Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. *Nature* **303**: 209-213.
- Hey, J., y J. Wakeley. 1997. A coalescent estimator of the population recombination rate. *Genetics* **145**: 833-846.
- Holsinger, K.E., P.O. Lewis y K.D. Dipak. 2002. A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology* **11**: 1157-1164.
- Holsinger, K.E., y L.E. Wallace. 2004. Bayesian approaches for the analysis of population genetic structure: An example from *Platanthera leucophaea* (Orchidaceae). *Molecular Ecology* **13**: 887-894.
- Hudson, R.R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**: 1-44.
- Hudson, R.R. 2001. Two-locus sampling distribution and their application. *Genetics* **159**: 1805-1817.
- INIA. 1985. *Edmundo L. Taboada Ramírez: una semblanza, 1906-1983.* INIA, SARH (publicación especial), México.
- Itakura, K., T. Hirose, R. Crea, A.D. Riggs, H.L. Heynker et al. 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* **198**: 1056-1063.
- Kimura, M. 1953. "Stepping-stone" model of population. *Ann. Rep. Natl. Inst. Genet. Japan* **3**: 62-63.
- Knowles, L.L. 2004. The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology* **17**: 1-10.
- Knowles, L.L., y W.P. Maddison. 2002. Statistical phylogeography. *Molecular Ecology* **11**: 2623-2635.
- Kuhner, M.K., J. Yamato y J. Felsenstein. 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* **140**: 1421-1430.
- Kuhner, M.K., J. Yamato y J. Felsenstein. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**: 429-434.
- Ledesma, I., y A. Barahona. 2003. The institutionalization of biology in Mexico in the early 20th Century. The conflict between Alfonso Luis Herrera (1868-1942) and Isaac Ochoterena (1885-1950). *Journal of the History of Biology* **36**: 285-307.
- Levin, D.A. 1981. Dispersal versus gene flow in plants. *Annals of the Missouri Botanical Garden* **68**: 233-253.
- Lisker, R. 1981. *Estructura genética de la población mexicana. Aspectos médicos y antropológicos.* Salvat, México.
- Lynch, M., y T. Crease. 1990. The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* **7**: 377-394.
- Lyons-Weiler, J., y M.C. Milinkovitch. 1997. A phylogenetic approach to the problem of differential lineage sorting. *Molecular Biology and Evolution* **14**: 968-975.
- Maddison, W.P., y D.R. Maddison. 2004. MESQUITE: A modular system for evolutionary analysis. Disponible en <<http://mesquiteproject.org/mesquite/download/download.html>>.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209-220.
- McVean, G., P. Awadalla y P. Fearnhead. 2002. A Coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231-1241.
- Moritz, C., y D.P. Faith. 1998. Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology* **7**: 419-429.
- Nason, J.D., J.L. Hamrick y T.H. Fleming. 2002. Historical vicariance and postglacial colonization effects on the

- evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. *Evolution* **56**:2214-2226.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70**:3321-3323.
- Nei, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* **41**:225-233.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, Nueva York.
- Neigel, J.E. 2002. Is F_{ST} obsolete? *Conservation Genetics* **3**:167-173.
- Neigel, J.E., R.M. Ball y J.C. Avise. 1991. Estimation of single generation migrations distances from geographic variation in animal mitochondrial DNA. *Evolution* **45**:423-432.
- Newton, A.C., T.R. Allnutt, A.C.M. Gillies, A.J. Lowe y R.A. Ennos. 1999. Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology and Evolution* **14**:140-145.
- Nielsen, R., y J. Wakeley. 2001. Distinguish migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**:885-896.
- Panchal, M., y M.A. Beaumont. 2007. The automation and evaluation of nested clade phylogeographic analysis. *Evolution* **61**:1466-1480.
- Pearse, D.E., y K.A. Crandall. 2004. Beyond F_{ST} : Analysis of population genetic data for conservation. *Conservation Genetics* **5**:585-602.
- Petit, R.J. 2008. The coup de grâce for the nested clade phylogeographic analysis?, *Molecular Ecology* **17**:516-518.
- Piry, S., A. Alapetite, J.M. Cornuet, D. Paetkau, L. Baudouin et al. 2004. GeneClass2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**:536-539.
- Pritchard, J.K., M. Stephens y P.J. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945-959.
- Ramos-Onsins, S.E., y J. Rozas. 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**:2092-2100.
- Riddle, B.R., D.J. Hafner y L.F. Alexander. 2000a. Comparative phylogeography of Baileys' pocket mouse (*Chaetodipus baileyi*) and the *Peromyscus eremicus* species group: Historical vicariance of the Baja California Peninsular Desert. *Molecular Phylogenetics and Evolution* **17**:161-172.
- Riddle, B.R., D.J. Hafner, L.F. Alexander y J.R. Jaeger. 2000b. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences of the United States of America* **97**:14438-14443.
- Rieseberg, L.H., y J.M. Burke. 2001. The biological reality of species: Gene flow, selection and collective evolution. *Taxon* **50**:235-255.
- Rocha-Olivares, A., J.W. Fleeger y D.W. Foltz. 2001. Decoupling of molecular and morphological evolution in deep lineages of a meiobenthic harpacticoid copepod. *Molecular Biology and Evolution* **18**:1088-1102.
- Rogers, A.R., y H.C. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**:552-569.
- Rousset, F. 2001. Inferences from spatial population genetics, en D.J. Baldwin, M. Bishop y L. Cannings (eds.), *Handbook of statistical genetics*. John Wiley, Nueva York, pp. 239-270.
- Rozas, J., J.C. Sánchez-Del Barrio, X. Messeguer y R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**:2496-2497.
- Salamanca, F., y S. Armendares. 1995. The development of human genetics in Mexico. *Archives of Medical Research* **26**:55-62.
- Salazar-Mallén, M., C. Arteaga, E. Ugalde y A. Vélez-Orozco. 1952. Blood agglutinogens of Mexicans. *Annals of Eugenics* **16**:351-355.
- Schaal, B.A., J.F. Gaskin y A.L. Caicedo. 2003. Phylogeography, haplotype trees, and invasive plant species. *Journal of Heredity* **94**:197-204.
- Schaal, B.A., D.A. Hayworth, K.M. Olsen, J.T. Rauscher y W.A. Smith. 1998. Phylogeographic studies in plants: Problems and prospects. *Molecular Ecology* **7**:465-474.
- Slatkin, M. 1981. Estimating levels of gene flow in natural populations. *Genetics* **99**:323-335.
- Slatkin, M. 1985. Gene flow in natural populations. *Annual Reviews in Ecology and Systematics* **16**:393-430.
- Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**:264-279.
- Slatkin, M. 1994. Gene flow and population structure, en L.A. Real (ed.), *Ecological genetics*, Princeton University Press, Princeton, pp. 3-18.
- Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**:457-462.
- Slatkin, M., y N.H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* **43**:1349-1368.
- Taboada, E. 1938. *Apuntes de genética*. Escuela Nacional de Agricultura, Chapingo.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595.
- Templeton, A.R. 1995. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping or DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the apoprotein E locus. *Genetics* **140**:403-409.
- Templeton, A.R. 1998. Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. *Molecular Ecology* **7**:381-397.

- Templeton, A.R. 2001. Using phylogeographic analysis of gene trees to test species status and processes. *Molecular Ecology* **10**:779-791.
- Templeton, A.R. 2004. Statistical phylogeography: Methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**:789-809.
- Templeton, A.R. 2008. Nested clade analysis: An extensively validated method for strong phylogeographic inference. *Molecular Ecology* **17**:1877-1880.
- Templeton, A.R., E. Boerwinkle y C.F. Sing. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117**:343-351.
- Templeton, A.R., C.F. Sing, A. Kessling y S. Humphries. 1988. A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. *Genetics* **120**:1145-1154.
- Templeton, A.R., K.A. Crandall y C.F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**:619-633.
- Templeton, A.R., y C.F. Sing. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134**:659-669.
- Templeton, A.R., E. Routman y C.A. Phillips. 1995. Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140**:767-782.
- Velázquez-Arellano, A., S. Cederbaum, F. Salamanca-Gómez, M. Rodríguez y A. Velázquez. 1986. A new frontiers in human genetics and their implications. *Gaceta Médica de México* **122**:123-134.
- Wakeley, J. 2004. Metapopulation models for historical inference. *Molecular Ecology* **13**:865-875.
- Wakeley, J. 2008. *Coalescent theory: An introduction*. Roberts & Company Publishers, Greenwood Village, Colorado.
- Whitlock, M.C., y D.E. McCauley. 1999. Indirect measures of gene flow and migration: F_{ST} not equal to $1/(4Nm + 1)$. *Heredity* **82**:117-125.
- Weir, B.S. 1996. *Genetic data analysis II: Methods fo discrete population genetic data*. Sinauer Associates, Sunderland.
- Weir, B.S., y C.C. Cockerham. 1984. Estimating F -statistics for the analysis of population structure. *Evolution* **38**:1358-1370.
- Workman, P.L., y J.D. Niswander. 1970. Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. *The American Journal of Human Genetics* **22**:24-29.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97-101.
- Wright, S. 1943. Isolation by distance. *Genetics* **28**:114-138.
- Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* **15**:323-354.