

# Flujo génico en poblaciones naturales\*

MONTGOMERY SLATKIN\*\*

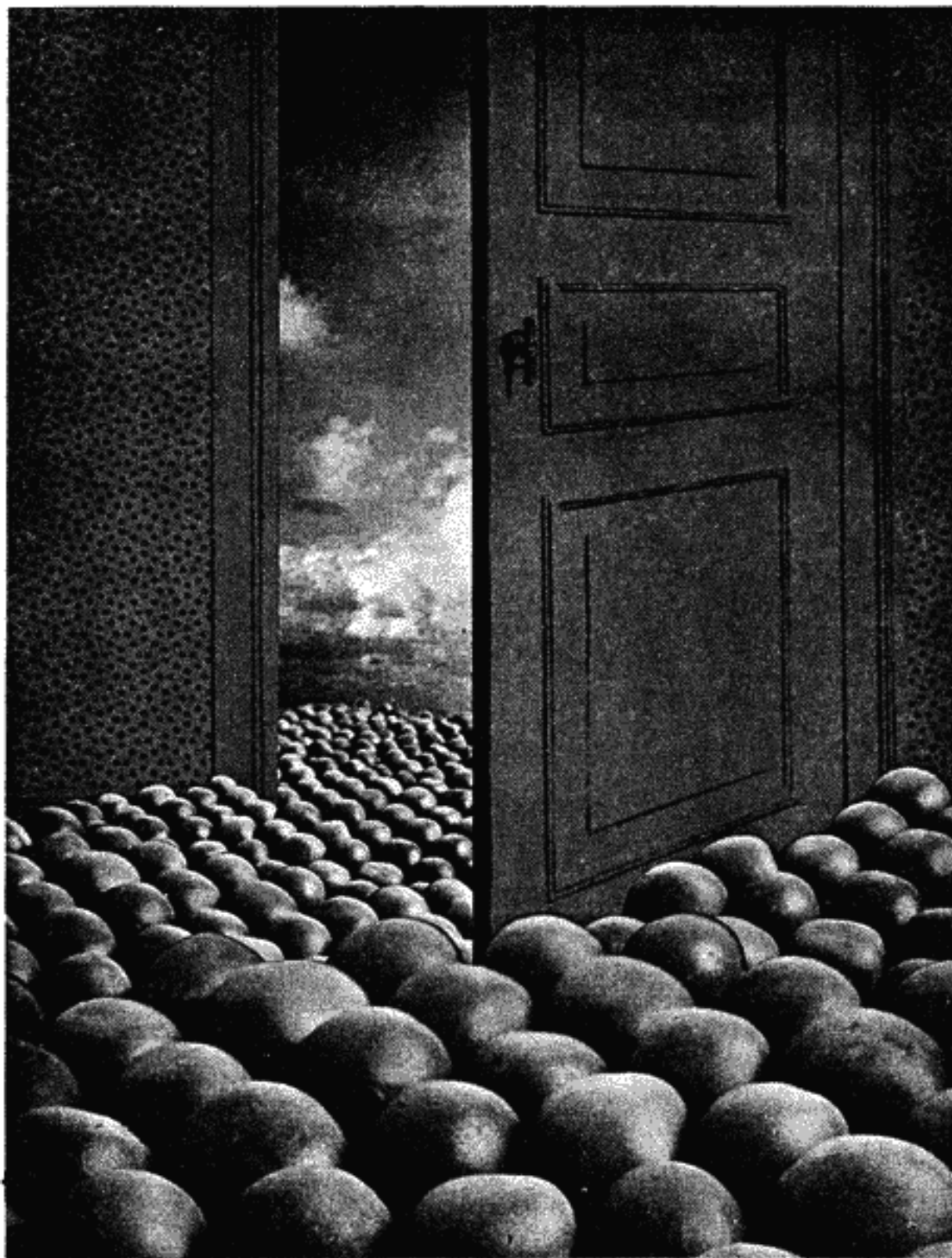
**E**l objetivo general de la genética evolutiva es entender el papel que las diferentes fuerzas tienen en la evolución de caracteres fenotípicos. Las principales fuerzas son la selección natural, la deriva génica, la mutación, la recombinación y el flujo génico. Recientemente, los biólogos moleculares han descubierto nuevas fuerzas de evolución génica tales como la transposición y la conversión génica. Sin embargo, aún no se sabe si estas fuerzas son importantes en la evolución de caracteres observables o no.

En esta ocasión voy a centrarme en el flujo génico como una fuerza evolutiva. Siempre que hay movimiento de genes de una parte del área de distribución de una especie a otra hay flujo génico. Un tipo de flujo génico opera cuando individuos de una población se mueven a otra. Este es el tipo de flujo génico que ocurre en una especie con poblaciones estables que persisten durante largos periodos de tiempo. Existen otros tipos de flujo génico que ocurren cuando se establecen poblaciones nuevas. En muchas especies, incluyendo malezas y especies parásitas, las poblaciones no persisten durante largos periodos de tiempo; por el contrario, hay una extinción constante de poblaciones y una colonización por nuevas poblaciones. El flujo génico puede ocurrir durante la colonización por nuevas poblaciones, porque los genes de una población que ha colonizado recientemente pueden ser una mezcla de los genes de otras dos poblaciones o más. Durante la expansión geográfica de una especie también hay flujo génico. Tanto

la colonización como la expansión de los rangos de distribución de las especies, son motivo de estudio para los ecólogos; más que para los biólogos evolucionistas. Es por lo que estas dos formas de flujo génico son ignoradas ocasionalmente.

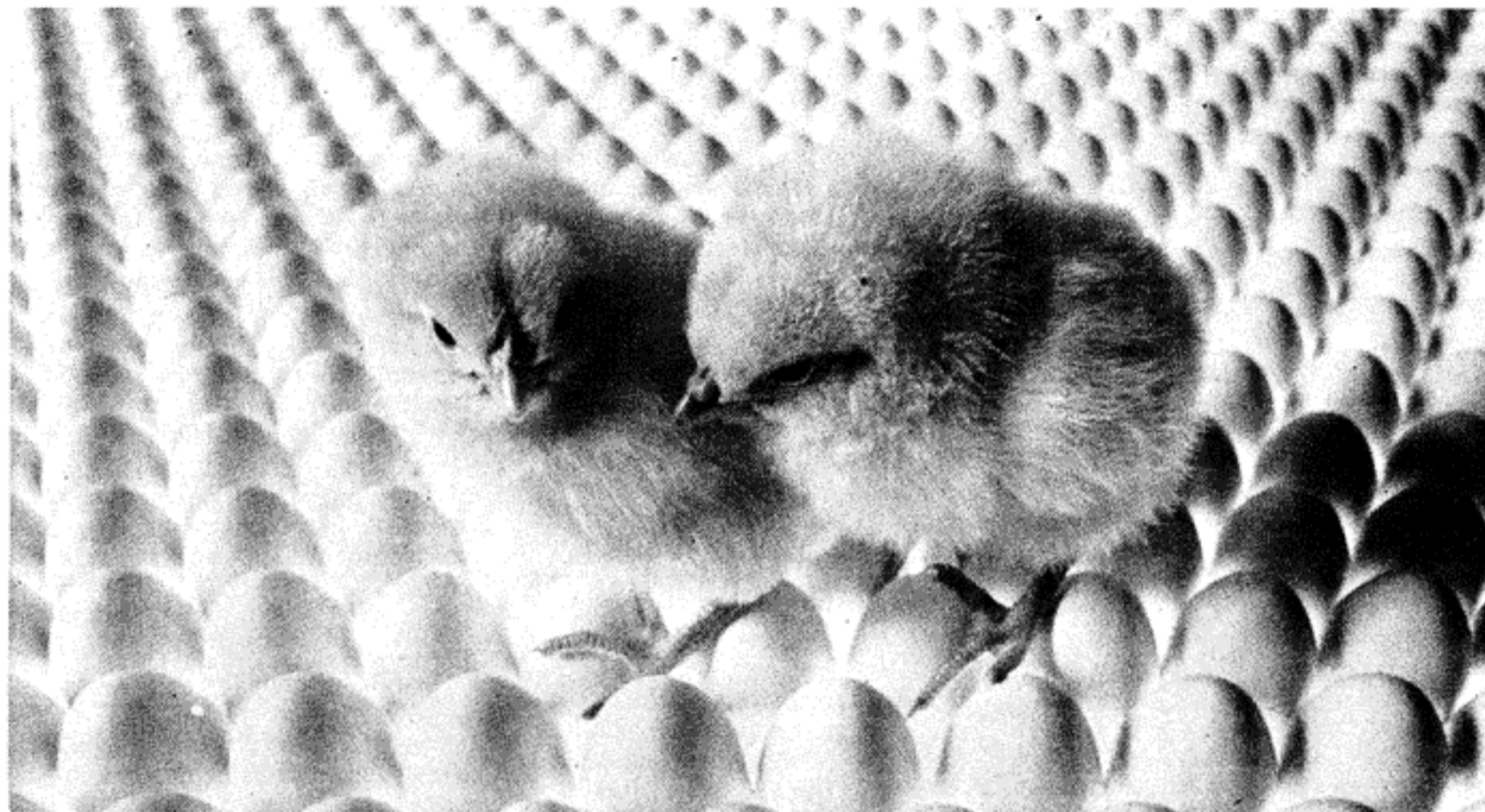
## CONSECUENCIAS GENÉTICAS DEL FLUJO GÉNICO

El flujo génico tiende a reducir las diferencias genéticas entre poblaciones locales. Qué tanto se reducen las diferencias genéticas entre las poblaciones, depende



\* Conferencia impartida en la Facultad de Ciencias UNAM, el 10 de Julio de 1989.

\*\* Universidad de California en Berkeley, EE.UU.



de las fuerzas que están actuando para causar diferenciación entre las poblaciones locales. La selección natural que favorece la adaptación local en diferentes condiciones ambientales locales, va a tender a actuar en favor de alelos diferentes en un locus genético, si dichos alelos causan adaptaciones locales. Si no existe flujo génico, habrán diferentes alelos en diferentes localidades. J.B.S. Haldane fue el primero en preguntarse cuánto flujo génico era necesario para prevenir la evolución de adaptaciones locales. Este autor examinó el siguiente ejemplo que ilustra el resultado principal: supongamos que en una determinada población un alelo de un locus particular tiene una adecuación de  $1+s$  en relación a otro alelo, pero que en cada generación una fracción  $m$  de esta población es sustituida por inmigrantes de una población en la que el alelo desventajoso se encuentra fijo. Haldane demostró que la frecuencia en equilibrio del alelo desventajoso es aproximadamente  $1-m/s$  si  $m < s$  y  $0$  si  $m > s$ . En otras palabras, solamente en los casos en que la tasa de inmigrantes sea menor a la intensidad de selección podrán haber adaptaciones locales. Se ha obtenido el mismo resultado con modelos más realistas y generales.

En la ausencia de flujo génico, la deriva génica conducirá también a diferencias locales en las frecuencias alélicas. En 1931, Sewall Wright propuso un modelo simple del balance entre el flujo génico y la deriva génica. Este autor asumió la existencia de un gran número de poblaciones locales, cada una de ellas con  $N$  individuos, y asumió que en cada

generación una fracción  $m$  de individuos eran reemplazados por un conjunto de individuos, escogidos al azar del resto de las poblaciones. Este es el famoso "modelo de islas" para la estructura poblacional. Se le conoce con ese nombre, porque se puede considerar que cada población se encuentra en una isla de un gran archipiélago. Wright demostró que la varianza entre las poblaciones para la frecuencia de un alelo,  $\sigma_p^2$ , es aproximadamente:

$$\sigma_p^2 = \frac{\bar{p}(1-\bar{p})}{1+4Nm}$$

donde  $p$  es la frecuencia alélica promedio en la población. Si  $4Nm$  es mucho mayor que uno,  $\sigma_p^2$  es pequeña, lo que implica que hay poca diferencia genética entre las poblaciones. Si  $4Nm$  es mucho menor que uno,  $\sigma_p^2$  es casi  $p(1-p)$ , lo que implica que la frecuencia alélica en la población es uno o cero.

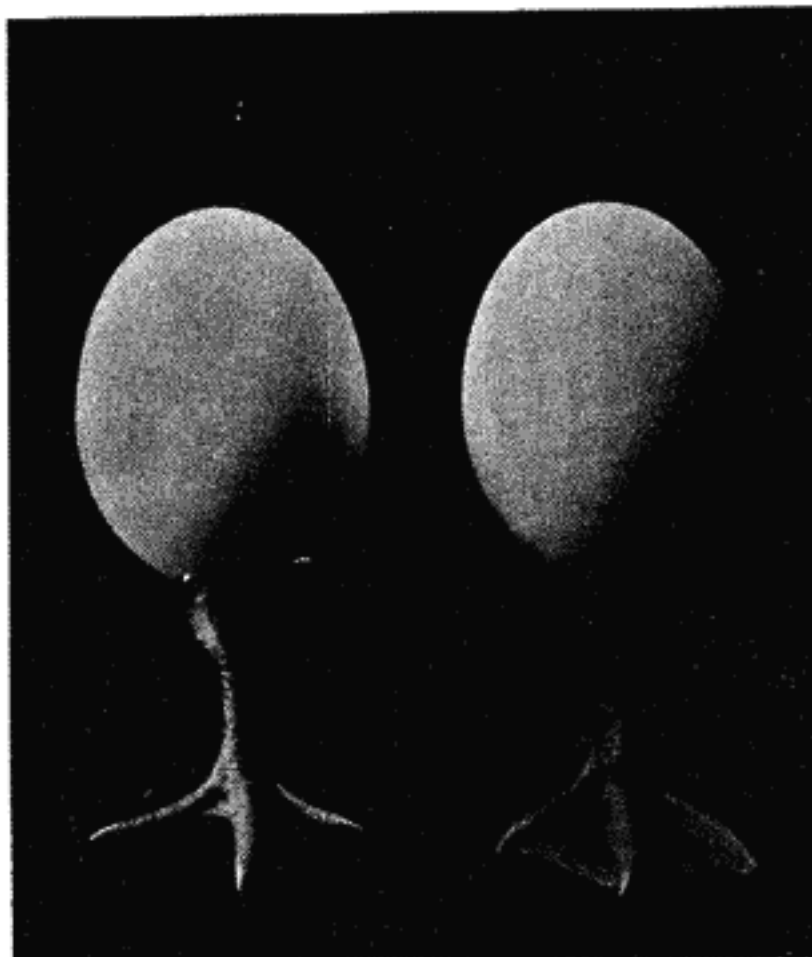
Como hemos visto, el flujo génico interactúa de manera diferente con la selección natural que con la deriva génica. En el caso de la selección natural el resultado depende del cociente  $s/m$ , y en el caso de la deriva génica el resultado depende del producto  $Nm$ . Por esta diferencia, es imposible decir si el flujo génico es importante o no en una especie particular. La respuesta dependerá de qué otras fuerzas evolutivas se consideren. Dada una tasa de inmigración  $m$  los alelos de loci fuertemente seleccionados a favor (valores altos de  $s$ ), tendrán altas frecuencias, aun si las tasas de flujo génico son elevadas, mientras que en las mismas poblaciones, las frecuencias de ale-

los de loci neutrales o de aquellos con coeficiente de selección bajo, estarán determinadas primordialmente por el flujo génico de otras poblaciones.

#### EL FLUJO GÉNICO COMO FUERZA EVOLUTIVA

La dificultad teórica de decidir la importancia relativa del flujo génico en cada especie, no ha impedido que los biólogos evolucionistas argumenten acerca de la importancia global del flujo génico. Como en la mayoría de las áreas de la biología evolutiva, también en este tema existen dos opiniones extremas. La primera de ellas es que el flujo génico es una fuerza importante en casi todas las especies. Ernest Mayr, en su libro *Especies animales y Evolución* (1963), sostiene que el flujo génico entre las poblaciones de una misma especie, provoca que éstas estén generalmente ligadas entre sí y conformen una sola unidad evolutiva. Mayr argumentaba que la interrupción del flujo génico es el primer paso hacia la formación de especies distintas, lo que constituye la teoría de "especiación alopátrica". En este sentido, Mayr continuaba la tradición de Darwin, quien enfatizó la importancia del aislamiento para la formación de nuevas variedades y de nuevas especies. En libros posteriores, Mayr ha cambiado de opinión, pero aún se pronuncia a favor del flujo génico.

La visión contraria sostiene que el flujo génico no es importante para determinar las frecuencias génicas de la mayoría de las poblaciones, en casi todas de las especies. En 1969 Paul Ehrlich y Peter Raven publicaron un artículo trascendente



en *Science*, en el que se manifestaban en contra de lo que en aquel tiempo era la visión ortodoxa de Mayr. Ehrlich y Raven afirmaban que la mayoría de los trabajos de campo demostraban que era poco frecuente que los individuos se dispersaran distancias largas, demasiado raro como para que el flujo génico causado por la dispersión fuera importante. El ejemplo de la mariposa *Euphydryas editha*, que Ehrlich ha estudiado extensivamente, fue uno de los que utilizaron en su artículo. Ehrlich estudió varias poblaciones de esta mariposa y encontró que, aun cuando las mariposas tenían la capacidad de volar a grandes distancias, raramente se movía de una población a otra que se encontraba a solamente 100 m de distancia. Además encontró que las épocas reproductivas eran distintas entre las poblaciones y esto impedía que las mariposas que lograban llegar a otras poblaciones se reprodujeran. Con base en estas observaciones Ehrlich concluyó que en esta especie, que se encuentra en gran parte del Suroeste de los Estados Unidos, el flujo génico no podría ser importante en lo absoluto.

La pregunta de qué tan importante es el flujo génico en las poblaciones naturales es interesante en sí misma. Es importante también para entender el papel potencial de otras fuerzas evolutivas. Si las poblaciones no están ligadas entre sí por flujo génico, entonces es necesario recurrir a explicaciones alternativas, para poder entender la similitud morfológica de poblaciones de la misma especie, en diferentes regiones geográficas. Si en efecto, el flujo génico no está operando en las poblaciones de la mariposa, tal como argu-

menta Ehrlich, entonces ¿qué fuerzas están determinando que las mariposas de diferentes poblaciones sean tan parecidas? ¿Acaso la selección natural favorece el mismo fenotipo en diferentes localidades o existen "restricciones de desarrollo" que impiden el cambio morfológico bajo diferentes circunstancias?

#### MÉTODOS PARA ESTIMAR EL FLUJO GÉNICO

Existen dos formas generales para estimar el flujo génico en poblaciones naturales: métodos indirectos y métodos directos. Los métodos directos implican observaciones detalladas de la especie. En algunas especies es posible observar el movimiento de adultos individuales o de semillas, o es posible marcar individuos y recapturarlos después de que ha ocurrido la dispersión. Ehrlich pudo marcar a los individuos y después encontrar la fracción de ellos que se había movido entre poblaciones. En otras especies es posible introducir algún alelo raro, como un carácter dominante que se expresa en un carácter claramente observable, como un color raro, y luego se puede seguir el movimiento de este alelo por la población. Técnicas moleculares modernas, tales como la secuenciación de ADN y las huellas genéticas, permiten la asignación de paternidad en algunas especies, y por lo tanto, permiten calcular las distancias de dispersión de la progenie.

Por otro lado, un método indirecto es aquel que utiliza observaciones de las frecuencias alélicas u otras características medibles de las especies para deducir qué cantidad de flujo génico debe haber ocurrido para producir los patrones espaciales observados. Los métodos indirectos dependen de modelos matemáticos que predicen qué patrones se observarían en las diferentes tasas de flujo génico. Wright introdujo uno de estos métodos basándose en su modelo de islas. Wright definió  $F_{ST}$  como la cantidad de consanguinidad causada por las diferencias en las frecuencias génicas entre las poblaciones y encontró que en un solo locus:

$$F_{ST} = \frac{\sigma_p^2}{p(1-p)}$$

donde, como en la fórmula anterior,  $p$  y  $\sigma_p^2$  son la media y la varianza de las frecuencias alélicas entre poblaciones. Como mencioné anteriormente, Wright había predicho el valor de  $\sigma_p^2$  en equilibrio para el modelo de islas, de tal manera que pudo predecir el valor de  $F_{ST}$  que se esperaría para un locus neutral:

$$F_{ST} = \frac{1}{1 + 4Nm}$$

Un método indirecto para estimar  $Nm$ , se puede derivar calculando  $F_{ST}$  para una especie a partir del promedio y la varianza de las frecuencias alélicas y después resolviendo esta ecuación para encontrar:

$$Nm = \frac{1}{4} \left( \frac{1}{F_{ST}} - 1 \right)$$

Hay que notar que este método no estima  $m$  independientemente, sino que estima el producto de  $Nm$ . Por el momento no hay ningún método indirecto para calcular  $m$  por separado.

El procedimiento para utilizar este método es simple. Un estudio de campo para una especie nos proporciona una tabla de frecuencias alélicas para varios loci, generalmente estudiados por medio de electroforesis. Después se estima un valor de  $F_{ST}$  con el cual se pueda estimar  $Nm$ . Hay diferentes formas de estimar  $F_{ST}$  a partir de este tipo de tablas, pero generalmente todos los estimados deben dar valores muy similares de  $Nm$ . Los dos estadísticos utilizados más comúnmente son el estadístico  $G_{ST}$  de Nei y el estadístico de Weir y Cockerham.

Yo he desarrollado otro método indirecto que se basa en las frecuencias de alelos raros. Tiene propiedades muy similares al  $F_{ST}$  de Wright, y por lo tanto, los voy a analizar conjuntamente. Para que estos métodos sean precisos deben proporcionar estimaciones de  $Nm$  también precisas, para diversas condiciones. Las poblaciones hacen caso omiso de las suposiciones de los matemáticos y por lo tanto no se ajustan a los modelos idealizados que se proponen; no se encuentran en islas que permitan el flujo génico a tasas iguales entre todos los pares de islas; no son todas del mismo tamaño ni permanecen siempre del mismo tamaño. Más aún, es muy poco probable que todos aquellos loci que se puedan muestrear sean completamente neutrales. Para que cualquier método indirecto proporcione estimaciones precisas de  $Nm$  no debe ser demasiado sensible a desviaciones de los supuestos en los cuales está basado.

Tabla 1

$m(4Nm)$	$\bar{p}(1)$	$\theta$	$G_{ST}$
0.001 (0.512)	1.31 (0.53)	1.80 (0.92)	1.85 (0.85)
0.01 (5.12)	5.61 (0.32)	6.73 (0.40)	6.36 (0.35)
0.1 (51.2)	33.3 (0.10)	289.1 (0.90)	43.82 (0.09)

Estimaciones de  $4Nm$  usando diferentes métodos, con datos simulados para el "modelo de islas" de estructura de población. Para el modelo de alelos infinitos, se corrieron 10 juegos de 10 réplicas. Los números entre paréntesis son los coeficientes de variación de los estimadores de  $4Nm$  para cada juego de réplicas. Se usaron los estimadores que se discuten en el texto, basándose en los mismos datos:

$p(1)$  indica que se usó la frecuencia promedio de alelos privados;  $\theta$  indica que se usó el estimador de  $F_{ST}$  de Weir y Cockerham (1984); y  $G_{ST}$  indica que se usó el estimador de  $F_{ST}$  de Nei (1973). En todos los casos hubo 100 demos con 128 individuos por demo. Las muestras de 25 individuos se tomaron de 10 demos escogidos al azar.

### MÉTODOS INDIRECTOS DE $Nm$

En los últimos años me he dedicado a investigar si es posible obtener estimaciones precisas de  $Nm$  a partir de valores de  $F_{ST}$  y de los alelos raros. Mi enfoque para resolver este problema se basa en el uso de simulaciones de flujo génico, deriva génica, mutaciones y selección natural. Utilizo estos modelos de simulación para derivar datos hipotéticos, a los cuales se les pueden aplicar diferentes métodos indirectos de estimación de flujo génico. De esta forma puedo evaluar la validez de los diferentes métodos, porque conozco el tamaño real de las poblaciones y las tasas de migración que han generado las frecuencias alélicas. No voy a describir todo este trabajo con detalle, pero voy a presentar algunos de los resultados que he obtenido.

La primera pregunta que me hice es si estos métodos son precisos cuando las poblaciones satisfacen las suposiciones a partir de las cuales fueron derivados. Simulé poblaciones asumiendo una estructura de islas en la que cada población tenía  $N$  individuos. Asumí que en cada generación una fracción  $m$  de cada población era sustituida por individuos escogidos al azar del resto de las poblaciones. En cada generación, cada gene tenía una probabilidad  $\mu$  de mutar. También partí de que todos los alelos eran neutrales. Posteriormente asumí que 25 individuos se escogían de 10 de las poblaciones escogidas al azar y que se les aplicaban diferentes métodos indirectos para estimar  $Nm$ . Estos tamaños de muestra son los que generalmente se utilizaban en estudios de campo.

Se muestran algunos resultados en la Tabla 1. En esta tabla se han utilizado tres métodos diferentes. Utilicé dos métodos para estimar  $F_{ST}$ , la  $G_{ST}$  de Nei y Weir y la  $\theta$  de Cockerham, y utilicé mi método que usa las frecuencias de los alelos que se encuentran sólo en una de

las poblaciones que se muestrean,  $p(1)$ . Si el valor de  $Nm$  está cerca de uno, cualquiera de estos métodos proporciona estimaciones precisas de  $Nm$ . Para valores de  $Nm$  muy pequeños, los tres métodos sobreestiman a  $Nm$  por un factor de 2. Para valores muy grandes de  $Nm$  mi método y el de  $G_{ST}$  tienden a subestimar un poco a  $Nm$ , mientras que a  $\theta$  lo sobrestima mucho.

Estos resultados muestran que el uso de estos métodos indirectos presenta problemas estadísticos no anticipados, pero funciona bien para parámetros de mucho interés. Indican de manera bastante precisa si el flujo génico es más fuerte o más débil que la deriva génica, aunque la precisión disminuye a medida que la diferencia en la magnitud de las fuerzas se hace más grande.

La siguiente pregunta es si los estimadores indirectos se vuelven menos precisos cuando la selección natural actúa sobre estos loci. Para responder a esta pregunta asumí en mis simulaciones, que cada heterocigoto tenía una adecuación relativa de  $1+s$  con respecto a todos los homocigos. Se muestran algunos resultados en la Tabla 2. Si  $s$  es positiva, la selección tiene un efecto muy pequeño y si es negativa hay una tendencia a sobreestimar a  $Nm$ . Por lo tanto, si sólo se utiliza alguno de estos métodos, se pueden obtener estimaciones sesgadas de  $Nm$ , en el caso en que haya habido presiones de selección. Sin embargo, para que dicho sesgo fuese importante todos los loci muestreados deberían ser afectados por selección en la misma dirección; y esto es muy poco probable. Por la misma razón la selección que varía geográficamente y que produce clinas en la frecuencia alélica, es poco importante como fuente de sesgo.

La tercera pregunta se refiere al efecto de algún tipo de estructuración geográfica en la colección de poblaciones lo-

cales. El modelo de islas representa el extremo de flujo génico a larga distancia. Los individuos nacidos en cada una de las poblaciones pueden llegar a otra cualquiera con la misma probabilidad. En el extremo contrario, supongamos que las poblaciones locales se encuentran en un arreglo bidimensional en que el flujo génico ocurre solamente entre las poblaciones adyacentes. Este es el modelo de migración de "stepping stone", que representa el extremo inferior de las distancias de migración. Yo utilicé este modelo de migración en mis simulaciones para averiguar cómo las estimaciones de  $Nm$ , dependen de los valores de  $Nm$  y de las localidades muestreadas. Asumí que se muestreaban 9 poblaciones de un arreglo cuadrado con  $k$  poblaciones entre las parejas de poblaciones más cercanas de las muestreadas.

La Figura 1 muestra cómo los estimados de  $Nm$  dependen de  $k$  para cada valor particular de  $Nm$ . Encontré que cuando  $k$  es grande, se obtienen estimaciones precisas, pero la  $Nm$  se tiende a sobreestimar cuando  $k$  es pequeña. La conclusión es que se tiende a sobreestimar a  $Nm$  cuando se consideran únicamente poblaciones cercanas.

### MÉTODOS DIRECTOS E INDIRECTOS

Tanto los métodos directos para estimar el flujo génico, como los indirectos, tienen ventajas y desventajas. Los métodos directos tienen la ventaja de observar la dispersión y saber bajo qué condiciones ecológicas opera. Estos métodos tienen la desventaja de estar restringidos, tanto en tiempo, como en espacio. El área geográfica en la cual se puede observar la dispersión, es generalmente una fracción pequeña del rango de distribución de las especies. Los individuos que se dispersan fuera de dicha área no son considerados. Generalmente se estudia la dispersión únicamente para una o dos temporadas reproductivas, pero la dispersión a larga distancia puede ocurrir esporádicamente y no ser detectada por

Figura 1.



estudios de corto plazo. Además, los métodos directos únicamente se llevan a cabo durante épocas consideradas como típicas o normales para una especie. Sin embargo pueden ocurrir dispersiones a larga distancia, únicamente durante épocas excepcionales, tales como huracanes, grandes tormentas o incendios.

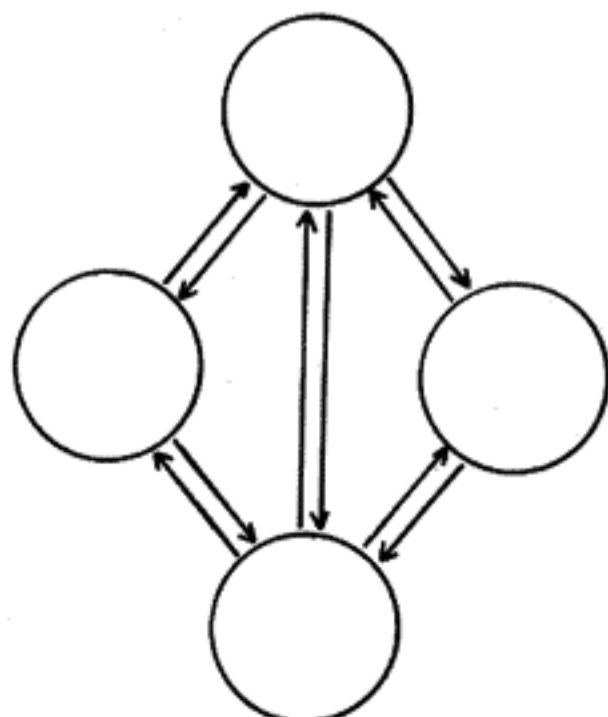
Los métodos indirectos tienen la ventaja de que en ellos las estimaciones de flujo génico de las tasas promedio son para muchos años, no solamente para una o dos generaciones. Por lo tanto estiman la tasa promedio de dispersión y no la tasa durante un periodo particular. Tienen la desventaja de que dependen de modelos matemáticos que no se pueden probar independientemente. Yo creo que no es correcto decir que un tipo de método es mejor que otro. Cada uno de ellos proporciona diferente tipo de información, y la comparación de resultados al aplicar ambos métodos puede proporcionar alguna visión acerca de las especies que se estudian.

#### APLICACIONES

Voy a discutir brevemente un ejemplo de una especie a la cual se le aplicaron ambos métodos (directo e indirecto). Es la especie que ya he mencionado: la mariposa de Ehrlich. Sus estudios directos han mostrado convincentemente que la dispersión entre poblaciones es en realidad muy rara. Él también ha usado electroforesis para estimar las frecuencias alélicas de varios loci de la población de esta especie, por toda California. Yo he calculado el  $F_{ST}$  de estos datos y encontré que es aproximadamente 3.7. Este valor es muy grande para ser consistente con las observaciones de Ehrlich.

La diferencia entre las conclusiones del método directo e indirecto es importante, yo pienso, ya que nos dice algo muy útil acerca de esta especie. En el

Modelo de islas: la migración es la misma desde y hacia cada subpoblación, independientemente de la distancia.



m (4Nm)	s	$\bar{p}(1)$	$\theta$	$G_{st}$
0.01 (5.12)	0	5.31 (0.22)	6.93 (0.21)	6.60 (0.19)
	+0.01	9.46 (0.50)	10.95 (0.77)	9.10 (0.61)
	-0.01	5.01 (0.24)	6.20 (0.15)	5.99 (0.14)
0.1 (512)	0	31.68 9(0.09)	154.53* (0.31)	42.18 (0.15)
	+0.01	33.70 (1.25)	44.18 (0.20)	44.18 (0.20)
	-0.01	24.26 (0.10)	228.68 (0.37)	44.00 (0.06)
0.001 (0.512)	0	0.94 (0.24)	1.33 (0.31)	1.41 (0.31)
	+0.01	1.57 (0.60)	1.45 (0.76)	1.52 (0.72)
	-0.01	0.48 (0.16)	1.00 (0.13)	1.07 (0.04)

Estimados de 4Nm, usando diferentes métodos con datos simulados para el "modelo de islas" de estructura de población. En todos los casos se corrieron 10 juegos de 10 réplicas. Los números entre paréntesis son los coeficientes de variación de los estimadores de 4Nm. En todos los casos  $d = 100$ ,  $d_{sam} = 10$ ,  $N = 128$ ,  $N_{sam} = 25$ , y se usó el modelo de alelos infinitos de mutación con  $\mu = 0.0001$

\* Se ignoró una estimación negativa de Nm

momento actual no hay esencialmente flujo génico. Sin embargo, hubo flujo génico en un pasado muy cercano, ya que su efecto todavía está presente. La explicación más probable de esto es que la especie ha pasado por un rango de expansión reciente durante algunos años, probablemente durante alguno de los periodos en los que el Valle de California tenía una mayor precipitación.

Algunas otras especies muestran diferencias entre los métodos directos e indirectos. Dobzhansky y Wright llevaron a cabo un estudio directo de dispersión con la mosca de la fruta, *Drosophila pseudoobscura*, en el Oeste de los Estados Unidos y encontraron que la dispersión ocurrió a través de unos cientos de metros; sin embargo, la especie es genéticamente uniforme en un rango de cientos de kilómetros.

No todas las especies muestran estas diferencias. El mejillón *Mytilus edulis*, tiene larvas que pueden flotar en el océano por varios meses antes de establecerse. Aunque no es posible llevar a cabo un estudio directo de dispersión de esta especie, es razonable asumir que las distancias lejanas de dispersión de las larvas son frecuentes. Mi análisis de frecuencias alélicas de esta especie, confirma que las tasas de flujo génico son altas. En contraste, hay varias especies de salamandras en donde los estudios directos de dispersión muestran que ésta es muy rara y en los cuales los valores de mis estimadores indirectos de Nm son mucho menores que uno.

#### CONCLUSIÓN

Finalmente, mi conclusión personal es que hay métodos indirectos útiles para

estimar la tasa de flujo génico en poblaciones naturales. Estos métodos, combinados con información obtenida de métodos directos, son útiles para determinar si las similitudes genéticas en las poblaciones reales son debidas al flujo génico real o a la dispersión pasada y el rango de expansión.

La aplicación de estos métodos a los datos obtenidos en las poblaciones naturales, muestran que algunas especies no están en equilibrio genético. En lugar de esto, parece que los eventos históricos han ocurrido como una expansión de rango reciente y que la historia actual es responsable principalmente de la distribución espacial actual de las frecuencias génicas. No está claro si esta situación es típica de muchas especies, porque ha habido muy pocos estudios de campo en los cuales se puedan obtener los valores de los estimadores directos e indirectos del nivel de flujo génico.

Modelo de "Adoquín" ("Stepping-Stone"): la migración se da entre subpoblaciones vecinas o adyacentes, exclusivamente.

