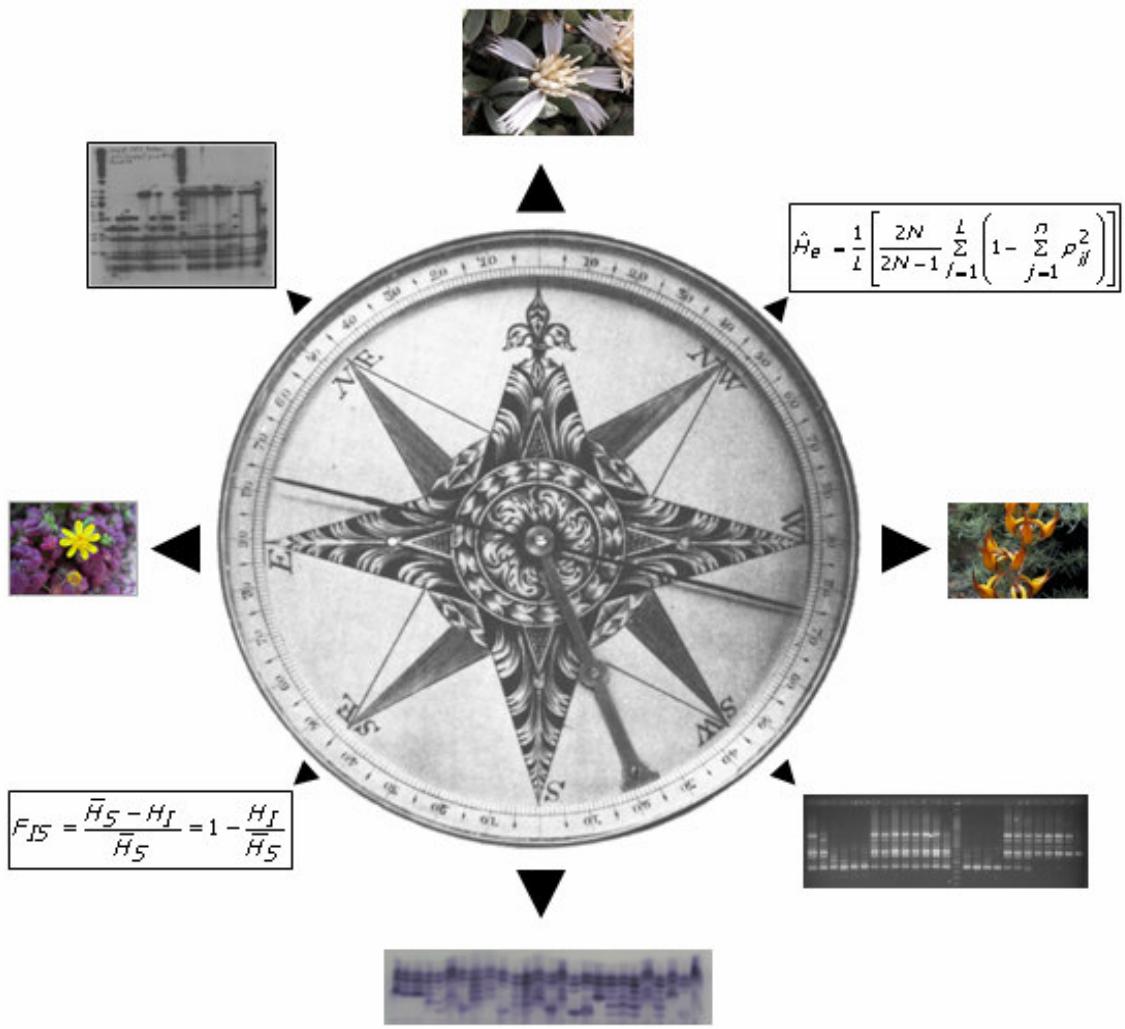


# BRÚJULA PARA BOTÁNICOS DESORIENTADOS EN LA GENÉTICA DE POBLACIONES

Juli Caujapé Castells



**EXEGEN**  
EDICIONES



---

## BRÚJULA PARA BOTÁNICOS DESORIENTADOS EN LA GENÉTICA DE POBLACIONES

---

© Juli Caujapé-Castells 2006

Departamento de Biodiversidad Molecular y Banco de ADN  
Jardín Botánico Canario "Viera y Clavijo"  
Ap. de correos 14 de Tafira Alta  
35017 Las Palmas de Gran Canaria (España)  
email: julicaujape@grancanaria.com

© EXEGEN ediciones  
<http://www.exegen.org>

La fotografía de la brújula fue descargada de internet a través de *Google images* en  
<http://images.google.com/imgres?imgurl=http://scholar.lib.vt.edu/theses/available/etd-05222003-124212/unrestricted/COMPASS.JPG>

Las Palmas de Gran Canaria (España)



## SOBRE LA DESORIENTACIÓN

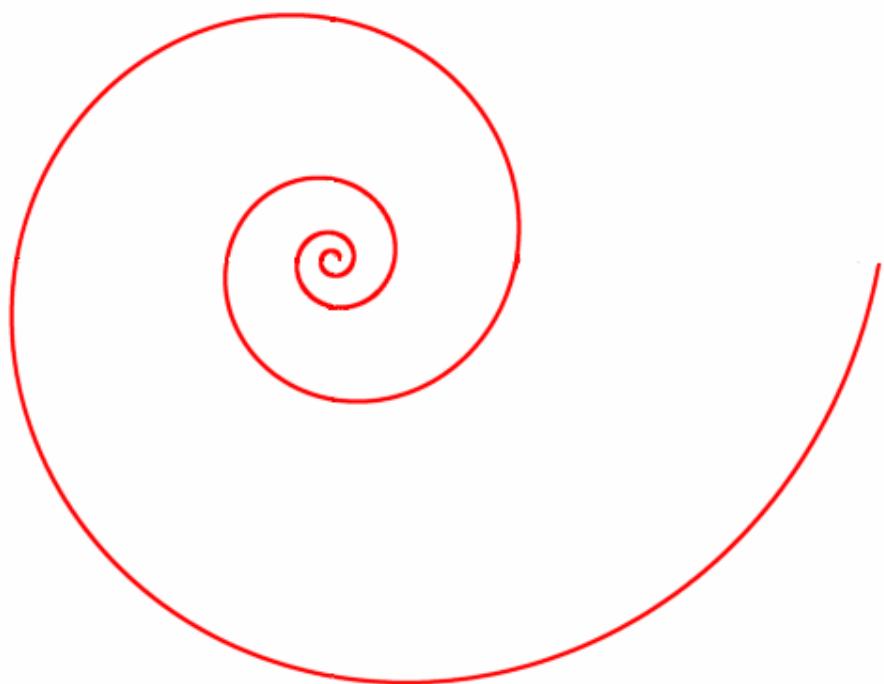
Creo que nadie se consideraría experto en fenómenos que cambiaron sin cesar, que pudieran desaparecer en cualquier momento (muchas veces sin motivo aparente) y cuya aparición fuera imposible de observar o predecir. Y eso es exactamente lo que pasa con los organismos vivos y con los procesos que los generan e incrementan o disminuyen su número. Cada una de las especies que conforman la manifestación visible de la biodiversidad es una singularidad, un fenómeno irrepetible que ha surgido una sola vez en la historia de la vida en nuestro planeta. Y el comportamiento de las singularidades no es predecible porque, como indica su nombre, no podemos observarlas un número de veces estadísticamente razonable. Existen muchas especies desconocidas porque se extinguieron antes de que la ciencia las descubriera, nadie ha observado nunca cómo ha surgido una especie de las que conocemos, y algunos de los métodos que se consideran más fiables para investigarlas cuantitativamente se basan en variables que son invisibles (por ejemplo, los marcadores moleculares). La impredicibilidad y nuestra ignorancia envuelven a la biodiversidad con un denso velo, y los intentos de la investigación para levantarla un poquito y ver lo que oculta pueden considerarse auténticas expediciones a la naturaleza de una realidad que se nos presenta necesariamente misteriosa.

No hay "expertos", y nuestra situación de partida es aproximadamente la siguiente: apenas tenemos una vaga idea del número de seres vivos que habitan el planeta con nosotros, no sabemos de cuántas maneras pueden producirse la especiación o la extinción, solamente estamos empezando a atisbar cómo interactúan los genomas de los diferentes seres vivos, y ninguna de nuestras maneras de representar la estructura de la diversificación biológica refleja la multidimensionalidad del problema. Y por si esta biodiversa miopía fuera poca, nuestras inferencias siempre parten de un muestreo muy limitado en el espacio, en el tiempo y en los genomas. Por todo esto y por más, no deja de ser sorprendente que nos atrevamos a elaborar teorías sobre el origen, la evolución y la estructura de la biodiversidad. Pero más sorprendente todavía es el hecho de que algunas teorías funcionen considerablemente bien, al menos durante un tiempo. La genética de poblaciones es una de ellas.

¿Cómo podemos llegar a saber algo partiendo de tan poco? Tal vez uno de los resultados más sólidos que ha producido la historia de la investigación biológica es que la biodiversidad solamente puede comprenderse a través de una diversidad de aproximaciones. Por esto, cuando describo la actividad del Departamento de Biodiversidad Molecular del Jardín Botánico Canario "Viera y Clavijo", siempre digo que utilizamos la información que contiene la molécula de ADN, pero también la multidisciplinariedad y la creatividad, para solucionar problemas que afectan al origen, diversificación y conservación de la Flora Canaria. Sin información molecular, los genéticos de poblaciones no tendríamos datos que discutir. Sin aproximaciones multidisciplinares (que implican muchos otros tipos de datos valiosos), no podríamos interpretar adecuadamente esa información y las conclusiones de muchas investigaciones moleculares serían pura deliciosa. Sin creatividad, no imaginariamos nuevas teorías que nos permitan saltar más allá de los datos y propongan otras vías de investigación molecular que las corroboren o refuten. La tríada información-multidisciplinariedad-creatividad da lugar a un círculo virtuoso, que no se cierra donde empezó, sino que se aleja a cada vuelta al menos un poco más allá de su punto de inicio. Y, gracias a esa permanente expansión, explora lo desconocido.

En realidad, todos estamos desorientados en la biodiversidad de la que somos parte, porque nadie es capaz de integrar todos los campos requeridos para comprenderla y gestionarla: genética, biología reproductiva, demografía, ecología... pero también (especialmente en lo que afecta a la conservación) economía, sociología y, cómo no, política. Aunque una de mis líneas de investigación molecular desde mis inicios en la Estació Internacional de Biología Mediterrània de Blanes (allá en 1990) ha sido la genética de poblaciones, hay muchas razones de peso para evitar considerarme un experto en un tema de por sí tan amplio y tan rápidamente expansivo. Por ello, esta brújula solamente puede ser una explicación muy sencilla de las dificultades más comunes de las que he sido testigo a través de la parte de mi trabajo que implica interactuar con botánicos y botánicas que usan las herramientas de la genética de poblaciones para sus investigaciones. Y ya que no puede ser un tratado, una de mis mayores esperanzas es que contribuya a desarrollar la multi-disciplinariedad y la creatividad en alguna de las personas que la usen.





## ÍNDICE

Sobre la desorientación	3
Prefacio	6
1. Antes de estimar la variabilidad genética poblacional	9
1.1. Antes de ir a la población	11
Introducción	10
¿Cuándo ir a la población?	11
¿Qué llevar?	12
1.2. En la población	17
Fase I. Reconocimiento espacial de la población	17
Fase II. Diseño del muestreo	18
¿Cuántas poblaciones? ¿Cuántos individuos?	19
Fase III. Recolección de muestras	22
Normas prácticas de muestreo de aplicación general (DNA e isoenzimas)	23
Normas prácticas de muestreo de aplicación específica para isoenzimas	24
Normas prácticas de muestreo de aplicación específica para DNA	24
1.3. Despues de ir a la población	25
2. Sobre los parámetros y tests más usados en genética de poblaciones	31
Introducción	32
2.1. Conceptos básicos	34
El teorema de Hardy-Weinberg	35
La deriva genética	39
El efecto fundador	43
El efecto cuello de botella	45
La paradoja del tamaño poblacional	46
El efecto Wahlund	48
El tamaño poblacional efectivo	50
El modelo de estructura poblacional en isla	53
El modelo de estructura poblacional en pasadera	54
2.2. Cálculos y tests más usados en genética de poblaciones	56
2.2.1. Descriptores de la variabilidad poblacional	57
Número de alelos por locus	57
Número medio de alelos por locus	57
Número efectivo de alelos por locus	57
La diversidad genotípica	58
La probabilidad de pérdida alélica	58
La heterozigosidad observada	59
La heterozigosidad esperada	60
El índice de fijación	61
La tasa de reproducción cruzada	62
La consanguinidad	63
La depresión por consanguinidad	66
El tamaño de la mínima población viable	70
El desequilibrio en el ligamiento	71
El test de cuello de botella	73
El test de neutralidad de Ewens-Watterson	75
La autocorrelación espacial	76
El test de Mantel	78
2.2.2. Descriptores de la variabilidad genética entre las poblaciones	80
Los estadísticos de estructura poblacional de Wright	81
Los estadísticos de estructura poblacional de Nei	86
La identidad de Nei y la distancia de Nei	89
Estimación del intercambio genético entre poblaciones	91
El número medio de migrantes por generación	92
Programas más usados para la estimación de la diversidad genética poblacional	94
3. Una aplicación para maximizar el muestreo de variabilidad genética en colecciones <i>ex situ</i>	95
4. Una aplicación para resolver casos taxonómicos conflictivos en ausencia de caracteres morfológicos discriminantes	114
BIBLIOGRAFÍA	123



## PREFACIO



Szent Györgyi

El científico húngaro Albert Szent-Györgyi Nagyrapolt, premio Nobel de Fisiología y Medicina en 1937 por sus descubrimientos sobre los procesos de combustión biológica y el ciclo del ácido fumárico, contribuyó también al mundo del pensamiento al escribir que “el descubrimiento científico consiste en ver lo que todos han visto pensando lo que nadie ha pensado” (ver Good 1962).

Pero la tecnología ha evolucionado mucho desde la época del eminent fisiólogo húngaro y, hoy en día más que entonces, también es cierto que los científicos podemos ver cosas que nadie ve, tal vez pensando lo que mucha gente pensaría si tuviera acceso a esa información. Los marcadores moleculares están dentro de ese conjunto de “visiones exclusivas”.

Puesto que la actividad científica consiste esencialmente en explicar las cosas, muchas veces me he propuesto encontrar las palabras adecuadas para describirle brevemente a alguien sin nociones de biología (por ejemplo, un niño) de qué trata la parte molecular de mi trabajo. A pesar de mi obstinación para conseguir una definición sencilla y a la vez rigurosa, lo mejor que me ha salido suena algo así como

“Estimado niño, investigo fenómenos que han ocurrido una sola vez en la historia de la vida en la tierra usando la información que reside en elementos invisibles que contienen los organismos”

Esotérica frase que parece transportarnos a épocas legendarias en que hechiceros y druidas eran los únicos que podían atravesar la oscuridad del mundo terrenal mediante su “mágica” comprensión de señales esquivas para el resto de mortales. No obstante, a pesar de lo que pueda parecer, los marcadores moleculares no son una piedra filosofal, sino una realidad palpable que, con sus defectos y limitaciones, puede ofrecer aplicaciones tan poco esotéricas como ayudar a comprender la realidad ya diseñar estrategias para la conservación de la biodiversidad.

Si quedan aún personas (incluidos algunos biólogos) a quienes les puede parecer que la actividad molecular tiene un punto de hechicería, mucha de la culpa puede ser nuestra, de los que nos dedicamos a la interpretación de la naturaleza utilizando información molecular. En una parte (afortunadamente cada vez más pequeña), porque se fomenta todavía cierta corriente de pensamiento que defiende que lo que tiene trascendencia científica debe parecer también difícil de entender. Para los científicos que adoptan esta actitud críptica (frontalmente incompatible con su trabajo de *explicar* las cosas), hacer que su quehacer parezca incomprensible tiene claras ventajas: a costa de asustar sin motivo, se aseguran un estatus privilegiado. Son druidas disfrazados.



Druida

Otra parte de culpa la tiene sin duda la necesidad imperiosa de reciclarse continuamente, publicar en revistas especializadas de impacto internacional, asistir a congresos, buscar financiación para proyectos, dirigir tesis y otras muchas actividades relacionadas. Estos quehaceres, imprescindibles e inexcusables para cualquiera que se llame científico hoy en día, exigen el conocimiento de lenguas extranjeras, jergas especializadas y técnicas complejas cuyo estudio, dominio y aplicación quita tiempo y ganas para dedicarse, entre otras cosas, a la divulgación. Pero algunos de los investigadores que adquieren cierta reputación por los éxitos científicos obtenidos en su campo de conocimiento utilizan a veces su *curriculum vitae* como un sofá: como no pueden dejar de publicar artículos de impacto, se procuran una cómoda excusa para no tener que divulgar. Son vagos prestigiosos.



Vagos en un sofá



Seguramente haya muchas más razones para evitar dar explicaciones no especializadas sobre nuestro trabajo, pero lo cierto es que todos los científicos moleculares que conozco tienen mucho más de "vago" que de "druida". Cualquiera de ellos sería capaz de (aunque quizás por lo visto debería estar obligado a) explicar con palabras sencillas y rigurosas las bases de su actividad, sus aplicaciones y sus logros. El desarrollo de nuestras capacidades divulgativas es una forma de dejar de alejarnos de la sociedad a la que pertenecemos y, a la vez, conseguir transmitirle la importancia de lo que hacemos. Al fin y al cabo, como dice Ramón Folch (1998), la ciencia surgió para explicar las cosas, y no para convertirse en una especie de esoterismo de precisión.

Aunque esta brújula no sea una obra de divulgación *sensu stricto*, y seguramente muchos de los conceptos que en ella se tratan no estén explicados de la mejor manera posible, sí es un intento de librarme de parte de mi vagancia social para orientar a botánicos interesados en genética de poblaciones en algunos de sus aspectos básicos, desde el muestreo genético y la estimación de parámetros hasta algunas aplicaciones en la conservación *ex situ* y la taxonomía. No significa mucho, pero es ése primer paso que hace más fácil dar los siguientes. Puesto que mi trabajo en los últimos ocho años se ha desarrollado en Canarias, esta obra se ha inspirado en mi interacción con la espectacular biodiversidad de este archipiélago y con (sobre todo) el personal científico y técnico del Jardín Botánico Canario "Viera y Clavijo" (JBCVC). Por ello, la mayoría de ejemplos de aplicación han sido elaborados pensando "en Canario".

La redacción de una brújula de este tipo es muy diferente a la de un artículo científico, ya que por definición la brújula va dirigida a botánicos con formación molecular inexistente o incipiente, y debe explicar aspectos que se dan por sentados en foros especializados. Por ello, su intención de partida fue (y sigue siendo) poner el estándar de conocimientos lo más bajo posible, para que las explicaciones sean accesibles a casi cualquier persona con formación universitaria en botánica e interés en la aplicación de técnicas de genética de poblaciones.

Como tantas cosas en la vida, el origen de esta brújula se debe a una singularidad. En Abril de 2003 fui convocado junto con otros colegas a una reunión en la que se discutió sobre el uso de las técnicas moleculares para la caracterización de organismos vivos, y la persona que convocó esa reunión me instó a que elaborara unas indicaciones de muestreo de poblaciones de plantas vasculares para la estimación de su variabilidad genética con técnicas moleculares. Debía estar listo en un mes. Por motivos ajenos a mi persona, aquellas indicaciones nunca fueron requeridas, y el resultado de mi injustificada prisa se parece mucho a la primera parte de este texto, que da unas orientaciones para muestrear poblaciones de plantas vasculares en ausencia de información sobre su variabilidad genética. Vaya pues mi más sincero agradecimiento a la persona que me propuso el reto de gestar las ideas que contiene esta primera parte ya que, sin su estímulo, es muy posible que este texto nunca hubiera nacido.

Como casi todos los partos prematuros del siglo XXI, aquel pequeño texto pudo crecer y desarrollarse, en gran medida gracias a las investigaciones que el Departamento de Biodiversidad Molecular del JBCVC desarrolla en colaboración con los diferentes departamentos del centro. Así, la segunda parte de la brújula, dedicada a la explicación de los parámetros y tests más utilizados en genética de poblaciones vegetales, está sobre todo motivada por mi interacción con el equipo del departamento de biología reproductiva y micromorfología del JBCVC, dirigido por las Dras. Julia Pérez de Paz y Rosa Febles, gracias a quienes inicié mi actividad en el JBCVC en 1999 colaborando desde la perspectiva molecular en varios de sus proyectos multidisciplinares de investigación. No obstante, a lo largo de este tiempo, también han hecho aportaciones a esta parte de la brújula muchos/as otros/as biólogos/as con los que he tenido la fortuna de poder interaccionar.

La tercera parte trata de una aplicación específica para el muestreo de variabilidad genética para conservación *ex situ* en bancos de germoplasma, y hubiera sido imposible sin la colaboración e interés de la responsable del Banco de Semillas del JBCVC (Alicia Roca) y su grupo, cuya misión (la conservación de la diversidad genética de las poblaciones naturales de plantas Canarias fuera de sus áreas de distribución) entraña perfectamente con una de las posibles aplicaciones prácticas de los marcadores moleculares. En esta tercera parte se propone un método para diseñar estrategias de



conservación *ex situ* de la biodiversidad vegetal, asumiendo un muestreo que siga las directrices expuestas en la primera parte.

Finalmente, la cuarta parte de la brújula está motivada por mis incursiones en la complejidad taxonómica de la biodiversidad vegetal Canaria, que he aprendido a conocer a través de mi interacción con los/as investigadores/as antes mencionados/as, pero sobre todo con Águedo Marrero (responsable del herbario del JBCVC e inmenso conocedor de las floras Macaronésicas), David Bramwell (impulsor de la biogeografía Canaria y actual director del Jardín Botánico Canario) y José Naranjo y Bernardo Navarro (artífices del Atlas de Flora Amenazada de Gran Canaria). En esta parte se propone una aplicación de las técnicas moleculares poblacionales para resolver dudas taxonómicas cuando los caracteres morfológicos no son suficientemente resolutivos, lo cual sucede con frecuencia, y no solo en Canarias.

Como todos los seres vivos, este texto está sujeto a variaciones en el tiempo, que intentaré que solamente contribuyan a mejorarlo. Sin duda, muchas de estas modificaciones tendrán su origen en las investigaciones moleculares que estamos llevando a cabo actualmente en el Jardín Botánico Canario "Viera y Clavijo". Pero también es probable que surjan nuevas técnicas y metodologías de análisis estadístico que superen, o incluso ridiculicen, muchos de los apartados actuales. Por el momento, me conformaría con que esta brújula cumpla su objetivo y oriente a los intrépidos botánicos que en ocasiones se pierden en la selva de la genética de poblaciones, cada vez más densa.

Juli Caujapé Castells  
Las Palmas de Gran Canaria  
Marzo de 2004



## 1. ANTES DE ESTIMAR LA VARIABILIDAD GENÉTICA POBLACIONAL



## INTRODUCCIÓN

La estructura genética es la huella heredable que la interacción continuada entre factores estocásticos (asociados a la contingencia histórica) y biológicos imprime en las poblaciones de organismos. A partir de esta definición, es obvio que para una correcta estimación de la estructura genética de las poblaciones naturales de plantas hemos de partir de un muestreo representativo; de otra forma, nos arriesgamos a que las inferencias que desarrollemos ofrezcan una estimación sesgada de la diversificación poblacional. El muestreo es una parte crucial de cualquier investigación sobre diversificación poblacional, de manera que la calidad de nuestro muestreo determinará el grado de bondad de nuestras estimaciones.

A menos que nuestros organismos de trabajo sean virus o bacterias, el muestreo para la estimación de la variabilidad genética parte siempre de la importante limitación de no tener acceso a una misma población en diferentes momentos de su tiempo de existencia. Pero, incluso situándonos en el presente, seríamos afortunados si esta fuera nuestra única fuente de incertidumbre. Unas veces desconocemos totalmente los factores históricos que han influido en la variación genética de una especie, otras veces no tenemos información sobre sus atributos biológicos (ploidía, sistema reproductivo, relaciones filogenéticas), frecuentemente nuestras estimaciones del tamaño poblacional son erróneas (por la extensión de la población, la incidencia de la reproducción clonal o la existencia de bancos de semillas en el subsuelo) y, en muchas ocasiones, ni siquiera conocemos bien el área de distribución actual de las especies que estamos investigando.

Aunque después de leer este párrafo nos sintamos como náufragos ciegos, el objetivo de estimar la variación genética no significa navegar a la deriva; simplemente nos debe obligar a ser conscientes de las muchas fuentes de error que amenazan a nuestras estimaciones y a trabajar en la peor de las hipótesis: que la variabilidad genética se puede hallar en el sitio más inesperado. Solamente concibiendo así el muestreo podemos estar mínimamente seguros de poner las bases adecuadas para estimar adecuadamente la variación genética de las poblaciones de interés.

Como es fácil imaginar, todo lo anterior significa también que es muy difícil dar indicaciones de validez universal para muestrear poblaciones naturales con el fin de cuantificar su variabilidad genética a través de técnicas moleculares. Las poblaciones de plantas vasculares (especialmente las Canarias!) son inherentemente variables y, además, existen muy pocas especies o familias para los que los conocimientos moleculares podrían darnos ciertas pistas para el muestreo. No obstante, sí es posible ofrecer algunas directrices para intentar muestrear los individuos de manera que las estimaciones de diversidad genética resultantes sean fiables.

Esta primera parte de la brújula trata de ofrecer algunas generalizaciones para conseguir representar la variabilidad genética de las poblaciones. Seguir estas directrices es importante para aplicar correctamente la metodología de conservación genética desarrollada en la segunda parte.



## 1.1. ANTES DE IR A LA POBLACIÓN

Es conveniente coordinarse con el laboratorio de análisis molecular para saber qué tipo de técnica va a emplearse con las plantas a muestrear; el tipo de técnica molecular condiciona el material que hay que llevar al campo, la conservación de las muestras y la rapidez de su transporte al laboratorio de análisis. Se debe consultar con la persona responsable del análisis genético cualquier duda o cuestión que se tenga sobre el muestreo.

Es asimismo conveniente recopilar información sobre la especie diana (especialmente si está publicada en revistas de impacto o libros) porque nos será de gran ayuda para la interpretación de los datos genéticos que se obtengan y para su publicación en revistas de impacto. Disponer de un mapa de distribución de la especie a muestrear (aunque sea preliminar) ayuda a planificar la ruta de recogida de material.

### Cuándo ir a la población?

Aunque hay diferencias dependiendo de si las muestras van a analizarse mediante isoenzimas o mediante técnicas de DNA, es siempre preferible visitar las poblaciones diana cuando la planta se halla en su pico productivo previo a la floración. Como fechas orientativas generales para Canarias, se puede decir que el muestreo podría empezar en Octubre-Noviembre después de las primeras lluvias y finalizar en Febrero-Marzo.

### Isoenzimas

**Muestrear antes de la floración es imprescindible**, ya que es en esta fase cuando las hojas son jóvenes y verdes, la planta se halla en su pico productivo y los niveles de los enzimas del metabolismo basal se hallan a la máxima concentración y, por tanto, se visualizan mejor en los geles.

Si el muestreo se realiza durante la época de floración o fructificación, la resolución de la mayoría de los isoenzimas puede verse muy atenuada o incluso suprimida, con lo cual se hace imposible la caracterización genética de las poblaciones mediante esta técnica. De manera general, esto es así porque cuando las plantas florecen dedican todos los recursos a la reproducción, y muchos de los enzimas del metabolismo basal que se producen en abundancia antes de este estadio fenológico dejan de producirse (o se producen en mucha menor cantidad) a partir de su inicio.

### DNA

**Es aconsejable realizar los muestreos para DNA en el estadio fenológico previo a la floración.** Aunque hay que admitir que la época de muestreo no es de importancia tan crítica en DNA como en isoenzimas, se da el caso de que algunos endemismos canarios producen metabolitos secundarios que luego acumulan en las hojas y otras partes vegetativas. Para los taxones con estas características, retrasar el muestreo de hojas hasta la época de floración o la fructificación de las plantas suele ir en menoscabo del rendimiento de la extracción de DNA, puesto que los mencionados metabolitos interfieren en el proceso de extracción o incluso se complejan con el DNA, dificultando los posteriores análisis.

**Siempre que sea posible, muestrear hojas adultas.** Según la experiencia del Banco de ADN de Kew (Mark Chase y Michael Fay, comunicación personal) el ADN de las hojas jóvenes de algunos grupos de plantas es imposible de amplificar debido a la presencia de compuestos químicos que no se hallan en otros tejidos. Estos compuestos probablemente impidan la predación por parte de insectos o animales antes de que las hojas se hayan endurecido.



## Qué llevar?

Ahorra mucho tiempo y esfuerzo tener todo el material necesario para el muestreo preparado (Tabla 1) desde el día anterior. Aparte de esto, es también aconsejable:

1. Ordenar correlativamente las bolsas que contendrán las muestras de hoja por el código correspondiente (que debe estar ya dentro de cada bolsa en sitio visible desde el exterior) y disponerlas en un recipiente que facilite su extracción individual en orden. A estos efectos, puede servir un chaleco o un bolso de cintura (como los que utilizan algunas personas para llevar el dinero) que tenga al menos dos compartimentos separados: uno para las bolsas vacías y otro para las bolsas con muestras.
2. La mejor alternativa de las que hemos ideado en el JBCVC hasta el momento para preparar las bolsas de muestreo es la que se ejemplifica en las Figuras 0a y 0b. Las bolsas (de plástico transparente, con cierre hermético) se preparan con antelación introduciendo las etiquetas de papel con los códigos individuales en el fondo de la bolsa, que luego se sella por calor. Posteriormente, con la ayuda de un troquelador adecuado, se pueden hacer unos agujeros flanqueando el código que nos pueden servir para almacenar las bolsas en orden en una libreta del tipo Multifin. Esta libreta puede llevarse al campo y, después del muestreo, introducirse directamente en el ultracongelador (en caso de que la muestra sea para isoenzimas), en el congelador de  $-20^{\circ}\text{C}$  o almacenarse a temperatura ambiente, si esta opción es la adecuada. Una ventaja de este método es que las hojas que se introduzcan en la bolsa nunca nos impedirán ver el código, al estar sellado el compartimento que lo contiene. Otra ventaja de este sistema es que la tinta con la que está escrito el código de cada planta no va a deteriorarse, ya que nunca entrará en contacto con el hielo o la escarcha que se forma en el interior del ultracongelador o la nevera. Este sistema de almacenamiento se ideó en el Jardín Botánico Canario "Viera y Clavijo" (JBCVC) gracias a la colaboración de Eugenio Reyes y Juan Luis Sánchez con el autor de esta brújula.

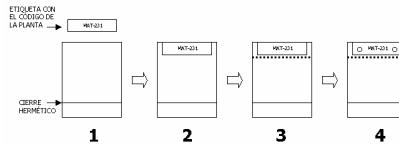


Figura 0a. Las cuatro fases para la elaboración de las bolsas de muestreo. (1) tener a mano las etiquetas con los códigos y las bolsas de plástico con cierre hermético en uno de sus extremos; (2) introducir la etiqueta en el fondo de la bolsa; (3) sellar al calor el espacio que contiene la etiqueta; y (4) troquelar dos agujeros a la separación adecuada para ordenar las bolsas en una libreta de anillas tipo Multifin (ver Figura 0b).

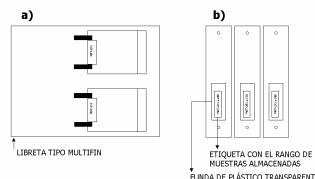


Figura 0b. Almacenamiento de las bolsas de muestreo ordenadas correlativamente en una libreta de anillas tipo Multifin. a) Libreta de anillas abierta. Una libreta mediana de 4 anillas puede almacenar dos filas de 25 bolsas conteniendo hojas; b) esquema de las libretas Multifin ordenadas correlativamente según el rango de muestras que contienen. Las libretas pueden ser almacenadas vertical u horizontalmente en el ultracongelador de  $-80^{\circ}\text{C}$ , congelador, frigorífico o a TA, facilitando considerablemente la localización de un individuo cualquiera en un tiempo mínimo.



3. En caso de muestrear para técnicas de DNA, poner el gel de sílice necesario dentro de las bolsas antes del muestreo. También puede verterse dentro de las bolsas según vayan utilizándose, dependiendo de lo que la(s) persona(s) que realice(n) el muestreo considere(n) más cómodo.
4. Es conveniente confeccionar una libreta con hojas de muestreo cuyo anverso y reverso sea como se muestra en las Figuras 1 y 2, respectivamente. Esta libreta facilita mucho la recopilación sistemática de información y evita tener que ir por ahí con hojas sueltas, que son fácilmente extraviadas.

Tabla 1. Material necesario y aconsejable para el muestreo de hojas de plantas vasculares para análisis con técnicas moleculares. Cuando el material es específico para isoenzimas o DNA, se simboliza mediante [ISO] o [DNA], respectivamente.

MATERIAL NECESARIO	UTILIDAD
<input type="checkbox"/> Permiso de muestreo (en caso de necesitarse)	Certificación de la autorización legal de recolecta
<input type="checkbox"/> DNI	Confirmación del permiso de muestreo
<input type="checkbox"/> Brújula	Orientación de la población
<input type="checkbox"/> GPS	Ubicación de la población
<input type="checkbox"/> Altímetro	Medida de la altitud de la población
<input type="checkbox"/> Mapa detallado de la zona a muestrear	Localización de la población
<input type="checkbox"/> Prismáticos	Supervisión de los alrededores, delimitación
<input type="checkbox"/> Tijeras pequeñas	Corte de hojas o ramas (en caso necesario)
<input type="checkbox"/> Cinta métrica	Medida de las distancias para mapa (Fig. 2)
<input type="checkbox"/> Claves de determinación (si se tienen)	Resolución de posibles dudas taxonómicas
<input type="checkbox"/> Nevera de campo [ISO]	Almacenamiento de muestras
<input type="checkbox"/> 3 ó 4 bloques refrigerantes (en la nevera) [ISO]	Enfriamiento y conservación de muestras
<input type="checkbox"/> Gel de sílice [DNA]	Secado de muestras para evitar su deterioro
<input type="checkbox"/> Libreta de muestreo (ver Figs. 1 y 2)	Sistematización de los datos poblacionales
<input type="checkbox"/> Bolsas de plástico con cierre hermético <sup>1</sup>	Almacenamiento de las hojas muestreadas
<input type="checkbox"/> Bolso conteniendo las bolsas en orden	Agilización de las recolecciones
<input type="checkbox"/> Etiquetas de papel para pliegos (ver Fig. 3)	Designación de los pliegos testigo
<input type="checkbox"/> Bolsas herméticas sin marcar	Recolección de posibles muestras adicionales
<input type="checkbox"/> Prensa de campo con periódicos viejos <sup>2</sup>	Recolección de pliegos testigo
<input type="checkbox"/> Bolígrafo	Anotaciones
MATERIAL ACONSEJABLE	
<input type="checkbox"/> Teléfono móvil o "walkie" (1 por persona)	
<input type="checkbox"/> Libreta de campo personal	
<input type="checkbox"/> Botella o cantimplora con agua	
<input type="checkbox"/> Algo de comida	
<input type="checkbox"/> Chaleco y chubasquero	
<input type="checkbox"/> Navaja de campo	
<input type="checkbox"/> Botas	
<input type="checkbox"/> Jersey	
<input type="checkbox"/> Gorra o sombrero	
<input type="checkbox"/> Crema protectora solar	
<input type="checkbox"/> Algo de dinero	

1: en orden correlativo para agilizar la colecta

2: Si se prefiere por comodidad, pueden llevarse bolsas para introducir el material recolectado para este fin, y hacer los pliegos al llegar al edificio de trabajo.



FECHA:	CÓDIGO:
ESPECIE:	
POBLACIÓN:	
ALTITUD: UTM:	
PENDIENTE (APROX.):	
TIPO DE TERRENO:	
ORIENTACIÓN DE LAS PLANTAS:	
RECOLECTOR/A:	
 <b>Localización aproximada</b>	
<b>BREVE DESCRIPCIÓN DE LA POBLACIÓN (plantas acompañantes, etc...)</b> <div style="border: 1px solid black; height: 100px; width: 100%;"></div>	
<b>TAMAÑO POBLACIONAL ESTIMADO</b> <input type="radio"/> < 100 <input type="radio"/> 100 a 500 <input type="radio"/> 500 a 1.000 <input type="radio"/> 1.000 a 5.000 <input type="radio"/> > 5.000 <input type="radio"/> > 10.000	
<b>INDIVIDUOS EN FLOR</b> <input type="radio"/> Ninguno <input type="radio"/> 10%-50% <input type="radio"/> 50%-80% <input type="radio"/> @100%	
<b>INDIVIDUOS CON FRUTO</b> <input type="radio"/> Ninguno <input type="radio"/> 10%-50% <input type="radio"/> 50%-80% <input type="radio"/> @100%	
<b>DISTRIBUCIÓN GENERAL DE LOS INDIVIDUOS</b> <input type="radio"/> Continua con núcleos físicos <b>Comentarios adicionales sobre la distribución?</b> <div style="border: 1px solid black; height: 100px; width: 100%;"></div> <input type="radio"/> Continua sin núcleos físicos	
<input type="radio"/> Discontinua en núcleos	
<input type="radio"/> Discontinua sin núcleos	
<input type="radio"/> Sin patrón definido	
<b>HAY HETEROGENEIDAD AMBIENTAL VISIBLE?</b> <input type="radio"/> SI <input type="radio"/> NO	
<div style="border: 1px solid black; height: 100px; width: 100%;"></div>	
<b>NÚMERO DE MUESTRAS RECOLECTADAS PARA ANÁLISIS</b> <input type="text"/>	
<b>OBSERVACIONES:</b> <div style="border: 1px solid black; height: 100px; width: 100%;"></div>	

Figura 1. Anverso de la hoja de muestreo. Todos los apartados de esta hoja, excepto el número de muestras recolectadas para análisis, han de rellenarse al final de la fase I del proceso de muestreo (ver Fig. 4), correspondiente al reconocimiento de la población. Aunque el ejemplo es para Gran Canaria, el recuadro superior derecho está pensado para insertar la silueta de la isla o región correspondiente.



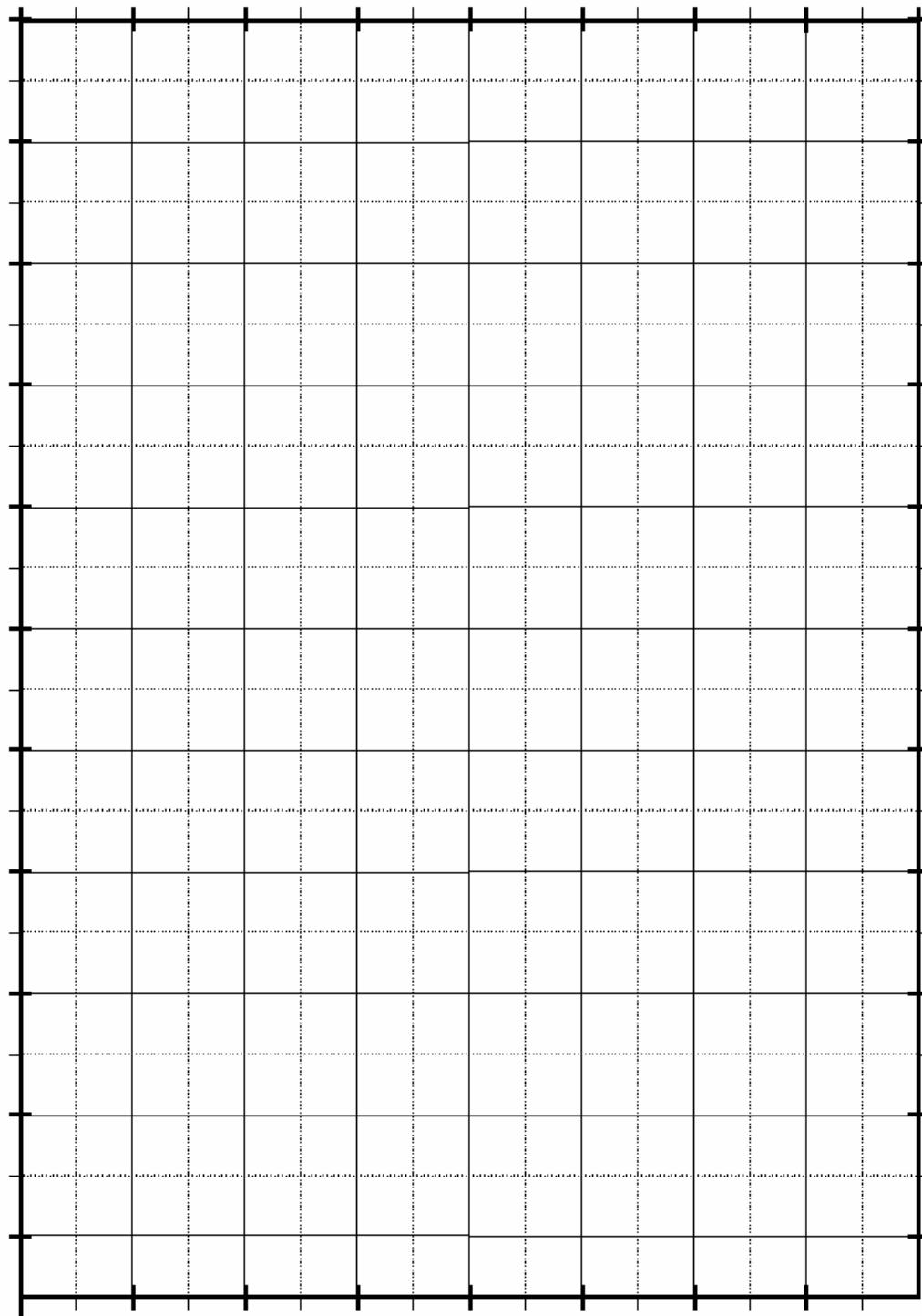


Figura 2. Reverso de la hoja de muestreo. Este entramado es ideal para utilizarlo para hacer el mapa del muestreo. Las subdivisiones se han dejado sin unidades para que el responsable del muestreo las seleccione en función de la extensión ocupada por la población. El dibujo no ha de ser exacto, pero sí tiene que mostrar de manera fidedigna la posición relativa de las plantas muestreadas.



ESPECIE:	ESPECIE:
FECHA:	FECHA:
LOCALIDAD:	LOCALIDAD:
ALTITUD:	UTM:
PENDIENTE (APROX.):	PENDIENTE (APROX.):
TIPO DE TERRENO:	TIPO DE TERRENO:
RECOLECTOR/A:	RECOLECTOR/A:

ESPECIE:	ESPECIE:
FECHA:	FECHA:
LOCALIDAD:	LOCALIDAD:
ALTITUD:	UTM:
PENDIENTE (APROX.):	PENDIENTE (APROX.):
TIPO DE TERRENO:	TIPO DE TERRENO:
RECOLECTOR/A:	RECOLECTOR/A:

ESPECIE:	ESPECIE:
FECHA:	FECHA:
LOCALIDAD:	LOCALIDAD:
ALTITUD:	UTM:
PENDIENTE (APROX.):	PENDIENTE (APROX.):
TIPO DE TERRENO:	TIPO DE TERRENO:
RECOLECTOR/A:	RECOLECTOR/A:

Figura 3. Hoja de etiquetas para pliegos de herbario. Estas etiquetas facilitan la sistematización de la información correspondiente a las poblaciones visitadas. Pueden llevarse en hojas separadas troquelables para incluirlas en el pliego correspondiente.



## 1.2. EN LA POBLACIÓN

Una vez puesto a punto todo lo anterior, ya podemos desplazarnos a la población y ejecutar el muestreo. Dependiendo del número de efectivos estimados en la población diana y de su distribución espacial, tal y como se ejemplifica en el organigrama de la Figura 4.

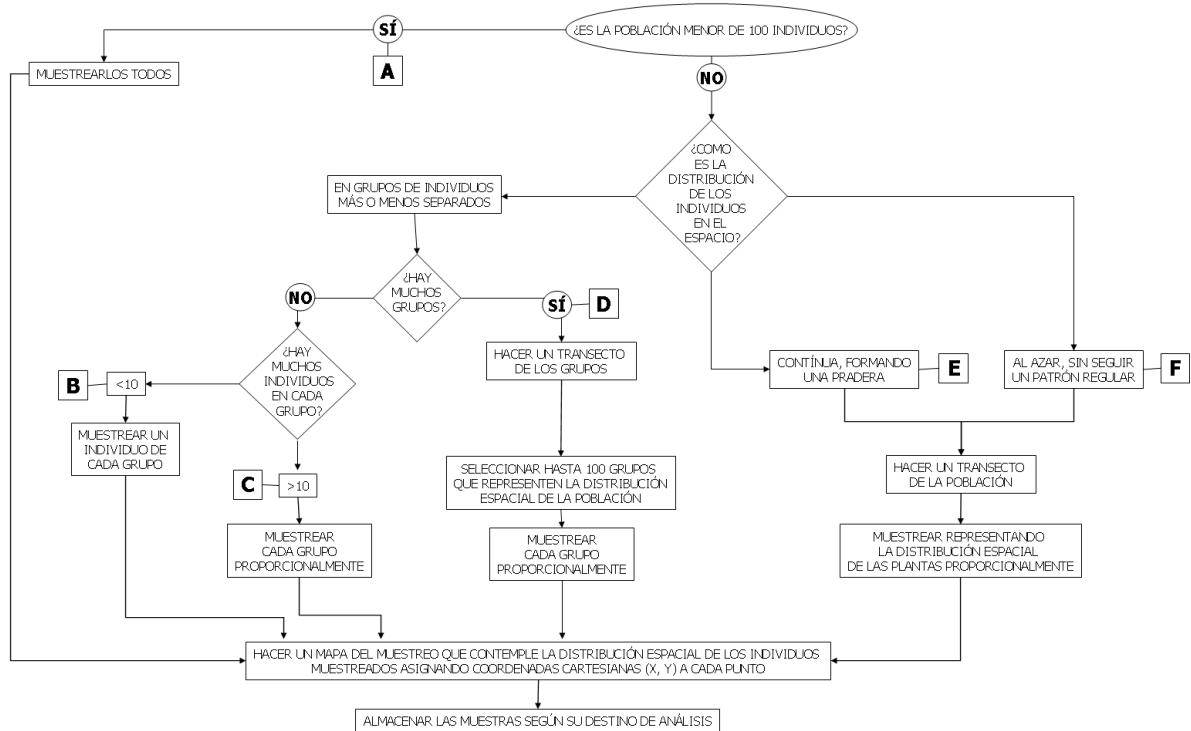


Figura 4. Organigrama de muestreo según seis posibles tipos poblacionales (codificados de A a F e inscritos en cuadrados) según el tamaño poblacional y a la distribución de los individuos en las poblaciones. Los tipos poblacionales denominados B, C, D, E y F se ilustran en las Figs. 8 a 12.

### Fase I: Reconocimiento espacial de la población

Llegados a la población diana, lo primero que hemos de hacer es reconocerla para estimar su extensión y el número de individuos que la componen, tomando todo el tiempo que se considere necesario en estas apreciaciones. Hay que tener en cuenta que en muchas ocasiones se dan núcleos de individuos (alejados del núcleo principal) que pueden contener variación genética diferente (especialmente si existe heterogeneidad ambiental). Por otro lado, a veces las poblaciones son muy extensas y solamente nos restringimos a una pequeña área de acción por "pereza", lo cual puede ir en detrimento de la estimación de los niveles de variación genética (ver las Figuras 4 y 5). Por estos motivos, esta es quizás la fase más importante, puesto que condicionará la bondad de la caracterización genética de la población según los marcadores moleculares seleccionados. Hay que hacer hincapié en que, para las finalidades del trabajo que nos ocupa, debemos representar bien toda la extensión de la población con objeto de estimar de manera consistente su variabilidad genética.

Durante esta fase conviene llenar todos los apartados de la hoja de muestreo (ver Figura 1) excepto, como es lógico, el número de muestras recolectadas, que sólo se podrá llenar al final del muestreo.



Debemos determinar si hay menos de cien individuos visibles (en cuyo caso se muestrearán todos) o si la población consta de un número mayor de efectivos (ver apartado 2.2.1). En este último caso, deberemos identificar a grandes rasgos la distribución de los individuos en el espacio poblacional (ver los ejemplos gráficos en las Figs. 8 a 12) para tener una idea del muestreo a realizar en la Fase III.

**Realizaremos siempre un pliego testigo recogiendo varias plantas representativas de los diferentes morfologías y tamaños que se hallan en la población e indicando todos los datos que se requieren en la etiqueta** (Fig. 3). Si las plantas no estuvieran en flor (este será el caso más frecuente si muestreamos para aloenzimas), sería conveniente regresar al lugar en la época adecuada para recoger plantas con flor y fruto, especialmente si las poblaciones muestreadas plantean dudas a nivel de reconocimiento morfológico.

### **Fase II: Diseño del muestreo**

Cuando la población sea mayor de 100 individuos, el muestreo que realizaremos obedecerá a un modelo estratificado. En esta técnica de muestreo aleatorio, la población se divide primero en grupos mutuamente exclusivos y posteriormente se seleccionan individuos al azar dentro de cada grupo. En nuestro caso, los criterios para delimitar los grupos estarán basados, en primer término, en la distribución espacial de los individuos en la población según las Figuras 4, 9, 11 y 12 (en grupos, continua, o sin seguir un patrón determinado). En segundo término, nuestro criterio de delimitación se basará en la densidad relativa de los individuos dentro de cada grupo (en líneas generales, muestrearemos más en los grupos que contengan más individuos).

No hay que olvidar que nuestro muestreo estratificado estará orientado a maximizar la probabilidad de capturar una porción significativa de la variación genética de la población. Por tanto, intentaremos abarcar toda el área de la población, para lo cual es conveniente trazar uno o varios transectos para recoger las muestras. En las Figuras 10, 11 y 12 se ejemplifican algunos transectos posibles para las poblaciones de tipo D, E y F. Si las poblaciones ocupan una franja amplia de terreno (como se quiere ejemplificar en estas figuras), se aconseja realizar transectos en cruz como los dibujados, muestreando predominantemente a lo largo de ellos, pero también incorporando en la muestra algunos de los individuos que queden fuera (ver Figs. 10, 11 y 12). En caso de que la población siga una tendencia de dirección en el espacio (por ejemplo a lo largo de una carretera, del cauce de un barranco o de los márgenes de un río) se muestreará a intervalos regulares a lo largo de ella, representando toda su extensión visible. Si la población es muy extensa, es conveniente muestrear a intervalos regulares (de tamaño proporcional a la extensión de cada población).

Hay que admitir que raramente encontraremos en el campo situaciones exactamente iguales a las ejemplificadas mediante los esquemas. No obstante, sí es muy probable que podamos interpretar la conformación espacial de la población con su ayuda.



## ¿Cuántas poblaciones?, ¿Cuántos individuos?

Los niveles de variación genética de las poblaciones naturales vienen condicionados por muchos factores extrínsecos que además presentan un impacto diferente dependiendo de su interacción con (entre otros factores intrínsecos) la historia evolutiva, la biología reproductiva y la antigüedad del linaje de los organismos, y no existen unas normas estándar para saber el número de muestras idóneo para cada tipo de organismo. El poder diversificador de la naturaleza se manifiesta con gran fuerza también a este nivel y es difícil de traducir en unas indicaciones de muestreo claras y concisas.

Ante la diversidad de interacciones posibles entre factores extrínsecos e intrínsecos, el volumen de muestras a recolectar dependerá en gran medida de nuestros objetivos y del conocimiento que tengamos del organismo diana y de su contexto ecológico. Es muy importante tener presente que objetivos diferentes exigen muestreos también diferentes, y que un mismo muestreo no es adecuado para responder a todas las cuestiones que puede plantear la diversificación o la conservación de un grupo de poblaciones.

En la mayoría de los casos, partiremos de una ignorancia total en cuanto a los niveles y distribución de la variabilidad genética en las poblaciones diana. Por ello, hemos de procurar representarlas de la mejor manera posible en previsión de que contuvieran mayor heterogeneidad genética de la que pudieramos prever a través de indicadores poco fiables como el tamaño poblacional o la distribución geográfica restringida. Esto es tanto más así cuanto que la caracterización de estas poblaciones y su eventual diferenciación pueden depender en gran medida de los datos genéticos que obtengamos.

Como punto de partida general, puede establecerse que, en igualdad de condiciones, la diversidad genética de una muestra es proporcional al logaritmo del tamaño de la muestra (Brown y Briggs 1991). La Figura 5A ilustra esta relación, que significa que muestras más grandes tenderán a contener mayores niveles de variación genética hasta alcanzar un punto de inflexión (señalado mediante una flecha en la Fig. 5B), a partir del cual se necesitarán incrementos muy grandes en el tamaño de la muestra para aumentar significativamente la variación genética detectada. En el ejemplo de la figura, incrementar solamente en una unidad de  $x$  la variación genética detectada a partir del punto de inflexión (correspondiente a una variación genética de  $4x$ ) implica cuadruplicar el número de muestras ( $4n$ ). Este aspecto tan importante del muestreo se comenta en relación a la Figura 6.

Esto significa que si todas las poblaciones a muestrear fueran muy grandes, la incidencia de incrementos de muestreo (o de tamaño poblacional) en la detección de variación genética adicional serían poco sustanciales a partir del punto de inflexión. No obstante, puesto que muchas de nuestras poblaciones serán de tamaño pequeño, se traslucen dos implicaciones muy importantes para la Fase III del muestreo:

1. Puesto que el impacto del incremento del número de muestras en la proporción de variabilidad capturada es mucho mayor cuantos menos individuos contiene la población, **el número de muestras a recolectar en poblaciones pequeñas ha de ser el máximo posible y guardar estricta relación con el tamaño estimado de la población.**

La consideración de este aspecto se hace especialmente importante cuando nuestras poblaciones diana son endémicas o están amenazadas. La Figura 6 muestra que la detección de variación adicional en poblaciones pequeñas implica muy poco esfuerzo (13 muestras más en una población de 50 individuos o 24 más en una de 100 individuos). Además, vemos que solamente las poblaciones de tamaño igual o menor a 1.000 individuos nos permitirían coger una muestra de 100 individuos que representara una proporción de al menos el 60% de la variación genética poblacional.



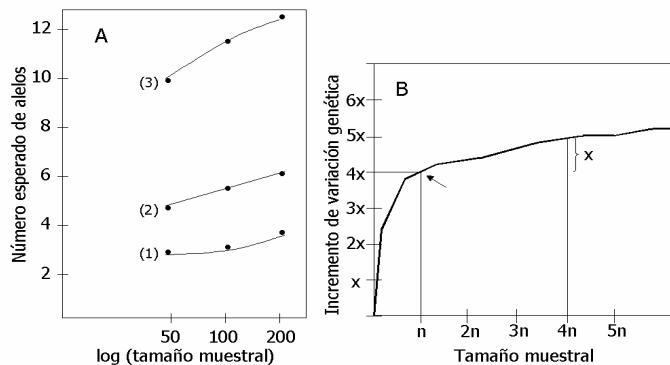


Figura 5. A: Número esperado de alelos en una muestra de tamaño  $n$  (ordenadas) en función del logaritmo del tamaño muestral (abscisas) en una población (1) que dista 0.4 generaciones de ser monomórfica, (2) estacionaria y (3) compuesta de un número elevado de alelos con frecuencias idénticas (adaptado a partir de Brown y Briggs 1991). B: Representación abstracta de la relación entre el tamaño de la muestra (abscisas) y el incremento de variación genética detectada (ordenadas).

2. En poblaciones muy grandes, no tiene sentido aumentar el tamaño de la muestra más allá de cierto límite, ya que no conseguiríamos incrementar sustancialmente la proporción de variabilidad genética detectada aunque dobláramos el número de muestras. Por tanto, **para poblaciones mayores de un cierto tamaño, es conveniente establecer un número de muestras umbral, siempre y cuando se represente bien la distribución espacial de las plantas en el área que ocupa la población.**

Por un lado, la Figura 6 muestra que incluso los tamaños de muestra necesarios para representar solo el 60% de variación en poblaciones grandes quedarían fuera del rango que es factible recolectar y analizar para la mayoría de trabajos de investigación. Por otro lado, una muestra de 100 individuos en estas poblaciones muestrearía una proporción de solamente el 20% de la variabilidad genética total (en poblaciones de 10.000 individuos) y de menos del 10% (en poblaciones de un millón de individuos), lo cual significa que el tamaño de muestra mínimo debe ser superior a 100 individuos (en la Figura 7 se sugiere un máximo orientativo de 400 individuos).

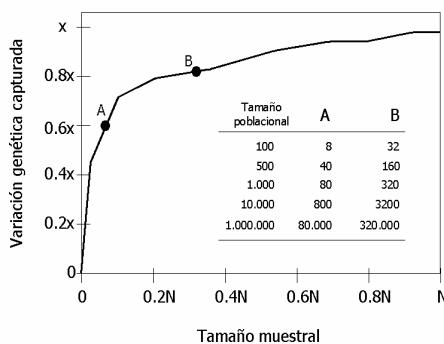


Figura 6. Proporción de variación genética capturada ( $x$  = variación genética total) en función del tamaño muestral ( $N$  = tamaño de la población). En la tabla del interior del gráfico se muestra el número absoluto de individuos a muestrear para capturar el 60% (A) y el 80% (B) de la variabilidad genética en función del tamaño poblacional.



Estas consideraciones son importantes a nivel general, ya que el tamaño poblacional tiene la ventaja de ser estimable por observación directa (a menos que no haya bancos de semillas en el subsuelo, elevada incidencia de la dormancia o niveles considerables de reproducción vegetativa). No obstante, en la naturaleza Canaria la homogeneidad es excepcional y, aunque en ocasiones el tamaño poblacional puede ser casi determinante del tamaño de la muestra, la mayoría de veces habremos de atender a otros factores que también tienen impacto en la variabilidad genética poblacional, pero que no son estimables de manera tan directa.

Dependiendo de la combinación de estos factores, especies diferentes (o poblaciones con-específicas) que comparten un mismo tamaño poblacional pueden tener niveles de variación muy dispares. A nivel general, estos factores pueden dividirse en *intrínsecos* (dependientes del linaje evolutivo de los taxones diana y de sus características biológicas) y *extrínsecos* (es decir, relacionados con el ambiente que les ha tocado vivir y con su historia, ya sea reciente o pasada). Dado que muchas veces los factores intrínsecos y la historia lejana de los taxones son desconocidos o poco conocidos, es conveniente considerar al menos los factores extrínsecos estimables antes de decidir cuantos individuos muestrear. Partiendo de las indicaciones generales del organigrama de muestreo (Fig. 4), solamente evaluando estas características de las poblaciones en conjunción con su tamaño estimado podremos decidir cuántas muestras recolectar. La Figura 7 pretende ser una síntesis orientativa de la influencia de estas características en la cantidad de muestras (poblaciones e individuos) a recolectar.

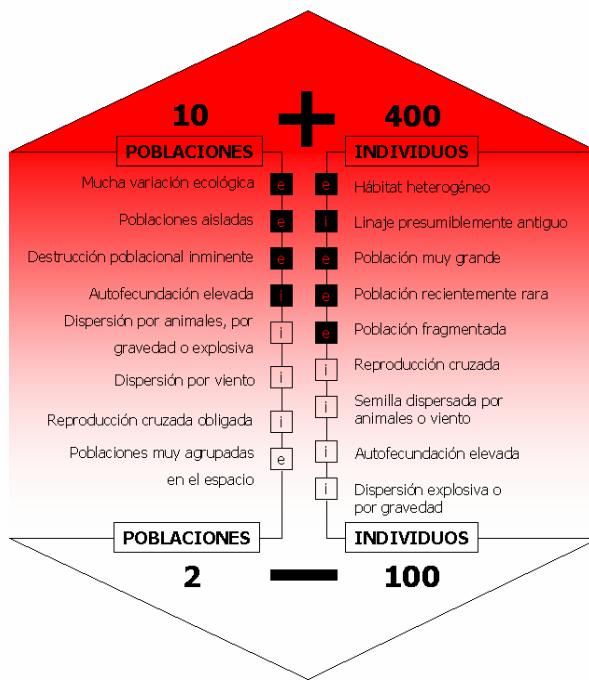


Figura 7. Jerarquía de indicadores extrínsecos (e) e intrínsecos (i) para determinar el número de poblaciones y número de individuos a muestrear en los taxones diana que consten de más de una población con más de 100 individuos. La convergencia de dos o más factores situados en el extremo superior de la flecha (color rojo más intenso) determinará un muestreo más amplio. En caso de que en una misma población diana converjan varios factores, prevalecerá el más crítico (es decir, el que obligue a recolectar un mayor número de muestras). Para decidir cuántos individuos muestrear, conviene considerar al menos los factores extrínsecos. Los recuadros rellenos de negro con las letras de color rojo indican los factores más influyentes. Los números arriba y debajo de las flechas son, respectivamente, el número máximo y mínimo de poblaciones e individuos a recolectar (orientativo). Inspirado en la Tabla A.2 del Center for Plant Conservation (1991).



### Fase III. Recolección de muestras

En esta fase, iremos recolectando hojas de los individuos seleccionados a lo largo de los transectos. Agilizaremos mucho la colecta de muestras si tenemos las bolsas ordenadas correlativamente (ver los apartados 1.1 y 1.2), de manera que el muestreo consista simplemente en sacar las bolsas, abrirlas, llenarlas de hojas (un individuo por bolsa), cerrarlas herméticamente y depositarlas en el sitio asignado al almacenamiento de las muestras.

La hoja a recolectar debe satisfacer los requisitos cualitativos y cuantitativos explicados en los apartados 2.3.1, 2.3.2 y 2.3.3. Si hay dos personas encargadas del muestreo, es muy conveniente que una de ellas se dedique a recoger las hojas y la otra a dibujar la distribución de los individuos recolectados en las coordenadas que estarán en el reverso de la hoja de muestreo correspondiente (ver Figura 2). Este dibujo no tiene porqué ser exacto (es decir, las distancias entre los individuos no tienen que medirse con precisión milimétrica), pero sí debe mostrar de manera fidedigna la distribución relativa de los individuos en el espacio poblacional.

Aunque alguien podría pensar que el dibujo no es imprescindible, conocer la distribución de los individuos en el espacio poblacional es fundamental si el intercambio de genes no es regular, o si una población consta de varias subpoblaciones con dinámicas independientes (esto es, poco flujo génico y poco intercambio de propágulos). Ninguno de estos aspectos tan relevantes sería revelado por los datos moleculares en ausencia de esta información espacial tan fácil de obtener. Por otra parte, dibujar el mapa de los individuos colectados no consume mucho tiempo adicional aunque sea una sola persona la que lo lleve a cabo, puesto que muchas de las poblaciones diana a las que se refiere esta brújula genética son endémicas o amenazadas y presumiblemente contienen pocos individuos.

Al final del muestreo se almacenarán las muestras según estas vayan a ser analizadas con isoenzimas (ver apartados 2.3.1 y 2.3.2) o técnicas de DNA (ver apartados 2.3.1 y 2.3.3) y se rellenará la celda de la hoja de muestreo (Figura 1) correspondiente al número de muestras recolectadas.



## Normas prácticas de muestreo de aplicación general (DNA e ISOENZIMAS)

---

a) Seguir las indicaciones generales de los apartados 2.1, 2.2 y 2.3.

Probablemente encontraremos en el campo situaciones intermedias entre dos o más de las que se contemplan, pero es importante tener estas indicaciones como pauta de actuación.

b) Fijarse bien en que todas las hojas recolectadas corresponden a la especie diana y en que no haya insectos u otros organismos pequeños adheridos a las hojas.

Es muy fácil que una aguja de pino o algún trozo de otra planta se cuele en la bolsa de muestreo. Si todas las hojas de la población mostraran síntomas de infección por parásitos o virus, limpiarlas con un paño limpio empapado en agua destilada y secarlas bien con un trozo de papel higiénico antes de almacenarlas.

c) No muestrear exclusivamente las plantas más sanas y vistosas de la población.

Aunque las plantas sanas y robustas llaman siempre más la atención, las plantas macilentas e inconspicuas también forman parte de la población y deben de ser muestreadas en aras de una correcta representatividad. Con frecuencia, las plantas deformes o menos atractivas poseen también alelos raros, por lo cual infra-representarlas podría provocar una estimación sesgada de los niveles de variación genética de la población.

d) Nunca muestrear una planta si la recolección de las hojas que necesitemos puede poner en peligro su supervivencia.

La mayoría de plantas transforman la energía solar en energía metabólica en las hojas, por lo que no debemos menoscabar su capacidad para desarrollar la fotosíntesis hasta extremos que dificulten su supervivencia (especialmente en caso de plantas amenazadas o con muy bajos tamaños poblacionales).

e) Seleccionar siempre las hojas más verdes y limpias de entre las disponibles en la planta.

Esta manera de proceder minimizará los posibles problemas de contaminación en las muestras.

f) La cantidad de hoja a muestrear por individuo varía, pero es aconsejable coger de 1 a 5 gramos (peso fresco).

g) En caso de tener que coger más de una hoja para asegurar la cantidad mínima de material por espécimen, debemos asegurarnos de que se muestrea siempre el mismo individuo.

Frecuentemente varios pies de diferentes plantas se hallan entremezclados, lo cual hace muy fácil equivocarse y representar la variación genética de dos plantas diferentes en una misma muestra. Este error tan simple invalidaría los datos genéticos obtenidos de esa muestra.

h) Procurar recolectar las hojas enteras, sin causarles heridas

El tamaño de las hojas de la mayoría de plantas permite depositarlas enteras dentro de la bolsa de muestreo. Hay que procurar desprenderlas del tallo o de la rama causando el mínimo de heridas, ya que en caso contrario se deteriorarían más rápidamente por los factores mencionados en el punto 7.

i) Siempre que el tamaño de la planta lo permita, coger al menos un duplicado de cada una de las muestras.

Este proceder se hace muy aconsejable especialmente en previsión de posibles deterioros de las muestras o accidentes en el laboratorio de análisis.

j) Si las plantas son leñosas o perennes, es conveniente marcar los individuos colectados en el campo con el mismo código que el que se les asigna para los análisis moleculares (en el JBCVC utilizamos cinta Dymo, pero hay métodos más seguros).

De esta manera nos aseguramos eventuales re-muestreos y disponemos de una información valiosísima para la gestión de los datos moleculares individuales. En efecto, con esta sencilla técnica de marcaje, las poblaciones naturales pueden actuar como auténticos *archivos de variabilidad genética*.



## Normas prácticas de muestreo de aplicación específica para ISOENZIMAS

---

- a) Nunca utilizar gel de sílice para conservar las muestras de hoja, con independencia de si estas son crasas o suculentas.
- b) Llevar siempre una nevera de campo de tamaño adecuado para el volumen de muestras que se pretenda recolectar, con dos o más bloques refrigerantes congelados dentro para preservar las muestras frescas hasta que lleguen al laboratorio de análisis.
- c) La rapidez de transporte de las muestras al laboratorio de análisis es crucial. En ningún caso deben transcurrir más de tres días desde que la muestra de hoja se recoge hasta que llega al laboratorio de análisis. La visualización de los isoenzimas se basa en que estos conserven su capacidad catalítica. Si las hojas pasan más de tres días en condiciones infra-óptimas (ya sea en la nevera de campo, a temperatura ambiente o en un frigorífico convencional), los enzimas se deterioran y pierden totalmente sus capacidades catalíticas. El transporte se realizará en neveras de campo con bloques refrigerantes congelados para preservar el frescor.

## Normas prácticas de muestreo de aplicación específica para DNA

---

- a) Recoger siempre hojas adultas que no hayan empezado la senescencia.  
Según la experiencia del Banco de DNA de Kew (Mark Chase y Michael Fay, comunicación personal) el DNA extraído de las hojas jóvenes de algunos grupos de plantas es imposible de amplificar debido a la presencia de compuestos químicos que no se hallan en otros tejidos. Estos compuestos probablemente impidan la predación por parte de insectos o animales antes de que las hojas se hayan endurecido.
- b) Asegurarse de que las hojas están completamente secas después de 24 horas.  
En caso contrario, la experiencia de la mayoría de laboratorios dicta que es muy probable que el DNA que se extraiga esté muy degradado. Quizás el requisito más importante para la buena preservación y amplificación del ADN sea un secado rápido de las hojas. Según Savolainen et al. (1995), si el secado se prolonga demasiado, la planta (que aun está viva) estará sujeta a un estrés hídrico, a la carencia de nutrientes y a heridas (entre ellas, las que hemos ocasionado al desprender la hoja de la planta, al doblarla para introducirla en la bolsa, etc...). Las heridas inducen rápidamente la producción de compuestos fenólicos y radicales libres, que pueden influenciar muy fuertemente el ambiente celular y, en consecuencia, perjudicar la calidad del ADN obtenido en la extracción (McKersie et al. 1988). Además, se ha visto que existe una disminución del contenido en ADN durante los primeros estadios de la senescencia foliar, que puede representar hasta el 20% del ADN total (Scott y Possingham 1983). Las respuestas metabólicas y celulares a la preparación de pliegos de herbario son similares a las debidas a la senescencia. Por todos estos motivos, si no estamos seguros de poder realizar la extracción de este tipo de hojas cuando aun están frescas, es preferible prescindir de muestreárlas.
- c) Si las hojas no son crasas ni suculentas, depositar cada muestra en una bolsa cerrada herméticamente que contenga gel de sílice libre de humedad y el mínimo de aire posible.  
Estas muestras se conservan bien a temperatura ambiente, siempre y cuando se almacenen en un sitio fresco resguardado de la humedad y de la luz. Si el color del gel de sílice virara a azul pálido (indicador de presencia de humedad) es necesario sustituirlo por gel de sílice seco (es decir, de color azul intenso).
- d) Si las hojas son crasas o suculentas, es muy aconsejable realizar la extracción el mismo día de la recolección.  
En caso de que esto no sea factible, hay dos opciones para conservarlas:
  - d1) Procurar que pierdan la mayoría del líquido antes de pasarlas al gel de sílice (se puede utilizar papel secante y una prensa para acelerar el proceso).
  - d2) En caso de que d1 no fuera posible por razones técnicas o de tiempo, habría que preservar las hojas en el gel de sílice dentro de una nevera de campo mientras dure el muestreo, pero almacenarlas en un congelador entre -20°C y -40°C (quitando antes el gel de sílice) en cuanto lleguen al primer lugar de destino.
- e) Si el primer lugar de destino no es el laboratorio de análisis genético, el transporte a este se realizará procurando que transcurra el menor tiempo posible desde la recolección.



### 1.3. DESPUÉS DE IR A LA POBLACIÓN

A efectos de agilizar el trabajo y facilitar el almacenamiento en los laboratorios de destino, es conveniente ordenar todas las muestras correspondientes a una misma población y agruparlas dentro de una bolsa hermética de volumen adecuado o en una fiambrera de plástico de cierre hermético. En ambos casos, hay que etiquetar convenientemente el recipiente que contiene las muestras para que la identificación de la especie y población se produzcan sin ambigüedades en el laboratorio de análisis.

En caso de que el destino de las muestras sea el análisis con DNA, el transporte al laboratorio de análisis no tiene porqué ser urgente, aunque se hace aconsejable la máxima rapidez posible.

Los protocolos analíticos para estimar la variación genética no se incluyen en esta brújula, ya que se explican exhaustivamente en los manuales que ha elaborado el Departamento de biodiversidad Molecular del JBCVC para aloenzimas (Caujapé-Castells et al. 1999) y para el Banco de ADN de la Flora Canaria que reside en esta institución (Caujapé-Castells, Jaén-Molina y Cabrera-García, 2006).



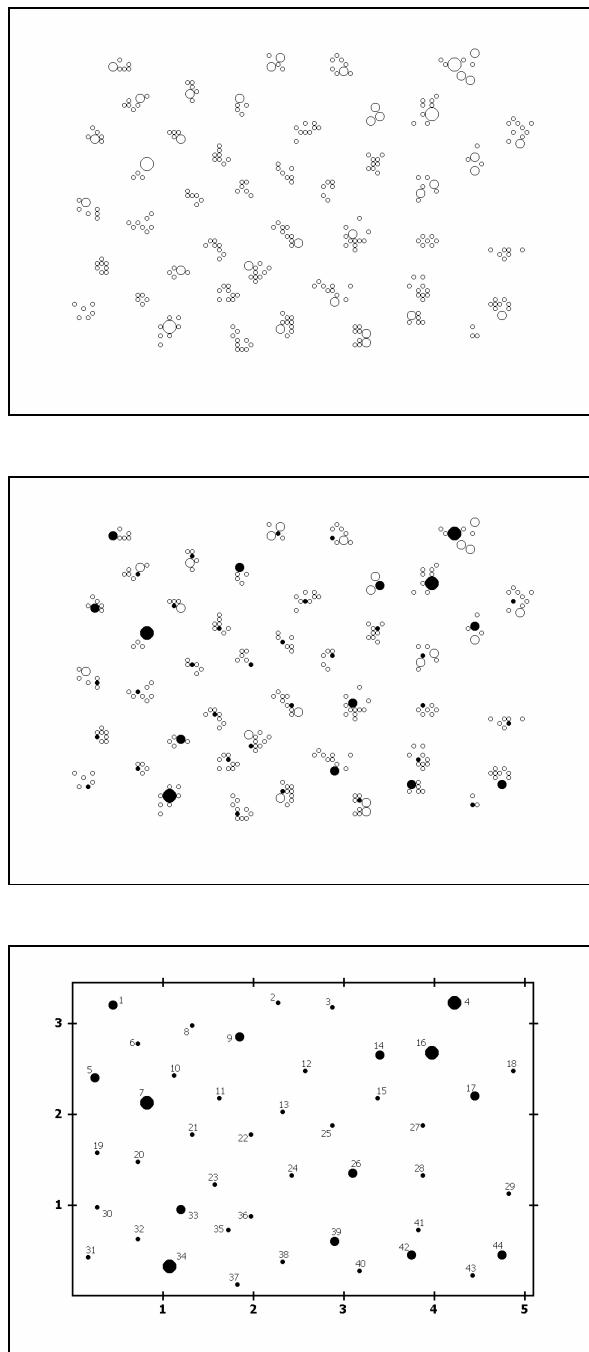


Figura 8. Ejemplificación de las tres fases temporales del muestreo esquematizadas en la Figura 4 para una población de tipo B (plantas distribuidas en pocos grupos, con menos de 11 individuos por grupo). Los círculos vacíos simbolizan las plantas de la población y los círculos rellenos corresponden a las plantas muestreadas. En la última representación se muestra el mapa que correspondería a este muestreo, con los códigos numéricos de cada planta al lado del círculo que la representa.



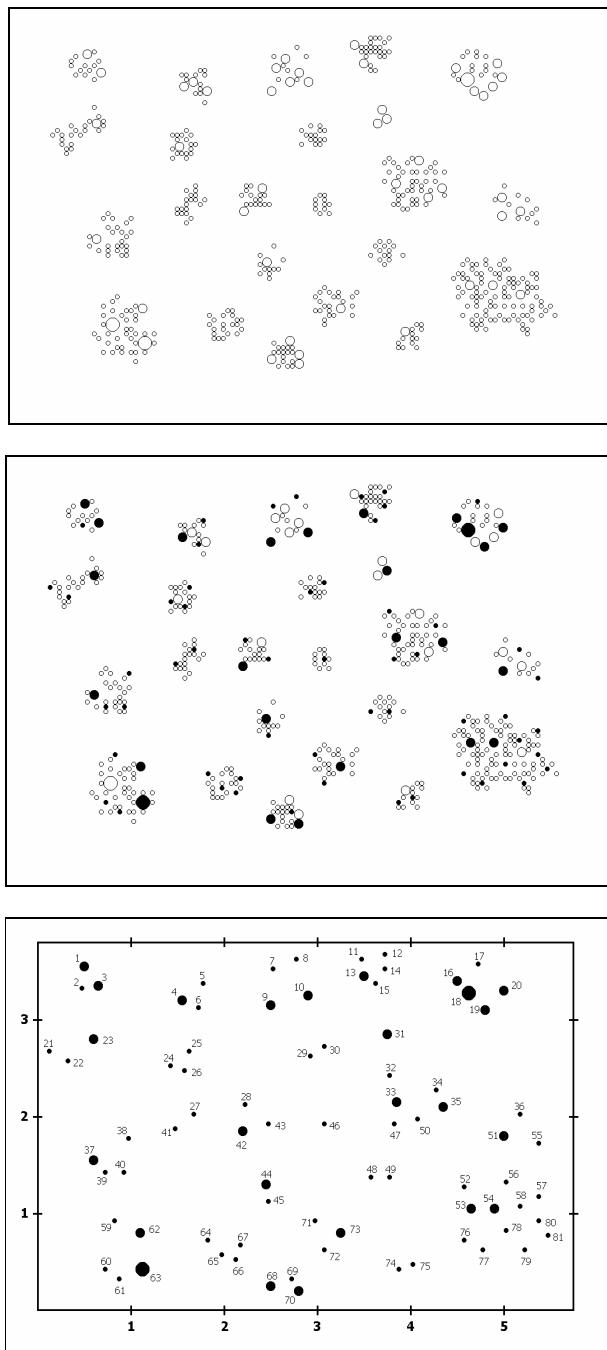


Figura 8. Ejemplificación de las tres fases temporales del muestreo esquematizadas en la Figura 4 en una población de tipo C (plantas distribuidas en pocos grupos, con más de 10 individuos por grupo). Los círculos vacíos simbolizan las plantas de la población y los círculos rellenos corresponden a las plantas muestreadas. Como se indica en el organigrama de la Figura 4, el muestreo de cada uno de los grupos se realizaría en proporción a su tamaño y representando el área que ocupan. En la última representación se muestra el mapa que correspondería a este muestreo, con los códigos numéricos de cada planta al lado del círculo que la representa.



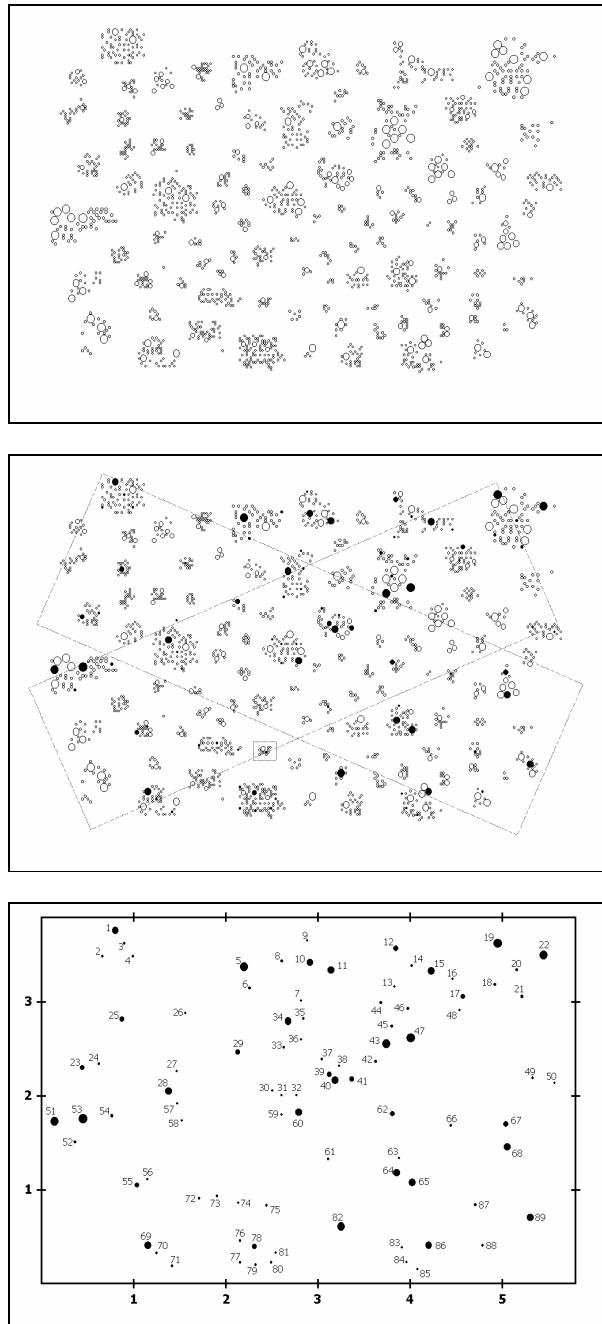


Figura 10. Ejemplificación de las tres fases temporales del muestreo esquematizadas en la Figura 4 en una población de tipo D (plantas distribuidas en muchos grupos, con menos de 11 individuos por grupo). Los círculos vacíos simbolizan las plantas de la población y los círculos rellenos corresponden a las plantas muestreadas. Los rectángulos dibujados en línea discontinua en la segunda representación se refieren a uno de los transectos posibles para muestrear la población. Nótese que aunque el muestreo se realiza eminentemente a lo largo del transecto, también se muestrean algunas plantas fuera de ellos, en aras de una correcta representación de la distribución espacial de los individuos. En la última representación se muestra el mapa que correspondería a este muestreo, con los códigos numéricos de cada planta al lado del círculo que la representa.



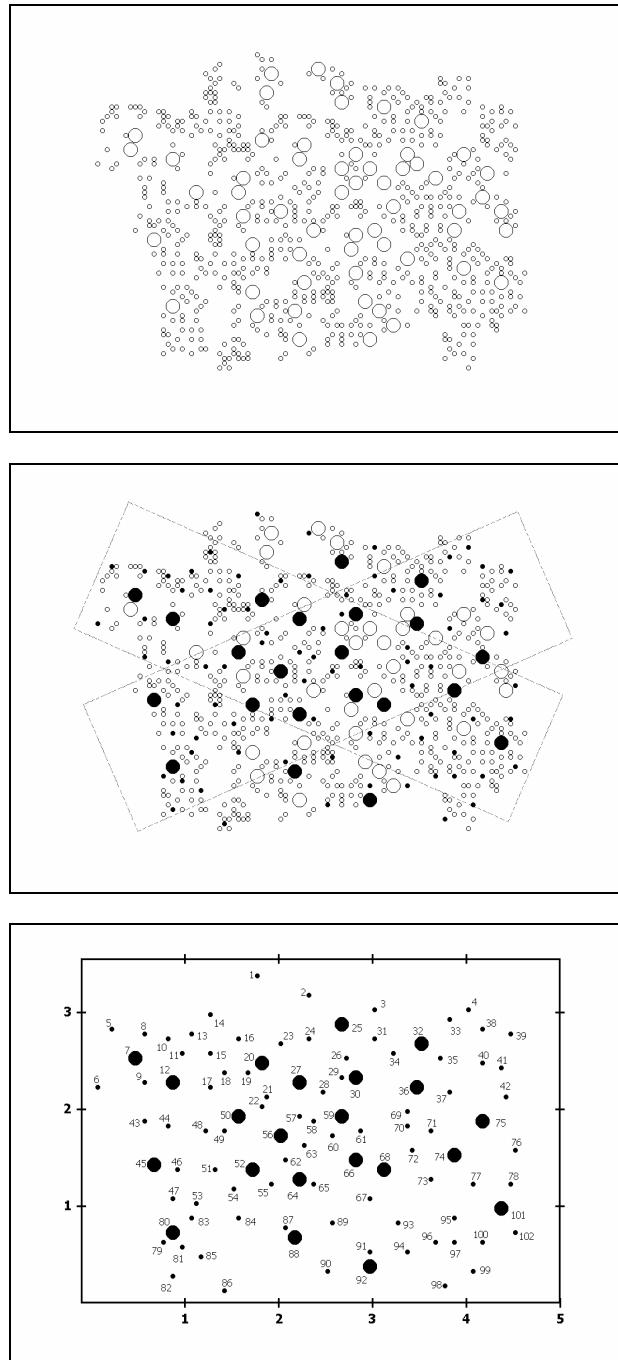


Figura 11. Ejemplificación de las tres fases temporales del muestreo esquematizadas en la Figura 4 en una población de tipo E (plantas distribuidas de manera continua, formando una especie de pradera). Los círculos vacíos simbolizan las plantas de la población y los círculos llenos corresponden a las plantas muestreadas. Los rectángulos dibujados en línea discontinua en la segunda representación se refieren a uno de los transectos posibles para muestrear la población. Nótese que aunque el muestreo se realiza eminentemente a lo largo del transecto, también se muestrean algunas plantas fuera de ellos, en aras de una correcta representación de la distribución espacial de los individuos. En la última representación se muestra el mapa que correspondería a este muestreo, con los códigos numéricos de cada planta al lado del círculo que la representa.



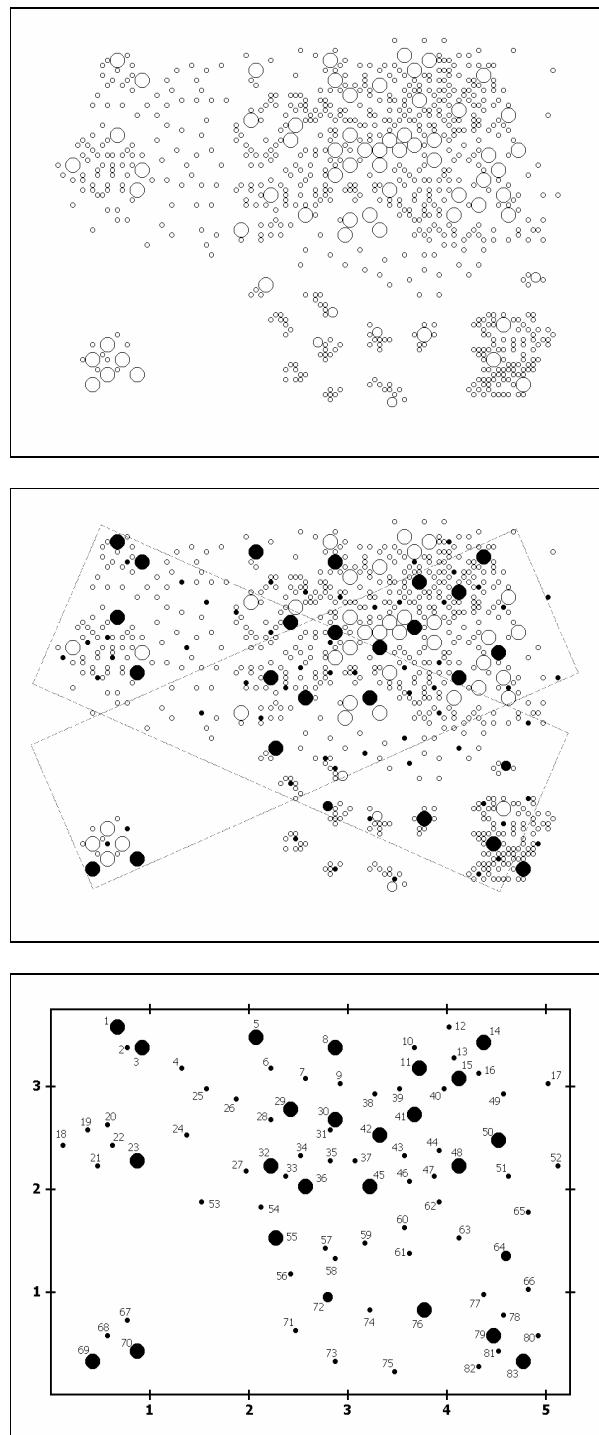


Figura 12. Ejemplificación de las tres fases temporales del muestreo esquematizadas en la Figura 4 en una población de tipo F (plantas distribuidas sin seguir un patrón regular). En este ejemplo, la población consta de zonas con distribución continua y gran densidad de individuos (parte superior derecha), grupos de tamaño variable (parte inferior derecha), plantas aisladas dispersas (especialmente en la mitad superior) y de un núcleo de individuos muy alejados del resto (parte inferior izquierda). Los círculos vacíos simbolizan las plantas de la población y los círculos llenos corresponden a las plantas muestreadas. Los rectángulos dibujados en línea discontinua en la segunda representación se refieren a uno de los transectos posibles para muestrear la población. Nótese que aunque el muestreo se realiza eminentemente a lo largo del transecto, también se muestran algunas plantas fuera de él, en aras de una correcta representación de la distribución espacial de los individuos. En la última representación se muestra el mapa que correspondería a este muestreo, con los códigos numéricos de cada planta al lado del círculo que la representa.

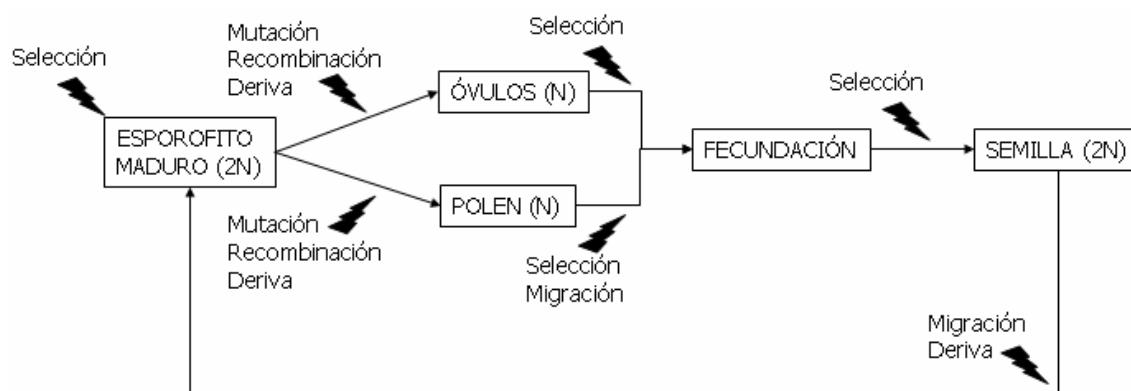


## 2. SOBRE LOS PARÁMETROS Y TESTS MÁS USADOS EN GENÉTICA DE POBLACIONES



## INTRODUCCIÓN

No le falta razón a Jorge Wagensberg cuando dice que la evolución de la vida es posible gracias a cierta capacidad para acumular innovaciones superfluas (Wagensberg 2006, reflexión 376). Efectivamente, en las poblaciones naturales de organismos lo delicuente se puede transformar en trascendente con sólo que se produzca un cambio (no necesariamente drástico) en el ambiente. Entonces, algunas de las innovaciones que en el estado anterior eran "superfluas" pueden dejar de serlo, y ofrecer soluciones efectivísimas a los nuevos problemas de supervivencia que el nuevo ambiente plantea (problemas que, antes del cambio, no existían). Es como si los organismos invirtieran parte de su energía vital en generar diversidad "a fondo perdido", como anticipo a lo impredecible, como un capital de riesgo evolutivo que puede dar increíbles réditos si el ambiente cambia y la suerte acompaña, pero que no supone ningún trastorno para su supervivencia en caso de que todo continúe como antes.



Esquematización del ciclo vital de las angiospermas, incluyendo la incidencia de las cinco fuerzas evolutivas que interactúan para generar la estructura genética de las poblaciones naturales de plantas.

Las innovaciones en la estructura genética de las poblaciones naturales se generan a través de cinco fuerzas evolutivas: mutación, recombinación, selección, migración y deriva. Como muestra el esquema superior, algunas de ellas actúan solamente en una de las fases del ciclo biológico de los organismos y (aunque no lo muestre el esquema superior) otras predominan sobre las demás en determinadas zonas geográficas, o en situaciones particulares.

Puesto que las innovaciones se producen al azar, es inevitable que algunas de ellas afecten drásticamente a genes que regulan funciones vitales. Lo más probable es que este tipo de innovaciones no aporten ventajas incuestionables, sino más bien todo lo contrario, por lo que el organismo que las hereda muere sin remisión, ya sea antes o después de nacer. Pero cuando las innovaciones no afectan a la supervivencia, tienden a comportarse como neutrales (esto es, no aportan grandes ventajas ni inconvenientes al organismo que las posee), y pueden expandirse en el espacio y en el tiempo, esperando a través de las generaciones a que llegue una oportunidad donde (quizás) poder demostrar su valor.

En la naturaleza, en las plantas vasculares, la heterogeneidad genética es una garantía, una especie de seguro de vida que otorga a las poblaciones que la poseen mayores posibilidades de sobrevivir a un cambio ambiental o a muchas catástrofes (siempre, claro está, que los cambios no sean en forma de una agresión física que aniquile todos los sistemas fisiológicos de los individuos, como una inundación, una erupción volcánica o una plaga).

La genética de poblaciones estudia las implicaciones del cambio genético en el tiempo evolutivo (a lo largo de las generaciones) y en el espacio (dentro y entre poblaciones en un momento puntual del tiempo). Para este cometido se sirve de formulaciones matemáticas que, en la práctica, se aplican a la información molecular obtenida para las poblaciones naturales sujeto de estudio.



Estas formulaciones se apoyan en modelos que nos otorgan un cierto poder de predicción a cambio de asunciones sobre la complejidad de la naturaleza. Y como cualquier modelo de cualquier ciencia es por definición una simplificación, antes de interpretar los parámetros resultantes de su aplicación hemos de ser conscientes de las limitaciones que imponen las asunciones que hemos aceptado. Nunca hay que olvidar que los parámetros en que nos fijamos para establecer los modelos se refieren a organismos vivos y a procesos inacabados cuya complejidad nunca puede ser descrita totalmente por una perspectiva reduccionista.

A lo largo de esta parte de la brújula, veremos que la información genética poblacional nos puede servir para inferir la historia de las poblaciones y también para hacer predicciones sobre su futuro.



## 2.1. CONCEPTOS BÁSICOS

---

Este capítulo introductorio trata de nueve conceptos básicos que serán recurrentes a lo largo de las partes 3 y 4 de esta brújula y cuya comprensión es imprescindible para la correcta interpretación de los valores de los parámetros más comúnmente usados en genética de poblaciones. Estos nueve conceptos son:

- El teorema de Hardy-Weinberg
- La deriva genética
- El efecto fundador
- El efecto cuello de botella
- La paradoja del tamaño poblacional
- El efecto Wahlund
- El tamaño poblacional efectivo
- El modelo de estructura poblacional en isla
- El modelo de estructura poblacional en pasadera (del inglés “stepping stone”)



**EL TEOREMA DE HARDY-WEINBERG** es una formulación matemática que muestra que las frecuencias de los diferentes alelos en poblaciones ideales de tamaño infinito no cambian en el tiempo a menos que actúen fuerzas evolutivas

La mayoría de inferencias sobre la diversificación de las poblaciones se realizan en base al sentido de las desviaciones de las frecuencias génicas observadas respecto de las esperadas bajo el equilibrio que define el teorema de Hardy-Weinberg. Por ello, comprender el teorema de Hardy-Weinberg es fundamental para empezar a comprender muchos aspectos de la genética de poblaciones de los organismos.

Por definición, el equilibrio Hardy-Weinberg describe la genética de una población que no evoluciona. Esta información proporciona la base comparativa necesaria para detectar cambios genéticos que pueden tener significación evolutiva en las poblaciones naturales.

### Conceptos preparatorios

Asumamos que estamos estudiando un locus dialélico (alelos *A* y *a*) en una población natural compuesta por *N* individuos diploides. Las frecuencias genotípicas de esta población tipo pueden expresarse según la Tabla 1.

Tabla 1. Estimación de las frecuencias genotípicas poblacionales de cada uno de los posibles genotipos en un locus dialélico. 'N' indica el tamaño poblacional total, y 'n' es el número de individuos de un genotipo determinado (entre paréntesis).

Genotipos posibles			
	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Número de individuos	$n(AA)$	$n(Aa)$	$n(aa)$
Frecuencia genotípica	$n(AA)/N = D$	$n(Aa)/N = H$	$n(aa)/N = R$

Según la Tabla 1, notaremos la frecuencia relativa de los individuos homocigotos *AA* como *D* (por Dominantes), la de los heterocigotos *Aa* como *H* (por Hetetrozigotos) y la de los homocigotos *aa* como *R* (por Recesivos). Por la definición de frecuencia relativa, tenemos que  $D+H+R=1$ .

La frecuencia alélica es la proporción de un tipo particular de alelo en relación a todos los alelos detectados en la población en el locus al que pertenece y se estima mediante la frecuencia relativa. En la población descrita en la Tabla 1, pueden calcularse las frecuencias alélicas de *A* y *a* a partir de las frecuencias genotípicas de manera que, si notamos *p* como la frecuencia del alelo *A* y *q* como la frecuencia del alelo *a*, entonces

$$p = \text{Nº de alelos } A / \text{Nº total de alelos} = [2n(AA) + n(Aa)]/2N = D + \frac{1}{2}H$$
$$q = \text{Nº de alelos } a / \text{Nº total de alelos} = [2n(aa) + n(Aa)]/2N = R + \frac{1}{2}H$$

(puesto que asumimos que los individuos son diploides, el número total de alelos analizados es  $2N$ )

Por la definición de frecuencia relativa, tenemos que  $p + q = 1$ .



La probabilidad de que todas las posibles combinaciones de los alelos *A* y *a* ocurran al azar es

$$(p + q)^2 = 1$$

y, por tanto,

$$p^2 + 2pq + q^2 = 1$$

Fijémonos que, en esta ecuación,

$p^2$  = frecuencia de los homozigotos dominantes (AA)

$2pq$  = frecuencia de los heterocigotos (Aa)

$q^2$  = frecuencia de los homozigotos recesivos (aa)

### Formalización del teorema de Hardy-Weinberg

- Consideremos una población monoica de individuos diploides con apareamiento aleatorio (o panmixia), cuyo tamaño es lo suficientemente grande como para que la influencia de los fenómenos estocásticos no sea apreciable.
- Asumamos que esta población no se ve afectada por la mutación, la migración o la selección natural y que las generaciones no se solapan en el tiempo.
- Consideremos un locus genético dialélico de esta población en el cual los alelos *A* y *a* presentan las probabilidades de detección  $P(A) = p$  y  $P(a) = q$ .

Siempre que todas las asunciones postuladas se mantengan, podemos enunciar el siguiente teorema:

"Después de una generación de apareamiento aleatorio, una población con frecuencias genotípicas  $D$  (para AA),  $H$  (para Aa) y  $R$  (para aa), llega a una situación de frecuencias génicas estables  $p^2$ ,  $2pq$ ,  $q^2$  donde  $p = D + \frac{1}{2}H$  y  $q = R + \frac{1}{2}H$ . Si las frecuencias genotípicas iniciales ya son de la forma  $p^2$ ,  $2pq$ ,  $q^2$ , entonces esas frecuencias ya son estables para todas las generaciones siempre y cuando las condiciones iniciales prevalezcan".

Esta formulación es conocida como el **teorema de Hardy-Weinberg** porque fue probada por primera vez (y casi simultáneamente) por el matemático inglés Godfrey Hardy (1908) y el médico alemán Wilhelm Weinberg (1908). En una población ideal que satisfaga las condiciones postuladas, se cumplirán las proporciones Hardy-Weinberg. La utilidad práctica de este teorema reside en que podemos evaluar la magnitud del cambio evolutivo en las poblaciones analizadas comparando las frecuencias genotípicas estimadas en las poblaciones naturales que estudiamos con las que esperaríamos según el equilibrio Hardy-Weinberg,

La Tabla 2 ofrece una ilustración gráfica de este teorema mediante la descomposición de las frecuencias genotípicas esperadas en la descendencia de cada uno de los apareamientos que tendrían lugar en una población panmíctica.

Cualquier persona mínimamente observadora que reflexione un poco acerca del funcionamiento de la naturaleza se dará cuenta de que las condiciones asumidas para definir la población ideal en el teorema de HW son extremadamente irreales: las poblaciones naturales jamás mantienen un tamaño constante, las generaciones sí se solapan en el tiempo y el apareamiento nunca es totalmente aleatorio. A lo largo del libro veremos



numerosos ejemplos donde las desviaciones que detectemos de estas condiciones irreales nos orientarán en nuestras inferencias sobre la diversificación evolutiva y la estructura genética de los organismos.

Empezar esta parte de la brújula hablando del teorema de Hardy Weinberg no es una casualidad, ya que el teorema de Hardy Weinberg funciona como una hipótesis nula, y la genética de poblaciones consiste en gran medida en investigar qué factores provocan la desviación significativa de esta hipótesis nula en las poblaciones naturales de organismos.

Tabla 2. Frecuencias de apareamientos y frecuencias genotípicas esperadas en la descendencia de una población infinita de individuos diploides con apareamiento al azar. Notemos que la frecuencia de cada uno de los genotipos permanece constante después de una generación de panmixia. Este es el resultado más destacable del teorema de Hardy-Weinberg.

Apareamiento	Frecuencia	Frecuencias genotípicas esperadas en la descendencia		
		AA	Aa	aa
AA x AA	$D^2$	= $D^2$	+	0
AAxAa y AaxAA	$2DH$	= $\frac{1}{2} 2DH = DH$	+	$\frac{1}{2} 2DH = DH$
AAxaa y aaxAA	$2DR$	= 0	+	$2DR$
AaxAa	$H^2$	= $\frac{1}{4} H^2 = H^2/4$	+	$\frac{1}{2} H^2 = H^2/2$
Aaxaa y aaxAa	$2HR$	= 0	+	$\frac{1}{2} 2HR = HR$
aaxaa	$R^2$	= 0	+	$0$
<b>SUMAS</b>	$(D+H+R)^2$	$D^2+DH+H^2/4 = [D+(H/2)]^2$	$DH+(H^2/2)+2DR+HR = 2[D+(H/2)][(H/2)+R]$	$H^2/4+HR+R^2 = [(H/2)+R]^2$
	▼	▼	▼	▼
	▼	▼	▼	▼
	<b>1</b>	<b>=</b>	<b><math>p^2</math></b>	<b><math>+ 2pq + q^2</math></b>

Para saber qué causas han actuado para provocar estas desviaciones, la ecuación de Hardy Weinberg es obviamente insuficiente, por lo cual tendremos que aplicar las formulaciones matemáticas que se discuten en esta brújula e interpretar sus resultados según nuestro conocimiento de la biología, ecología e historia del organismo en cuestión. Este proceso de exégesis es el núcleo de acción de la genética de poblaciones.

Es importante advertir que, a pesar que las violaciones de las condiciones del modelo de HW son muy frecuentes, a veces las proporciones HW se cumplen en (algunos loci de) algunos organismos. No obstante, es asimismo fundamental hacer hincapié en que el hecho de que una población natural *cumpla las proporciones HW* no implica directamente que *esté en equilibrio HW*. Notemos que afirmar que una población está en equilibrio HW implica el cumplimiento de todas las condiciones implícitas al modelo, mientras que la mera adecuación a las proporciones HW puede ser debida a factores aleatorios o relacionados con un mal muestreo. La confusión entre factores asociados al muestreo y factores asociados a características de los organismos emana directamente de nuestra ignorancia (o de la incompletitud de nuestros datos), y será un tema recurrente a lo largo de esta brújula. Es extremadamente improbable que el teorema de Hardy Weinberg se cumpla en condiciones naturales y, si las proporciones esperadas HW se adaptan a las frecuencias génicas observadas, es muy posible que se trate de un artefacto estadístico que de una realidad genética.

Si se detecta un exceso de heterozigotos, esto puede indicar selección sobredominante o reproducción cruzada obligada. En cambio un defecto de heterocigotos puede deberse a selección, a la presencia de alelos nulos que den la impresión falsa de defecto de



heterocigotos, consanguinidad o efecto Wahlund por la existencia de estructura poblacional que no hemos tenido en cuenta.

A modo de introducción de lo que se desarrollará en el resto de esta parte de la brújula, la Tabla 3 considera las perturbaciones más frecuentes de las condiciones de equilibrio Hardy-Weinberg y los posibles indicadores para su diagnóstico que se tratan en esta brújula.

Tabla 3. Perturbaciones naturales de las asunciones del Teorema de Hardy Weinberg (THW) y capítulos de la brújula donde se explican estas violaciones.

ASUNCIÓN THW	PERTURBACIONES DE LAS ASUNCIÓNES	VER EN LA BRÚJULA...
La población es infinita	Las poblaciones naturales tienen tamaños finitos. Este hecho se traduce (especialmente en las poblaciones pequeñas) en una pérdida generalizada de variabilidad genética y, en ocasiones, una alta incidencia de la consanguinidad.	<ul style="list-style-type: none"> <li>Deriva genética</li> <li>La paradoja del tamaño poblacional</li> <li>Tamaño poblacional efectivo</li> </ul>
Existe panmixia	Existen ciertos tipos de apareamientos predominantes, diferencias en los tiempos de floración, u otros factores atribuibles al sistema reproductivo de los individuos (p. e., plantas no hermafroditas o existencia de sistemas de autoincompatibilidad) o al muestreo.	<ul style="list-style-type: none"> <li>Efecto Wahlund</li> <li>Tamaño poblacional efectivo</li> <li>Estructura genética</li> <li>Consanguinidad</li> </ul>
No hay mutación	La diferenciación por mutación es tanto mayor cuanto más tiempo ha pasado desde la divergencia evolutiva entre las poblaciones consideradas.	<ul style="list-style-type: none"> <li>Distancias genéticas</li> <li>Alelos raros</li> </ul>
No hay migración	El intercambio de migrantes (en forma de granos de polen o semillas) es frecuente entre las poblaciones naturales de plantas vasculares.	Estimación del intercambio genético entre poblaciones
No hay selección	Diferenciación ecológica o predominio de algunos de los genotipos de la descendencia más que otros.	<ul style="list-style-type: none"> <li>La depresión por consanguinidad</li> </ul>
No hay error de muestreo	Mezcla de individuos de más de una población (por ignorancia), mayor probabilidad de que ciertos genotipos se incluyan en la muestra, muestreo poco representativo.	<ul style="list-style-type: none"> <li>Sección 1 de la brújula</li> <li>Efecto Wahlund</li> <li>Deriva genética</li> </ul>



**LA DERIVA GENÉTICA** es la alteración de la composición genética de las poblaciones debida a factores aleatorios.

Los cambios de frecuencias génicas en las poblaciones, ya sea en una dimensión temporal (a través de las generaciones) o espacial (en la distribución intra- e inter-poblacional de la variabilidad genética), son a la vez causa y efecto de la diversificación biológica. Sus efectos principales son el aumento de la diferenciación inter-poblacional, la pérdida de variabilidad genética y la predominancia de factores estocásticos en el proceso de diversificación de los organismos.

### LA DERIVA GENÉTICA EN EL ESPACIO

Notemos que una de las asunciones fundamentales del teorema de Hardy-Weinberg es que las poblaciones son de tamaño infinito. No obstante, las poblaciones naturales tienen siempre un tamaño finito (aunque éste pueda llegar a ser muy grande). Este detalle es importante ya que, simplemente por el error asociado al muestreo, un conjunto de gametos provenientes de una población finita raramente tendrá las mismas frecuencias génicas que las de la población que los originó. Cuando el cambio de frecuencias génicas es debido exclusivamente al azar, se engloba bajo la definición de deriva genética.

Es importante entender que en todas las poblaciones existen fluctuaciones aleatorias de frecuencias génicas y genotípicas a lo largo de las generaciones. En general, la magnitud de cambio evolutivo asociado a la deriva está inversamente relacionada con el tamaño poblacional ( $N$ ). La probabilidad de fijación de un alelo selectivamente neutral en una población es  $1/2N$ , y la probabilidad de pérdida de ese mismo alelo es de  $1-1/2N$ . Esto significa que **la deriva genética es más efectiva como fuerza diferenciadora en las poblaciones pequeñas, ya que es en ellas donde las desviaciones aleatorias de las frecuencias génicas presentan mayor posibilidad de amplificación**.

La Figura 1 ilustra la magnitud de los cambios dependiendo del tamaño poblacional inicial y de la frecuencia alélica inicial. La deriva genética modifica la composición genética de las poblaciones a través del tiempo y muestra que es muy difícil que las condiciones del equilibrio Hardy-Weinberg se mantengan, especialmente cuando las poblaciones son finitas. Si, además de finitas, las poblaciones son extremadamente pequeñas, la deriva genética puede constituir una fuerza de diferenciación más fuerte que las variables biológicas intrínsecas, que son las que determinan el cambio evolutivo en las poblaciones de gran tamaño (Figura 2).

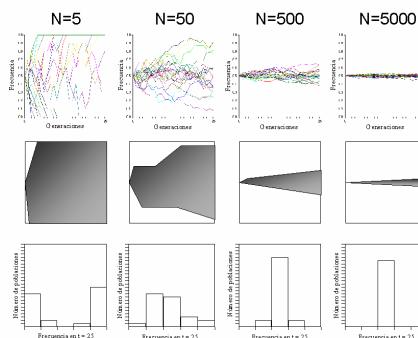


Figura 1. Efecto del tamaño poblacional en la magnitud de las desviaciones causadas por la deriva genética en 20 poblaciones con tamaño poblacional  $N$  a lo largo de 25 generaciones. Las regiones sombreadas muestran el grado de perturbación provocado por cada una de las situaciones simuladas. Todas las simulaciones asumen un locus dialélico con frecuencias  $p=q=0,5$ .



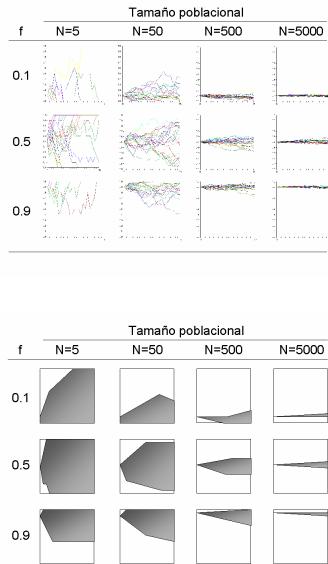


Figura 2. Efecto de la frecuencia alélica en la magnitud de las desviaciones causadas por la deriva genética. N y f son, respectivamente, los tamaños poblacionales y las frecuencias alélicas que se utilizaron para realizar las simulaciones. Las áreas sombreadas muestran el grado de perturbación provocado por cada una de las situaciones simuladas.

## LA DERIVA GENÉTICA EN EL TIEMPO

Cuando las poblaciones son pequeñas, las desviaciones de frecuencias provocadas por la deriva tienden a amplificarse con el paso de las generaciones. Puesto que la mayoría de poblaciones naturales que estudiamos bajo la perspectiva de la genética de poblaciones tiene cierta antigüedad (especialmente si son endémicas), es conveniente saber como influye la deriva en la diferenciación poblacional en el tiempo. En ausencia de otros estímulos como la selección natural, los factores estocásticos pueden adquirir un gran protagonismo en las diferencias genéticas observadas en las poblaciones naturales.

Aunque no es posible predecir cuando un determinado alelo se perderá, la probabilidad de que en un locus dialélico los alelos  $A$  y  $a$  se fijen en una población son iguales a sus frecuencias de partida  $p_0$  y  $q_0$  (Fisher 1930, Wright 1931). Por consiguiente, una proporción  $p_0$  de las poblaciones contendrán solamente el alelo  $A$ , mientras que una proporción  $q_0$  contendrán solamente el alelo  $a$ . De esta manera, es sencillo entender que si cualquiera de los dos alelos está a una frecuencia mucho menor que el otro, entonces su pérdida es más probable (siempre y cuando la deriva genética sea la única fuerza que opera).

En relación a este punto, Watterson (1962) postula que el tiempo medio para la fijación de un alelo ( $\bar{t}$ , también conocido como tiempo medio de absorción) viene dado por la fórmula

$$\bar{t} = -4N(p_0 \log_e p_0 + q_0 \log_e q_0)$$

Aunque desarrollar matemáticamente la formulación que da lugar a esta ecuación queda fuera del objetivo de esta brújula, sí es importante entender que esta expresión nos está diciendo que el tiempo medio de absorción depende de la frecuencia inicial de un alelo y del tamaño de la población donde se encuentra.



Las Figuras 3 a 5 ponen de relevancia algunas de las propiedades generales más importantes de la deriva genética en el tiempo.

1. **La deriva genética no tiene dirección.** A partir de una misma frecuencia de partida, un alelo bajo el influjo de la deriva puede incrementar su frecuencia hasta fijarse, disminuir su frecuencia hasta desaparecer o permanecer en la población a una frecuencia diferente de la de partida. En la Figura 3, este efecto se muestra con un alelo con frecuencia de partida 0.5. Mientras que en b) se manifiesta una dirección clara hacia la fijación (que se alcanza en todos los casos antes de la quinta generación), la distribución de frecuencias alélicas en a) después de 25 generaciones es totalmente aleatoria, con alelos fijados, alelos desaparecidos y alelos a frecuencias intermedias.

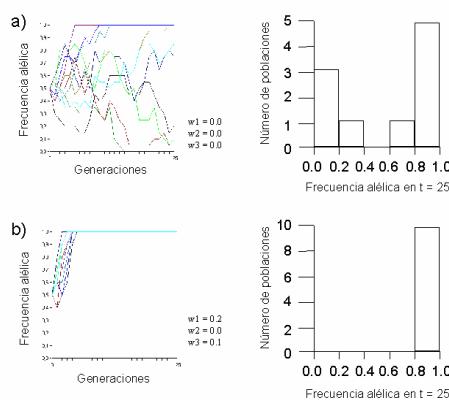


Figura 3. Simulaciones de cambio de frecuencia alélica en un locus influido solamente por la deriva genética (a) o por la acción de fuerzas selectivas (b). En la simulación (a) no se ha asignado peso a ninguno de los tres alelos considerados ( $w_1=w_2=w_3=0$ ). En cambio, en (b) se asignaron pesos arbitrarios. Ambas simulaciones consideraron 10 réplicas poblacionales con dos alelos cada una que partieron de una frecuencia de 0.5.

2. **Los efectos de la deriva genética son acumulativos.** La magnitud de las desviaciones se incrementa con el tiempo hasta que un alelo se fija (alcanza frecuencia 1) o desaparece (alcanza frecuencia 0). Como puede verse en la Figura 1, el proceso de fijación es tanto más rápido cuanto más pequeña es la población sobre la que opera la deriva. Cuando el alelo se ha perdido o se ha fijado, la actuación de la deriva genética ya no es posible a menos que se origine una mutación nueva en la población.

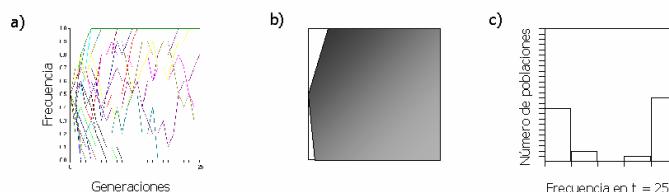


Figura 4. Simulaciones de cambio de frecuencia alélica en poblaciones de cinco individuos en un locus dialélico con frecuencia de partida 0,5 a lo largo de 25 generaciones. En a) se muestra el trayecto de cada población simulada, en b) el área de influencia de las perturbaciones provocadas por la deriva, y en c) la distribución de frecuencias después de las 25 generaciones. La simulación consideró 25 réplicas poblacionales con dos alelos cada una que partieron de una frecuencia de 0,5.



### 3. La deriva causa la pérdida de variación genética dentro de las poblaciones

debido a que las poblaciones tienden a fijarse para un alelo u otro, con la particularidad de que el tiempo de fijación es tanto menor cuanto más cercanas sean las frecuencias alélicas de partida a cero y a uno. La figura 5 ilustra este fenómeno mostrando que solamente en el caso de frecuencia de partida 0.5 existen unas pocas poblaciones donde el alelo no se ha fijado o ha desaparecido. La tasa de pérdida de variabilidad genética por deriva es de  $1/(2N)$  (donde  $N$  es el tamaño poblacional).

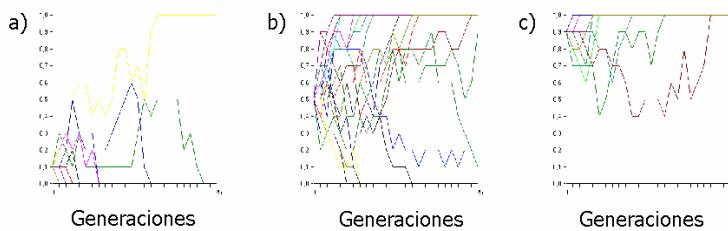


Figura 5. Simulaciones de cambio de frecuencia alélica por deriva partiendo de una frecuencia inicial de 0.1 (a), 0.5 (b) y 0.9 (c). Las simulaciones corresponden la evolución de 10 réplicas poblacionales a lo largo de 25 generaciones.

### 4. Por su naturaleza aleatoria, la deriva genética tiene como efecto un incremento de la diferenciación genética entre poblaciones.

La consanguinidad y la deriva genética suelen ser fenómenos interrelacionados, puesto que la deriva genética resulta en una pérdida de variación que acarrea un incremento en la probabilidad de que dos genes derivados de un ancestro común se junten en un zigoto.

## DERIVA GENÉTICA Y ADAPTACIONES

El trabajo empírico de las últimas décadas ha demostrado que la estructura genética poblacional significativa que exhiben muchos organismos puede fomentar las oportunidades de actuación de la deriva genética.

Aunque las condiciones en que la deriva influye en la evolución de los organismos han sido postuladas teóricamente (Fenster y Galloway 2000), hay solamente unos pocos casos bien documentados donde se demuestre que la deriva ha contribuido a la evolución de adaptaciones. Así pues, tenemos una comprensión limitada de cómo la deriva interactúa con la selección y de si la selección actúa en las interacciones génicas.

Aunque la deriva actúa en todas las poblaciones de organismos, existen dos casos extremos en los cuales su manifestación es particularmente notoria y puede tener implicaciones de gran calado para la diversificación biológica: el efecto fundador y el efecto cuello de botella.



**EL EFECTO FUNDADOR** es la pérdida de variación genética que se produce cuando una nueva población se forma a partir de un número muy pequeño de individuos provenientes de una población más grande

Este concepto fue propuesto por primera vez por Ernst Mayr (1963) para referirse al "establecimiento de una población nueva a partir de unos pocos fundadores ... que poseen solamente una pequeña fracción de la variación genética de la población suministradora". A corto plazo, el impacto de un efecto fundador dependerá de la cantidad de variabilidad que posee la población suministradora, del número de efectivos que intervienen en el efecto fundador y del número de veces que se producen eventos fundadores hacia la misma población receptora. A nivel general, la diferenciación entre la población suministradora y la receptora será máxima cuando la población suministradora sea muy variable, el evento fundador sea único y conste de muy pocos individuos. A largo plazo, los efectos diferenciadores de un evento fundador se amplificarán con el tiempo (por la aparición de mutaciones independientes) y cuanto más diferentes sean las condiciones en la población receptora (por el posible estímulo de la adaptación y la selección).

Obviamente, los efectos fundadores son de máxima importancia en el estudio de la diversificación de organismos insulares. Si unos pocos individuos fundan una nueva población, entonces (suponiendo que no hay extinción) debe de producirse una expansión en unas pocas generaciones. La deriva genética causada por el efecto fundador tiene dos efectos relacionados.

En primer lugar, una reducción de la varianza genética y la heterozigosidad que llevará a una acumulación de consanguinidad y a un incremento de la homozigosidad.

En segundo lugar, un cambio aleatorio en las frecuencias alélicas (incluyendo la pérdida y la fijación de alelos) y en las medias de los caracteres cuantitativos de las poblaciones fundadas que conllevará la divergencia genética respecto de la población suministradora y una reducción de la variabilidad.

Los eventos fundadores pueden causar también desviaciones de las proporciones Hardy-Weinberg y fluctuaciones aleatorias de los desequilibrios de ligamiento, pero ambos efectos son transitorios (Barton 1989). Si diferentes poblaciones poseen un diferente conjunto de alelos, entonces habrá una diferenciación genética que puede conducir al aislamiento.

Para Mayr (1954) la acumulación de consanguinidad en la población fundadora era la consecuencia más importante del efecto fundador, porque incrementa la probabilidad de cambio en los complejos génicos adaptativos. El razonamiento de Mayr (1954) es que, al ser los homocigotos más frecuentes en la población fundadora, estarán mucho más expuestos a la selección, que favorecerá especialmente a aquellos alelos que sean más viables en homozigosis. Este hecho llevó a este autor a postular que el mero cambio del *ambiente genético* puede variar el valor selectivo de un gen, y puede significar a la poste una auténtica revolución genética cuyas consecuencias afecten a la totalidad del genoma (y no solamente a los genes que la desencadenan).

La definición de Mayr establece una diferencia fundamental entre la especiación por efecto fundador y la especiación alopátrica. En la especiación por efecto fundador, el aislamiento conlleva una alteración del ambiente genético que provoca la especiación. Por el contrario, en la especiación alopátrica el aislamiento solamente permite que la especiación ocurra como un sub-producto de los procesos microevolutivos ordinarios (Templeton 1980).



Posteriormente, Carson (1975) amplía la definición de revolución genética al considerar que el genoma puede descomponerse en un sistema cerrado y un sistema abierto, y que la revolución genética solamente incumbría al sistema cerrado. Éste constaría de unos pocos loci (del orden de magnitud de 10) que regulan los procesos fundamentales del desarrollo, la fisiología y la historia vital de los organismos. Según esta teoría, los loci aloenzimáticos pertenecerían al sistema abierto y, por lo tanto, no serían los desencadenantes de las revoluciones genéticas sino solamente agentes pasivos que reciben y manifiestan sus consecuencias.

Más adelante, Templeton (1979) sustituye el término *revolución genética* por el concepto de *transiliencia genética*, que define como el cambio rápido de un complejo multilocus con influencia adaptativa en respuesta a una perturbación súbita del ambiente genético. Templeton (1979) muestra que la transiliencia genética puede conllevar aislamiento pre- y post apareamiento respecto de la población ancestral y que puede por tanto desencadenar la especiación.

Aunque los trabajos de Mayr establecen una conexión clara entre el efecto fundador y la especiación, Templeton (1980) nota que la teoría de Mayr no puede predecir qué efectos fundadores conllevan especiación y cuáles no.

Según Templeton (1980), la transiliencia genética ocurrirá con mayor probabilidad cuando

- (1) el incremento del tamaño efectivo de consanguinidad en la población fundadora ( $\Delta N_{ef}$ ) sea lo suficientemente grande como para crear un cuello de botella selectivo,  
y
- (2) la varianza del tamaño efectivo de la población ancestral [ $N_{ev(A)}$ ], del tamaño efectivo de la población fundadora [ $N_{ev(F)}$ ] y el tamaño de la población fundadora ( $N$ ) son tan grandes como sea posible siempre y cuando se cumpla que el incremento del tamaño efectivo de la consanguinidad en la población fundadora [ $\Delta N_{ef}$ ] sea lo suficientemente grande para que la población fundadora tenga suficiente variación genética como para responder al cuello de botella selectivo.

La mayor probabilidad de revolución genética se da cuando la población suministradora es grande y panmíctica, con tamaño efectivo y tamaño efectivo de consanguinidad grandes. Templeton (1980) también ofrece una tabla en la que lista los atributos de la población fundadora que favorecerán o disminuirán la probabilidad de transiliencia genética. Los parámetros de interés para poblaciones de plantas vasculares se reproducen en la Tabla 4.

ATRIBUTOS RELACIONADOS CON LA TRANSILIENCIA		
SI?		NO?
<input type="checkbox"/>	Alto número de descendientes	<input type="checkbox"/>
<input type="checkbox"/>	Estructura poblacional subdividida	<input type="checkbox"/>
<input type="checkbox"/>	Alto valor reproductivo de los fundadores	<input type="checkbox"/>
<input type="checkbox"/>	Nicho abierto que permite la expansión poblacional	<input type="checkbox"/>
<input type="checkbox"/>	Densidad inicial baja	<input type="checkbox"/>
<input type="checkbox"/>	Apareamiento selectivo	<input type="checkbox"/>
<input type="checkbox"/>	Generaciones solapadas	<input type="checkbox"/>
<input type="checkbox"/>	Número cromosómico alto	<input type="checkbox"/>
<input type="checkbox"/>	Longitud del mapa genómico grande	<input type="checkbox"/>
<input type="checkbox"/>	Pocos supresores de la recombinación	<input type="checkbox"/>

Tabla 4. Factores que pueden estimular (en caso de estar presentes) o dificultar (en caso de no presentarse) la transiliencia de un efecto fundador en poblaciones de plantas vasculares [adaptada a partir de la Tabla 2 de Templeton (1980)]. Cuantos más factores converjan en la columna 'SI', más probable es la transiliencia genética.



**EL EFECTO CUELLO DE BOTELLA** es la variación aleatoria de frecuencias alélicas que se produce en respuesta a una brusca reducción del tamaño poblacional

Un cuello de botella genético es un caso extremo de deriva genética porque la variación genética remanente difícilmente representará a la variación genética que la población poseía antes de experimentar el cuello de botella (excepto si la población presentaba niveles muy bajos de variación genética).

Según Nei (1975) los efectos de un cuello de botella en la variabilidad se aprecian mejor a través de los cambios operados en el valor de la heterozigosidad. En sus simulaciones, Nei (1975) encuentra que la reducción de heterozigosidad asociada a un cuello de botella no depende solamente del tamaño del cuello de botella, sino también de la tasa de crecimiento poblacional después del cuello de botella: a mayor tasa de crecimiento, más rápidamente se recuperará la población de la reducción de la heterozigosidad.

Es sencillo pronosticar que cuando se produce un cuello de botella genético, la disminución en el número de alelos por locus será mucho más significativa que la reducción de la heterozigosidad, ya que cualquier alelo constituye una unidad de diferenciación genética, pero los alelos que primero se perderán serán los menos frecuentes (que contribuyen poco al valor de la heterozigosidad).

Por el contrario, si los tamaños poblacionales originales se restauran, el número medio de alelos por locus incrementará más rápidamente que la heterozigosidad media. Esto es así porque cuando se incorpore un nuevo alelo (ya sea por mutación, selección o migración) a la población "convaleciente", éste estará a baja frecuencia. Por ello, mientras el número de alelos por locus se incrementará en una unidad (lo cual es un incremento bastante significativo), la heterozigosidad asociada a esa población apenas aumentará, puesto que el valor de este parámetro apenas se ve alterado por los alelos poco frecuentes (como son todos los de nueva aparición).

Para ejemplificar este aspecto, imaginemos una población de 100 individuos con tres alelos *a*, *b* y *c*. Si hay 50 copias de *a*, 49 de *b* y 1 de *c*, tendremos que  $p(a) = 0.50$ ,  $p(b) = 0.49$  y  $p(c) = 0.01$ .

Según la fórmula de la heterozigosidad esperada, (ver capítulo de parámetros)

$$H_e = 1 - \sum_{i=1}^n x_i^2$$

, donde  $x_i$  representa la frecuencia de cada alelo detectado. Por lo tanto, la contribución al valor de heterozigosidad esperada del alelo *c* es solamente de  $0.10^2 = 0.0001$  frente a  $0.50^2 = 0.2500$  del alelo *a* y  $0.49^2 = 0.2401$  del alelo *b*. A partir de este ejemplo imaginario, resulta obvio que la desaparición del alelo *c* afectaría muy poco al valor de heterozigosidad esperada ( $H_e$  con el alelo *c* es  $H_e = 0.5098$  y sin el alelo *c* es  $H_e = 0.5099$ ) y, en cambio, sería importante para el valor del número de alelos por locus, que se reduciría en un 33% (pasaría de 3 a 2).

El efecto cuello de botella tiene profundos efectos en la heterozigosidad promedio de la población. Basándose en simulaciones matemáticas basadas en el modelo de pasadera, Chakraborty y Nei (1977) predicen que la restauración del nivel original de heterozigosidad no se producirá hasta que hayan transcurrido  $t = 2N/(4Nv + 1)$  generaciones.

Es conveniente precisar que las reducciones de número de individuos en las poblaciones naturales no siempre van acompañadas de una brusca reducción de su variabilidad genética. Los cuellos de botella poblacionales solamente crean un cuello de botella genético si conducen a una reducción drástica de variabilidad genética. En este caso, la teoría elemental de genética de poblaciones predice que la adaptabilidad de la población y de sus individuos se verá reducida (Lande 1988). Por esta razón, los cuellos de botella genéticos son de gran trascendencia para la biología de la conservación.



**LA PARADOJA DEL TAMAÑO POBLACIONAL** muestra que los niveles de variación genética detectados empíricamente no concuerdan con las predicciones de la teoría neutral

Los tamaños poblacionales de las especies contemporáneas difieren en varios órdenes de magnitud. En consecuencia, la teoría de la genética de poblaciones predice que esas diferencias de tamaños poblacionales deberían reflejarse en los patrones de diversidad genética. Si la mayoría de la variación genética en las poblaciones naturales es neutral (Kimura y Ohta 1971), las fuerzas principales que la modulan son la deriva y la mutación. En esta situación, la deriva elimina variabilidad a una tasa de  $1/2N$  (donde  $N$  es el tamaño poblacional) y la mutación añade variación a una tasa  $2Nu$  (donde  $u$  es la tasa de mutación asociada a los alelos neutrales). Con el tiempo, se alcanzará un estado de equilibrio entre estas dos fuerzas opuestas en el cual la heterozigosidad puede ser descrita como

$$H = (4Nu)/(1 + 4Nu)$$

Es fácil darse cuenta de que esta expresión es extremadamente sensible al tamaño poblacional, de forma que si  $4Nu$  es pequeño, entonces no se espera ninguna heterozigosidad y, si  $4Nu$  es grande, entonces se espera heterozigosidad total.

No obstante, los datos empíricos no muestran ninguna evidencia de esta influencia esperada del tamaño poblacional en los niveles de variación genética. En efecto, el rango de niveles de heterozigosidad de los taxones para los que se dispone de datos solamente va del 5% al 18% (que corresponderían “en teoría” a diferencias de tamaño poblacional de solamente en un orden de magnitud) a pesar de que las diferencias entre los tamaños poblacionales son de varios órdenes de magnitud. Lewontin (1974) fue el primero que se dio cuenta de esta discrepancia entre la teoría y los datos, y la llamó la paradoja del tamaño poblacional. Desde hace un tiempo, también se la conoce como la paradoja de Lewontin.

Según Lewontin (1974), la única explicación a esta paradoja es que la teoría de la deriva genética estocástica está gravemente incompleta y que, en la teoría correcta, la heterozigosidad predecida sería independiente del tamaño poblacional. Esencialmente, falta una explicación acerca de la pérdida de la heterozigosidad esperada.

**SOLUCIONES PROPUESTAS A LA PARADOJA DEL TAMAÑO POBLACIONAL****El autoestop genético (del inglés “genetic hitchhiking”)**

Maynard-Smith y Haigh (1974) reaccionaron a las evidencias presentadas por Lewontin ofreciendo la explicación del autoestop genético, según el cual un alelo neutral (o, en algunos casos, incluso deletéreo) podría expandirse en una población si estuviera ligado a una mutación selectivamente ventajosa. Según los autores de esta teoría, el hitchhiking elimina variabilidad genética más rápido que la deriva cuando el tamaño poblacional aumenta. No obstante, uno de sus puntos débiles es que depende de la asunción de que debe existir un elevado desequilibrio en el ligamiento.

**El reclutamiento genético (del inglés “genetic draft”)**

La teoría del reclutamiento genético (Gillespie 2001, 2004) es una reformulación de la teoría del hitchhiking que usa un modelo estocástico con variables aleatorias para predecir la



ocurrencia de un evento de hitchhiking y para calcular la probabilidad de que un alelo neutral esté ligado a una mutación selectivamente ventajosa.

Esta teoría ofrece el tipo de explicación que pide Lewontin a través del proceso evolutivo denominado reclutamiento genético. El reclutamiento produce una dinámica poblacional que se parece a la inducida por la deriva en que elimina variabilidad genética de las poblaciones (aunque a mayor velocidad). No obstante, una diferencia importante es que, mientras la capacidad de la deriva para eliminar variabilidad genética disminuye a medida que el tamaño poblacional aumenta, el reclutamiento genético no exhibe dependencia respecto del tamaño poblacional. En consecuencia, según esta teoría, el efecto del reclutamiento en la variación genética eclipsa al de la deriva en poblaciones a partir de un cierto tamaño poblacional.

El reclutamiento soluciona la paradoja del tamaño poblacional al eliminar la variación genética de forma más rápida que la deriva. En esta teoría, la expresión para describir la heterozigosidad es

$$H = (4Nu)/(1 + 2N\mu E\{\gamma^2\})$$

, donde

$\mu = 4Nus$  ( $s$  es el coeficiente de selección)

$\gamma$  = frecuencia final del alelo reclutado.

Cuando  $E\{\gamma^2\} > 1/2N$ , el reclutamiento será más importante que la deriva para explicar la disminución de la heterozigosidad. Esto sucederá cuando el tamaño poblacional  $N$  es lo suficientemente grande (por ejemplo  $N > 10.000$ ).

Según ha expresado John Gillespie en varios foros científicos, la dinámica de los alelos raros es tan radicalmente diferente de la de los alelos comunes que se justificaría una denominación para la dinámica estocástica de los alelos comunes y otra para la de los alelos raros.

Por ejemplo, la varianza en el número de copias ( $X$ ) de un alelo raro con frecuencia  $p$  en un locus dialélico de un organismo diploide es

$$\text{Var}\{X\} = 2Npq$$

Si hay  $j$  copias de este alelo,  $p = j/2N$  debe ser muy próximo a cero y  $q \approx 1$ , y entonces

$$\text{Var}\{X\} \approx 2N \times (j/2N) = j,$$

que es independiente de  $N$ .



**EL EFECTO WAHLUND** es la reducción de la heterozigosidad observada que se produce cuando, por ignorancia o por error de muestreo, consideramos poblaciones independientes como si fueran una sola población.

No siempre tenemos conocimiento de si una población en estudio está subdividida y, a veces, incluso desconocemos si los organismos muestreados pertenecen a una o varias poblaciones. De hecho, si existe subdivisión poblacional pero el grado de divergencia entre las subpoblaciones es pequeño, la distorsión producida por efecto Wahlund será mínima. No obstante, a partir de ciertos niveles de divergencia genética entre sub-poblaciones, el hecho de no tener en cuenta la subdivisión tiene como consecuencia el incumplimiento del equilibrio HW por un exceso de homocigotos (o un defecto de heterocigotos) observados que parece causado por consanguinidad.

El efecto Wahlund no es un producto de la deriva, sino que está causado por nuestra ignorancia de la estructura genética espacial de las poblaciones naturales y de los niveles de flujo génico entre ellas. Su consecuencia principal es la detección de un exceso de homocigotos (o un defecto de heterocigotos). Si lo consideramos en este apartado es porque algunos de sus efectos pueden confundirse con los de la deriva genética. El impacto del efecto Wahlund será tanto mayor cuanto más diferenciadas genéticamente estén las subpoblaciones.

Si muestreamos en poblaciones que han divergido suficientemente por la ausencia de flujo génico entre ellas, observaremos un defecto de heterocigotos. Mientras la heterozigosidad esperada será elevada y reflejará la presencia de numerosos alelos, la mayoría de esos heterocigotos esperados nunca llegará a observarse porque la ausencia de flujo génico entre las subpoblaciones que hemos unido artificialmente impedirá su formación. Por ello, la heterozigosidad observada será significativamente menor que la esperada.

La noción de efecto Wahlund se percibe de forma más intuitiva imaginando que cada subpoblación está fijada para un alelo diferente y que no existe intercambio genético entre estas subpoblaciones. Si, por ignorancia, muestreamos todas las subpoblaciones como si fueran una sola subpoblación, observaremos que la diferencia entre la heterozigosidad esperada y la observada es significativa. Dicho de otra manera, observaremos un exceso de homocigotos, porque habrá heterocigotos esperados que nunca se observarán por la ausencia de intercambio genético.

Obviamente, cuando todas las subpoblaciones tienen exactamente los mismos alelos y las mismas frecuencias génicas, el efecto Wahlund es indetectable, aunque haya subpoblaciones "críticas" que no matengen intercambio genético.

Para un locus dialélico con frecuencias  $p$  y  $(1-p)$ , la proporción de heterocigotos observados en todas las subpoblaciones puede expresarse como la heterozigosidad media

$$H_S = \frac{1}{d} \sum_{k=1}^d H_{Sk}$$

donde  $d$  es el número de subpoblaciones y  $H_{Sk}$  es la heterozigosidad en la población  $k$ .

Si existe equilibrio HW en el interior de cada subpoblación, entonces  $H_{Sk} = 2p_k(1-p_k)$ ; luego



$$H_S = \frac{1}{d} \sum_{k=1}^d 2p_k(1-p_k) = \frac{1}{d} \sum_{k=1}^d (2p_k - 2p_k^2) = 2 \left( \frac{1}{d} \sum_{k=1}^d p_k - p_k^2 \right) = 2 \left( \bar{p} - \frac{1}{d} \sum_{k=1}^d p_k^2 \right) = 2 \left( \bar{p} - \bar{p}^2 \right)$$

Pero como  $\text{var}(X) = E(X^2) - E(X)^2$ , entonces

$$H_S = 2 \left( \bar{p} - \text{var}(p) - \bar{p}^2 \right) = 2\bar{p}(1-\bar{p}) - 2\text{var}(p)$$

Vemos pues que, si ignoramos la presencia de subdivisiones, observaremos un déficit de heterocigotos (o, lo que es lo mismo, un exceso de homocigotos) en una proporción igual a  $2 \text{var}(p)$ . Por la definición de varianza, este efecto será tanto más pronunciado cuanto más diferenciadas estén las subpoblaciones [la fuente de estas formulaciones puede consultarse en Elseth y Baumgardner (1981)].

El fenómeno contrario al efecto Wahlund consistiría en la llamada rotura de aislamiento ("isolate breaking"), que se produce cuando poblaciones aisladas contactan y se reproducen, después de lo cual puede producirse un exceso temporal de heterocigotos.



**EL TAMAÑO POBLACIONAL EFECTIVO** es el número de individuos reproductores que son responsables de la composición genética de la generación siguiente.

En cualquier población natural de casi cualquier tipo de organismos (quizás los virus serían la única excepción) hay individuos que mueren jóvenes sin alcanzar la edad mínima de reproducción, individuos que se reproducen normalmente, individuos que se reproducen más de lo normal e individuos que mueren a edad avanzada sin haberse reproducido.

Es fácil ver que, puesto que no todos los individuos de una población se reproducen con el mismo éxito, el tamaño poblacional efectivo será siempre menor o (solamente en casos muy excepcionales) igual al tamaño censal. Las condiciones en las cuales el tamaño poblacional efectivo se aproximaría más al tamaño real se darían cuando la proporción de cada sexo fuera igual, el tamaño poblacional constantemente elevado y cada individuo produjera el mismo número de descendientes.

Por estos motivos, resulta obvio que el concepto de tamaño poblacional efectivo y su estimación numérica son de importancia central en genética de poblaciones. Diferentes organismos (o los mismos organismos en situaciones diferentes) presentarán distribuciones de frecuencias alélicas también diferentes. Por poner dos ejemplos contrapuestos, un mismo tamaño poblacional real  $N$  no será igualmente efectivo para la persistencia de un alelo en una población con apareamiento aleatorio que en otra con alto índice de autogamia y reproducción vegetativa.

Estadísticamente, el tamaño poblacional efectivo ( $N_e$ ) cuantifica el tamaño de la población ideal en el que la varianza de la frecuencia alélica se mantendrá constante. A nivel estrictamente numérico, puede también definirse como el tamaño de una población ideal que nos daría el mismo valor para alguna propiedad cuantificable interesante que la población real que estamos analizando.

El tamaño poblacional efectivo es extremadamente importante para la conservación, ya que influye en la tasa de pérdida de heterocigosidad por generación.  
En una generación dada, la proporción de variación genética que subsiste es

$$P = 1 - \frac{1}{2N_e}$$

donde  $N_e$  es el tamaño poblacional efectivo.

Por lo tanto, si el tamaño poblacional efectivo es muy pequeño, la proporción de variación genética que se perderá en una sola generación será muy grande. En contraposición, tamaños poblacionales muy grandes no llevarán asociado un decremento sustancial de variación genética por generación.

Para una buena comprensión de la estimación del tamaño poblacional efectivo se hace necesario comprender la noción de varianza de la frecuencia alélica.



## VARIANZA DE LA FRECUENCIA ALÉLICA [a partir de Gale (1990)]

---

La mayoría de simulaciones se han desarrollado para el caso de un locus dialélico en una población infinita donde el cambio de frecuencias alélicas es solamente atribuible al azar. Para no entretenernos en la gran variedad de situaciones que se han estudiado a través de modelos matemáticos (lo cual no es el objeto de este libro), diremos solamente que existe un amplio abanico de fórmulas para cuantificar la varianza de la frecuencia alélica ( $V(p')$ ). Cuál de ellas es la de uso más apropiado dependerá del tipo de planta (¿monoica?, ¿dioica?, ¿anual?, ¿perenne?) y de otras circunstancias biológicas básicamente asociadas a su biología reproductiva (¿es el apareamiento aleatorio?, ¿se solapan las generaciones?, ¿hay reproducción clonal?).

Dependiendo de la concurrencia de unas u otras circunstancias biológicas, la distribución de las frecuencias alélicas variará, y con ella, la estimación del tamaño poblacional efectivo. No obstante, dentro de esta heterogeneidad, estas diferentes formulaciones comparten propiedades comunes, y es a éstas a las que nos referiremos para introducir el concepto de tamaño poblacional efectivo.

$V(p'/p)$  es en general inversamente proporcional al tamaño poblacional  $y$ , por tanto, será en principio muy pequeño (a menos que el tamaño poblacional no sea extremadamente pequeño). Por consiguiente, las desviaciones grandes de  $p'$  del valor esperado de  $E(p'/p)$  son, por norma general, improbables.

Todas las distribuciones condicionales de frecuencias alélicas tienen la misma media  $y$ , de manera bastante aproximada, los mismos momentos estadísticos. Así pues, la diferencia entre las diferentes distribuciones reside en la diferencia entre sus varianzas. Para poblaciones con apareamiento aleatorio y fecundidad desigual de alelos parentales, la varianza puede obtenerse tomando la llamada "varianza de Wright-Fisher"

$$V(p') = \frac{pq}{2N}$$

y reemplazando  $N$  por una constante, que notaremos  $N_e$  y que se conoce como el tamaño poblacional de varianza efectiva o, más comúnmente, tamaño poblacional efectivo.

El concepto de tamaño poblacional efectivo se debe a Wright (1931). Su utilidad es que permite, para muchas situaciones biológicas diferentes, calcular un valor de  $N_e$  tal que la varianza de la distribución de frecuencias alélicas sea siempre

$$V(p') = \frac{pq}{2N_e}$$

Es decir, es el valor para el cual la varianza de la distribución de frecuencias alélicas es la misma. El valor de  $N_e$  adquiere expresiones diferentes según la situación biológica de los organismos. Por poner solamente unos ejemplos ilustrativos,

Cuando hay desigualdades de fecundidad potencial debidas a causas genéticas, entonces, en la generación (t+1)



$$N_e = \frac{NEk}{1 + (1 + 3h^2)V(k) / Ek} \quad (\text{Nei y Murata 1966})$$

donde  $h^2$  es el coeficiente de heredabilidad asociado a la fecundidad.

Cuando en una población de tamaño constante  $N$  la proporción de sexos es desigual, entonces

$$N_e = \frac{4N_1N_2}{N}, \quad (\text{Wright 1931, 1939; Ewens 1979}),$$

donde  $N_1$  es el número de individuos masculinos y  $N_2$  el número de individuos femeninos.

En términos generales,  $N_e$  es siempre inferior a  $N$  y será tanto menor cuanto más desigual sea la proporción de sexos.

Cuando existe autofecundación,

$$N_e = \frac{N}{1 + s / (2 - s)} = N \frac{2 - s}{2} \quad (\text{Wright 1969}),$$

donde  $s$  es la proporción de semillas provenientes de autofecundación.

Cuando tenemos grados diferentes de autofecundación y fecundación cruzada en la misma población, entonces

$$N_e = \frac{NEk}{(s^2 / Ek)(1 + f) + (1 - f)} \quad (\text{Crow y Kimura 1970}),$$

donde  $N$  y  $f$  son los valores de estos parámetros en la población parental y  $s^2 = \frac{NV(k)}{(N - 1)}$ .



**EL MODELO DE ESTRUCTURA POBLACIONAL EN ISLA** describe la situación ideal de un conjunto de poblaciones del mismo tamaño en que cualquier población puede intercambiar igual número de migrantes por generación con cualquier otra población.

Asunciones:

- No hay selección
- No hay mutación
- Todas las poblaciones tienen el mismo tamaño ( $N$ ) y cada una de ellas intercambia un número  $m$  de estos individuos con las otras poblaciones (esto es, cada una de las poblaciones contribuye con  $m$  individuos al conjunto de migrantes)
- En todas las poblaciones,  $m$  de sus individuos llegan como migrantes por cada generación
- No hay aislamiento por distancia

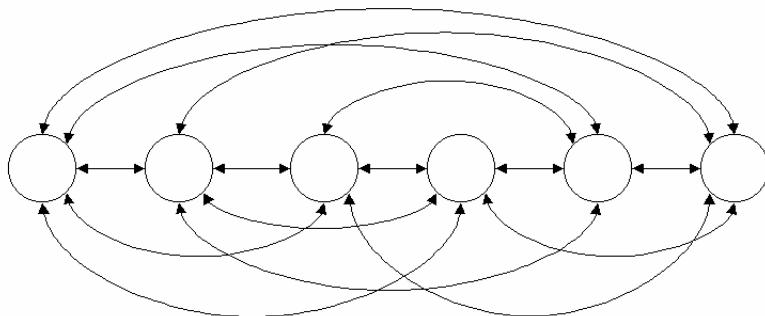


Figura 7. Esquematización del modelo de estructura poblacional en islas para el caso de una sola dimensión. Los círculos indican poblaciones (todas de tamaño  $N$ ), y las flechas indican todas las direcciones posibles de intercambio de migrantes entre cada una de las poblaciones.

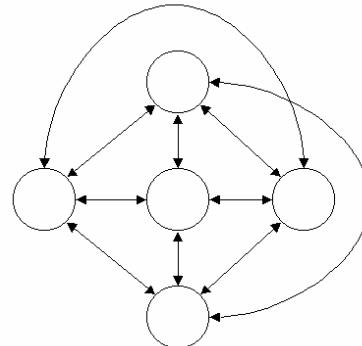


Figura 8. Esquematización del modelo de estructura poblacional por isla en el caso de dos dimensiones. Los círculos indican poblaciones (todas de tamaño  $N$ ), y las flechas indican todas las direcciones posibles de intercambio de migrantes entre cada una de las poblaciones.

No obstante, la naturaleza no responde al modelo de isla: las poblaciones son muy variables en tamaño tanto en el espacio como en el tiempo, las tasas de migración son también variables dependiendo de la población (y en la misma población dependiendo del momento), la mayoría de las veces la dispersión está estrechamente relacionada con la distancia, hay continuas extinciones y recolonizaciones, hay inmigración que afecta a las frecuencias alélicas, casi siempre hay selección y (dependiendo de la región del DNA considerada), la mutación puede ser significativa (p. e., en los microsatélites). A veces ni tan siquiera el modelo responde bien ante sí mismo en algunas situaciones (ver Whitlock y McCaulay 1999 y el capítulo sobre flujo génico para más detalles).



**EL MODELO DE ESTRUCTURA POBLACIONAL EN PASADERA** describe la situación ideal de un conjunto de poblaciones del mismo tamaño en que los migrantes solo se intercambian con las poblaciones más cercanas.

Wright (1943) propuso un modelo de estructura poblacional en que la población se distribuye de manera uniforme en un territorio grande, pero en que los padres de cualquier individuo provienen de una pequeña región adyacente. Puesto que el gran tamaño de la población impide que ciertos individuos alejados puedan ser los padres de la generación siguiente, este modelo se denomina de aislamiento por distancia.

En las poblaciones naturales, los individuos se distribuyen con frecuencia de manera discontinua, formando colonias que intercambian inmigrantes según su distancia de separación. Para analizar este tipo de situaciones, Kimura (1953) propuso el modelo de estructura poblacional en pasadera (del inglés "stepping stone").

### Asunciones

- Un solo locus dialélico
- Tiempo generacional discreto
- La frecuencia alélica cambia por mutación migración y deriva
- No hay selección
- Hay aislamiento por distancia
- El efecto de la dispersión a larga distancia se asume análogo al de la mutación

En su vertiente unidimensional, este modelo considera un entramado infinito de poblaciones en el cual cada generación un individuo puede migrar como máximo un paso en cada sentido mientras esa migración se restrinja exclusivamente a colonias adyacentes.

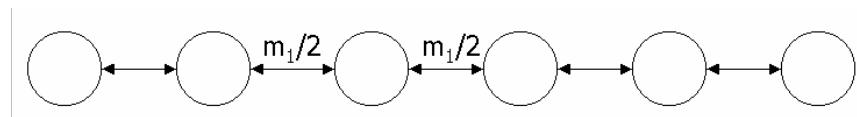


Figura 9. Representación esquemática del modelo de estructura poblacional en pasadera para el caso de una sola dimensión (adaptado de Kimura y Weiss 1964). Los círculos representan poblaciones,  $m_1$  es la tasa de migración a colonias adyacentes por generación, y  $m_1/2$  es la proporción de individuos intercambiados entre colonias adyacentes cada generación.

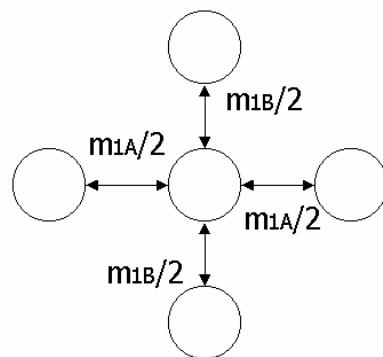


Figura 10. Representación esquemática del modelo de estructura poblacional en pasadera para el caso de dos dimensiones (adaptado de Kimura y Weiss 1964). Los círculos representan poblaciones,  $m_{1A}$  y  $m_{1B}$  representan las tasas de migración a colonias adyacentes en sentido horizontal y vertical (respectivamente) por generación, y  $m_{1A}/2$   $m_{1B}/2$  son las proporciones de individuos intercambiados entre colonias adyacentes en dirección horizontal y vertical (respectivamente) cada generación.



Aunque Kimura y Weiss (1964) modelan la pasadera para tasas de migración de un paso por generación, los resultados pueden extenderse a más de un paso por generación sustituyendo  $m_1$  por la varianza de la distancia de migración por generación. En el caso de una sola dimensión, si  $m_j/2$  es la proporción de individuos que intercambian dos colonias que están separadas por  $j$  pasos, entonces

$$\sigma_j^2 = \sum_{j=1}^{\infty} j^2 m_j$$

de manera que el coeficiente de correlación de las frecuencias génicas entre dos colonias separadas por  $k$  pasos viene dado por

$$r(k) = e^{-\sqrt{\frac{2m_{\infty}}{\sigma_m}}k} \quad (\text{Kimura y Weiss 1964})$$

En el caso de dos dimensiones, si  $m_{ij}/2$  es la proporción de individuos intercambiados por dos colonias, entonces las varianzas  $\sigma_{mj}^2$  se usarían en lugar de cada una de las  $m_{ij}$  para calcular los coeficientes de correlación, con

$$\sigma_{mA}^2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} i^2 m_{ij} \quad y \quad \sigma_{mB}^2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} j^2 m_{ij}$$

Según los cálculos de Kimura y Weiss (1964), el coeficiente de correlación entre frecuencias alélicas decae con la distancia, especialmente cuando esta es grande y de manera mucho más rápida cuanto mayor es el número de dimensiones del modelo de pasadera.

El paso del modelo de pasadera al modelo de población continua se obtiene cuando la distancia entre los grupos adyacentes tiende a cero, de forma que el modelo de poblaciones continuas se obtiene sustituyendo en el modelo de pasadera  $\sigma^2$  por  $m_1$  y  $(N_e/N)\delta$  por  $N_e$ , donde

$\sigma^2$  es la varianza de la distancia de migración por individuo por generación,  $N_e/N$  es el cociente entre el tamaño poblacional efectivo y el número de individuos real por población, y  $\delta$  es la densidad de individuos.



## 2.2. CÁLCULOS Y TESTS MÁS USADOS EN GENÉTICA DE POBLACIONES

---

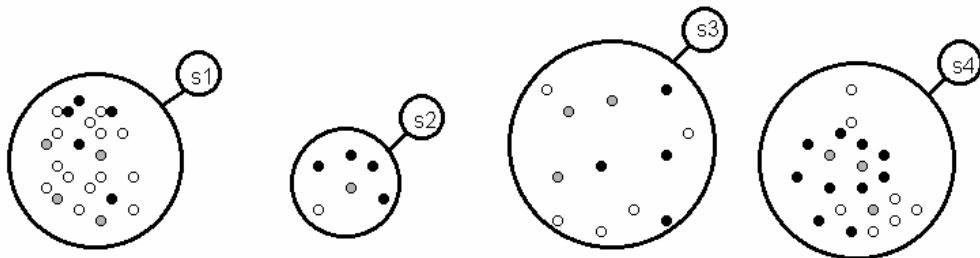


Figura 2.1. Esquematización de un sistema de 4 poblaciones (codificadas s1, s2, s3 y s4) consideradas para ejemplificar los cálculos de las frecuencias génicas, la heterozigosidad observada, la heterozigosidad esperada y los parámetros asociados a la estructura genética poblacional según Wright (1951) y Nei (1973). Todos los individuos (representados por círculos) se consideran diploides. La población S1 consta de 23 individuos, s2 de 6, s3 de 12 y s4 de 20. Los círculos negros simbolizan individuos con genotipo  $AA$ , los blancos el  $aa$  y los grises el  $Aa$ .



## 2.2.1 DESCRIPTORES DE LA ESTRUCTURA GENÉTICA DENTRO DE LAS POBLACIONES

### NÚMERO DE ALELOS POR LOCUS ( $A$ )

$$A = a_I \quad \forall I (I = 1, \dots, L) \quad (A \geq 1)$$

$a_I$  es el número de alelos detectados en el locus  $I$ . En el ejemplo que se ilustra en la Figura 2.1, este parámetro tendría el valor  $A = 2$  para cada una de las poblaciones consideradas.

### NÚMERO MEDIO DE ALELOS POR LOCUS ( $\bar{A}$ )

$$\bar{A} = \frac{1}{L} \sum_{I=1}^L a_I \quad 1 \leq \bar{A} \leq \max a_I$$

donde  $L$  es el número total de loci analizados en la población considerada. Para el ejemplo de la Figura 2.1, su valor sería el mismo que el de  $A$ , puesto que solamente se considera un locus.

### NÚMERO EFECTIVO DE ALELOS POR LOCUS ( $A_e$ )

El estimador original de este parámetro fue propuesto por Kimura y Crow (1964) a través de la fórmula

$$A_e = 1 + 4N_e u$$

, donde  $N_e$  es el tamaño poblacional efectivo y  $u$  la tasa de mutación.

Más adelante, Ohta y Kimura (1973a, b) propusieron otro estimador que tiene en cuenta el hecho de que no todos los alelos presentes en una muestra poblacional son detectados mediante electroforesis. Este estimador adquiere la forma

$$A_e = \sqrt{1 + 8N_e u}$$

La reformulación obedece al razonamiento de que los aloenzimas solamente nos revelan una fracción del número total de cambios de aminoácidos presentes en la población, ya que las alteraciones de aminoácidos con carga opuesta se anulan en términos de movilidad en el gel.

No obstante, parece que la visión de que un cambio positivo y otro negativo se cancelan mutuamente conduciendo al alelo al estado ancestral es extremadamente limitada y representa solamente una alternativa a las múltiples posibilidades bioquímicas. Según Johnson (1974), las dos formulaciones mencionadas representarían los límites del espectro de posibilidades de identificación alélica. En el primer caso, los cambios de carga opuesta no se anulan, y todos los alelos son detectados; en el segundo caso, todas las cargas opuestas se anulan, y solamente un número mínimo de alelos pueden identificarse. Por lo tanto, según este autor, la ausencia de una formulación satisfactoria para estimar  $A_e$  lleva a la conclusión de que este parámetro ostentará un valor intermedio entre  $1 + 4N_e u$  y  $\sqrt{1 + 8N_e u}$  y que, cuanto mayor sea la diferencia genética entre las poblaciones, mayor será la diferencia entre estos dos estimadores de  $A_e$ .



**LA DIVERSIDAD GENOTÍPICA** (también llamada riqueza genotípica) es la proporción de genotipos diferentes que hay en una muestra poblacional

La riqueza genotípica ( $R$ , Dorken y Eckert 2001) se estima con la fórmula

$$R = (G-1)/(N-1)$$

, donde  $G$  es el número de genotipos diferentes y  $N$  el tamaño muestral.

**LA PROBABILIDAD DE PÉRDIDA ALÉLICA** es la probabilidad de que una muestra de  $N$  poblaciones no contenga un alelo con frecuencia poblacional  $p$

Esta probabilidad se estima mediante la expresión

$$L = (1 - p)^{2N}$$

solamente para los alelos que satisfagan un criterio adecuado de rareza (Bengtsson et al. 1995). Este proceder selectivo de partida es lógico, ya que si se incluyen también en los cálculos los alelos más ubicuos, su bajísima probabilidad de pérdida enmascara a las estimaciones correspondientes a los alelos raros, que son los elementos que nos interesan a este nivel de análisis.

Aunque Bengtsson et al. (1995) recurren a las técnicas estadísticas multivariantes para seleccionar los alelos a analizar mediante esta metodología, es más práctico (y seguramente igual de efectivo) establecer un umbral de distribución y de frecuencia, como por ejemplo seleccionar los alelos que (1) tengan una frecuencia media menor que 0.5, y (2) estén presentes como máximo en el 50% de las poblaciones consideradas (Vilches et al. 2002, Caujapé-Castells y Pedrola-Monfort 2004, Sánchez et al., en prensa). La Tabla 1 ilustra este proceso de selección, que está implementado en el programa informático Transformer-3 (Caujapé-Castells y Baccarani-Rosas 2004) con la denominación "Viera y Clavijo".

Una vez culminado el proceso de selección de alelos, deben calcularse las probabilidades de pérdida observada ( $L_o$ ) y esperada ( $L_e$ ) para los alelos seleccionados (deberían ser un mínimo de 10 para garantizar la robustez de los datos).

La probabilidad de pérdida observada,  $L_o$ , se obtiene sustituyendo en la fórmula la frecuencia media de cada alelo en todas las poblaciones consideradas ( $p$ ) y el número de poblaciones donde ha sido detectado ( $N$ ). En este caso, pues, tanto  $p$  como  $N$  son variables para cada uno de los alelos analizados.

La probabilidad de pérdida esperada,  $L_e$ , es la probabilidad de pérdida que tendría un alelo si ostentara su frecuencia media  $p$  en todas y cada una de las poblaciones consideradas. Por lo tanto, para estimarla se fija el valor de  $N$  como constante (igual al número total de poblaciones consideradas), y la única variable es el valor de  $p$  para cada uno de los alelos incluidos en el análisis.

Los valores de  $L_o$  y  $L_e$  así obtenidos se utilizan posteriormente para hacer una regresión lineal donde las frecuencias medias de cada alelo incluido en el análisis se representan en el eje x y  $-\log L_o$  and  $-\log L_e$  se representan en el eje y (ver figura 2). De acuerdo con Bengtsson et al. (1995), puede obtenerse un valor de representatividad ( $R$ ) de una población de ese grupo en relación a los alelos raros detectados dividiendo la pendiente de la línea de regresión observada (basada en los valores de  $L_o$ ) por la pendiente de la línea de regresión esperada (basada en los valores de  $L_e$ ). Ver la sección III de la brújula para una aplicación de la probabilidad de pérdida en conservación de semillas.



**LA HETEROZIGOSIDAD OBSERVADA ( $H_O$ )** es la proporción de individuos heterocigotos en una muestra poblacional

Para una población, se calcula como la media inter-loci de la frecuencia genotípica de los individuos heterocigotos mediante la expresión

$$H_O = \frac{1}{L} \sum_{i=1}^L \frac{(N_{\text{heterocigotos}})_i}{N_i} \quad (0 \leq H_O \leq 1)$$

donde

$L$  es el número de loci analizados en la población objeto de estudio,  
 $(N_{\text{heterocigotos}})_i$  es el número absoluto de individuos heterocigotos para el locus  $i$ , y  
 $N_i$  es el número total de individuos analizados para el locus  $i$ .

La heterozigosidad observada representa la probabilidad de que un individuo escogido al azar de la población sea heterocigoto. Es un parámetro simple porque su estimación requiere conocer únicamente los genotipos de los individuos analizados. Para el único locus representado en el ejemplo de la Figura 2.1, las heterozigosidades observadas por población son:

$$H_O(s1) = \frac{4}{23} = 0.174$$

$$H_O(s2) = \frac{1}{6} = 0.167$$

$$H_O(s3) = \frac{3}{12} = 0.250$$

$$H_O(s4) = \frac{3}{20} = 0.150$$

Estos valores se hallan en la columna ( $H_O$ ) de la Tabla 1. Si dispusiéramos de datos para más de un locus con más de dos alelos (como es la norma habitual), el valor operativo de  $A$  sería el promedio de este parámetro teniendo en cuenta todos los loci analizados.



**LA HETEROZIGOSIDAD ESPERADA ( $H_e$ )** estima la probabilidad de que dos alelos extraídos al azar del conjunto de genes de una población sean diferentes.

Este parámetro es una función de las frecuencias génicas detectadas que, para poblaciones grandes ( $>50$  individuos) tiene la expresión

$$H_e = \frac{1}{L} \sum_{l=1}^L \left( 1 - \sum_{j=1}^n p_{jl}^2 \right) \quad (0 \leq H_e \leq 1)$$

donde  $p_{jl}$  es la frecuencia del alelo  $j$  en el locus  $l$  y  $L$  es el número total de loci analizados. Si el número de individuos de la población es  $<50$  (como es el caso del ejemplo de la Figura 1), entonces el estimador no sesgado de la heterozigosidad esperada es

$$\hat{H}_e = \frac{1}{L} \left[ \frac{2N}{2N-1} \sum_{l=1}^L \left( 1 - \sum_{j=1}^n p_{jl}^2 \right) \right] \quad (0 \leq \hat{H}_e \leq 1)$$

donde  $N$  es el tamaño de la población (Hedrick 2000). Para las cuatro subpoblaciones representadas en la Figura 2.1,

$(H_e)_{s1} = 0.420$	$(\hat{H}_e)_{s1} = 0.429$
$(H_e)_{s2} = 0.375$	$(\hat{H}_e)_{s2} = 0.409$
$(H_e)_{s3} = 0.497$	$(\hat{H}_e)_{s3} = 0.519$
$(H_e)_{s4} = 0.490$	$(\hat{H}_e)_{s4} = 0.502$

Este parámetro se interpreta, para cada locus analizado, como la probabilidad de muestrear dos genes diferentes en un muestreo con reemplazamiento del total de genes en una población. Se le denomina también “diversidad génica”, aunque esta denominación se hace desaconsejable porque puede inducir a malinterpretar que estamos refiriéndonos a variación genética “en general” y no solo al valor de  $H_e$ .

Es fácil imaginar que las estimaciones de la heterozigosidad media varían de organismo a organismo. Chakraborty y Nei (1977) estimaron que  $0 \leq H_e \leq 0,3$  y que no parecía haber ningún organismo bisexual que mostrara valores superiores a  $H_e = 0,31$  (Nei 1975, Selander 1976), a pesar de las grandes variaciones en el tamaño poblacional de diferentes organismos. La existencia de un límite superior para la heterozigosidad hizo que muchos autores cuestionaran la validez de la hipótesis de mutación neutral formulada por Kimura (1983): si las mutaciones no tienen valor selectivo (como arguye esta hipótesis), entonces no tiene porqué existir un límite superior de variación genética. Para explicar esta paradoja, Ohta (1973, 1974) propuso la hipótesis de que la mayoría de mutaciones no eran neutrales, sino ligeramente deletéreas, lo cual limitaba el valor de la heterozigosidad. No obstante, Chakraborty y Nei (1977) argumentan que el (hipotético) límite superior de la heterozigosidad puede también explicarse por las fluctuaciones del tamaño poblacional a lo largo del proceso evolutivo: si una población experimenta cuellos de botella ocasionalmente, la heterozigosidad media nunca será elevada.

$H_e$  ofrece una medida indirecta de heterogeneidad genética que se ve menos afectada por el error de muestreo que el número de alelos por locus ( $A$ ) porque su valor está mucho menos influenciado por las frecuencias de los alelos raros (Nei 1983).



**EL INDICE DE FIJACIÓN** cuantifica la disminución de la heterozigosidad en una población natural a consecuencia de apareamientos consanguíneos

Wright (1933) definió el índice de fijación ( $F_{IS}$ ) según la expresión

$$F_{IS} = 1 - \frac{H_O}{H_e}$$

o bien, según vimos en el apartado de subdivisión poblacional, como

$$F_{IS} = 1 - \frac{H_I}{H_S}$$

en todos los casos,  $-1 \leq F \leq 1$ , de manera que, cuando existe un exceso de heterocigotos,

$$H_O > 2pq \Rightarrow \frac{H_O}{H_e} > 1 \Rightarrow F_{IS} < 0$$

cuando la población cumple las proporciones de Hardy-Weinberg,

$$H_O = 2pq = H_e \Rightarrow \frac{H_O}{H_e} = 1 \Rightarrow F_{IS} = 0$$

cuando existe un defecto de heterocigotos,

$$H_O < 2pq \Rightarrow \frac{H_O}{H_e} < 1 \Rightarrow F_{IS} > 0$$

y, finalmente, cuando no hay heterocigotos en la población,

$$H_O = 0 \Rightarrow \frac{H_O}{H_e} = 0 \Rightarrow F_{IS} = 1$$

Por tanto, es claro que valores positivos de  $F_{IS}$  indican la posible existencia de consanguinidad, que será tanto mayor cuanto más se aproxime este valor a uno. La interpretación de consanguinidad a partir solamente del valor de  $F_{IS}$  requiere precaución (ver el apartado de deriva genética).



**LA TASA DE REPRODUCCIÓN CRUZADA** es una transformación simple del valor del índice de fijación que estima el grado de intercambio genético intra-poblacional.

El movimiento de genes dentro del espacio poblacional se relaciona sobre todo con el sistema reproductivo de los organismos y presenta implicaciones para la tasa de consanguinidad en una población natural. En la mayoría de poblaciones naturales no existe panmixia; no obstante, cuando la incidencia de apareamientos consanguíneos es elevada, entonces las condiciones del equilibrio HW se alteran de manera que las proporciones de homocigotos observadas son mayores que las proporciones esperadas. Este fenómeno, cuya explicación es la misma que la del efecto Wahlund, se produciría aunque todos los apareamientos consanguíneos se dieran entre individuos heterocigotos, ya que en este caso (asumiendo ausencia de selección) la proporción de heterocigotos esperada en la siguiente generación siempre sería la mitad de la existente en la generación parental.

Weir (1990) define el índice de alogamia o tasa de outcrossing ( $t$ ) a partir de la formulación de Wright del índice de consanguinidad. La expresión matemática de este parámetro es

$$t = \frac{1 - F_{IS}}{1 + F_{IS}} \quad 0 < t < 1$$

, donde  $F$  es el índice de fijación.

Si todos los individuos son homocigotos, entonces  $F_{IS} = 1 \rightarrow \frac{1 - F_{IS}}{1 + F_{IS}} = 0 \rightarrow t = 0$ .

Si hay apareamiento aleatorio, entonces  $F_{IS} = 0 \rightarrow t = 1$  (este es el valor que indica máxima alogamia).

Si hay exceso de heterocigotos, entonces  $F_{IS} < 0 \rightarrow \frac{1 - F_{IS}}{1 + F_{IS}} > 1 \rightarrow t > 1$ .

Por tanto, la interpretación de este índice es que una población se aproxima al apareamiento aleatorio a medida que el valor de  $t$  se acerca a 1.



**LA CONSANGUINIDAD** es la probabilidad de que los dos genes de un individuo diploide sean idénticos por descendencia

También puede ser definida como la proporción probable de los loci de un individuo que contienen genes que son idénticos por descendencia.

La consanguinidad puede medirse por la reducción de heterozigosidad según la fórmula

$$F = \frac{(H_o - H_e)}{H_o}$$

donde, como vimos en otro apartado de la brújula,  $H_o$  es la heterozigosidad observada.

En presencia de consanguinidad, las frecuencias genotípicas en un locus dialélico (que simbolizaremos D, H y R como en el capítulo de conceptos básicos), pueden expresarse según las ecuaciones

$$\begin{aligned} D &= p^2 + Fpq \\ H &= 2pq(1-F) \\ R &= q^2 + Fpq \end{aligned}$$

Estas expresiones matemáticas muestran que la frecuencia de heterozigotos disminuye respecto de la esperada según las proporciones de Hardy Weinberg en favor de cada uno de los dos tipos de homocigotos, y que esa disminución es tanto más ostensible cuanto mayor sea la consanguinidad. Por lo tanto, la consanguinidad conduce a una reducción de la heterozigosidad en comparación con lo esperado en una población con apareamiento aleatorio, donde  $H = 2pq$ . Es claro, pues, que el coeficiente  $F$  mide a la vez una probabilidad y una reducción en la heterozigosidad respecto de lo esperado por el teorema de Hardy-Weinberg.

En una población de tamaño  $N$ , la probabilidad media de que dos alelos escogidos al azar sean idénticos por descendencia respecto de la generación  $t-1$  es

$$F_t = \frac{1}{2N}$$

Por definición, esta es la tasa de consanguinidad debida a la deriva por generación.

Por definición, la heterozigosidad esperada después de una sola generación de consanguinidad es

$$H = H(1 - F)$$

Por extensión, después de  $t$  generaciones de deriva,

$$H_t = H_{t-1} - \frac{1}{2N}(H_{t-1}) = \dots = H_0 \left(1 - \frac{1}{2N}\right)^t$$

Vemos pues que la heterozigosidad se rebaja (como promedio) en un factor  $1/2N$  por generación. No obstante, hay que tener en cuenta dos factores:



- i)  $F$  puede también estar midiendo la consanguinidad como resultado de una subdivisión poblacional que no hemos tenido en cuenta por nuestra ignorancia de la estructura poblacional (efecto Wahlund), y
- ii) puede haber ventaja genética que prevenga la recombinación entre loci sometidos a selección (aunque normalmente se asume que la selección ha eliminado los alelos deletéreos)

### El tamaño poblacional y la consanguinidad

El nivel de consanguinidad ( $F$ ) en una población aumenta en el tiempo a una tasa que se relaciona con el tamaño poblacional efectivo por generación según la ecuación (Wright 1922):

$$\Delta F = \frac{1}{2N_e}$$

donde  $N_e$  es el tamaño poblacional efectivo.

Según Falconer y Mackay (1996) en poblaciones con apareamiento aleatorio, la consanguinidad y la pérdida de variabilidad genética ocurren como consecuencia del tamaño poblacional finito. En ausencia de mutación o selección la tasa de consanguinidad  $F$  puede predecirse por la expresión

$$F = \left[ 1 - \frac{1}{2N_e} \right]^t$$

, donde  $t$  es el tiempo en generaciones.

Una implicación directa de estas relaciones cuantitativas es que las poblaciones más pequeñas tenderán a ser más consanguíneas y que, además, la consanguinidad crecerá en ellas más rápidamente que en las poblaciones grandes. Notemos que, para poblaciones extremadamente pequeñas, el efecto del tamaño poblacional por sí solo puede enmascarar la influencia de otras variables biológicas que sí son de gran importancia en las poblaciones grandes (por ejemplo, el sistema reproductivo). Es esperable que una tasa lenta de consanguinidad favorezca más la eliminación selectiva de los alelos deletéreos que una tasa rápida de consanguinidad. Por ello una consanguinidad lenta tiende a ser menos deletérea que una consanguinidad rápida (Reed et al. 2003). Extrapolando aún más, esperaríamos que menores tasas de consanguinidad den lugar a menores tasas de extinción. Por estos motivos, evitar la consanguinidad es un objetivo primario en el manejo de especies amenazadas o en peligro (Frankham et al. 2002).

### Implicaciones para la conservación

Las mayores amenazas para el mantenimiento de la viabilidad de poblaciones de especies raras son la pérdida de variabilidad por efecto de la deriva genética y los efectos deletéreos de la consanguinidad. Los contrastes entre la distribución espacial y el tamaño poblacional probablemente den lugar a patrones diferenciados de estructura genética. No obstante, aunque hace algún tiempo se asumía que las plantas raras tendían a presentar niveles depauperados de variabilidad genética, hoy en día la relación entre rareza y variabilidad genética no se considera ni mucho menos directa (Stebbins 1980, Nickrent and



Wiens 1989). Los últimos trabajos publicados (Bataillon y Kirkpatrick 2002, Hedrick 2002) indican que las consecuencias genéticas del pequeño tamaño poblacional dependen en gran medida de las características históricas y reproductivas de las especies y, en especial, de su linaje evolutivo. En un trabajo relativamente reciente, Gitzendanner y Soltis (2000) muestran que cuando las comparaciones de niveles de variabilidad genética en términos de características biológicas (sistema reproductivo, mecanismos de dispersión, hábito de crecimiento) se realizan fuera de un contexto filogenético, las predicciones que nos ofrecen tienen un menor poder.

Hay linajes que son más propensos a perder o amplificar su variabilidad genética, ya sea por su antigüedad, por su sistema reproductivo, por su facilidad para la hibridación o por otras características intrínsecas o extrínsecas. En otras palabras, no vale decir que una población pequeña presenta niveles depauperados (o elevados) de variación genética si para establecer la comparación solamente nos basamos en especies o poblaciones que solamente comparten su pequeñez con la que es objeto de nuestro estudio. Las comparaciones con sentido son las que se realizan con poblaciones o especies filogenéticamente cercanas, porque solo en ellas las variables biológicas intrínsecas tendrán un comportamiento comparable. Por ello, a efectos de evaluación comparativa de los niveles de variabilidad en poblaciones pequeñas, es indispensable incluir poblaciones congenéricas (o conespecíficas) ampliamente distribuidas. Es también de gran importancia comprender el concepto de deriva genética y sus implicaciones para la diversificación de poblaciones extremadamente pequeñas.



**LA DEPRESIÓN POR CONSANGUINIDAD** es el decremento en el valor de una característica, normalmente relacionada con la adaptabilidad o la supervivencia, como resultado del apareamiento entre individuos emparentados

La intensidad de este fenómeno se mide normalmente mediante el coeficiente de depresión por consanguinidad,  $\delta$ , cuya estimación viene dada por

$$\delta = 1 - X_1 / X_0 ,$$

donde  $X_1$  y  $X_0$  son los valores medios de carácter resultantes después de una generación de inbreeding y outbreeding, respectivamente. Experimentalmente se han observado altos niveles de  $\delta$  en plantas con alto índice de autofecundación (Roff 1997, 2002).

### CAUSAS DE LA DEPRESIÓN POR CONSANGUINIDAD

Mientras los efectos de la depresión por consanguinidad parecen estar claros y bien investigados por trabajos empíricos y teóricos, las causas que la provocan son aún poco claras, en gran medida porque su investigación requiere conocimientos genéticos profundos de los que aún carecemos. Actualmente, existen dos hipótesis contrapuestas para explicar la depresión por consanguinidad.

**La hipótesis de la sobredominancia** arguye que los heterocigotos poseerían una fitness más elevada. Por consiguiente, la consanguinidad causaría una disminución de la fitness por su efecto de llevar a la homozigosidad. Parece ser que esta hipótesis fue propuesta por primera vez por East (1908). En cambio, la **hipótesis de la dominancia parcial** supone que la depresión por consanguinidad resulta de un incremento en la frecuencia de alelos deletéreos recesivos o parcialmente recesivos en homozigosis (Davenport 1908, Jones 1917).

Aunque han existido discrepancias históricas entre los proponentes de cada uno de estos modelos, la dominancia parcial es la explicación favorecida por los trabajos más actuales (Crow 1993, Charlesworth y Charlesworth 1999, Byers y Waller 1999), a pesar de que existen aún pocas pruebas empíricas de su validez.

No es el objeto de esta brújula dar una visión exhaustiva de las pruebas que llevan a apoyar la hipótesis de la dominancia parcial, son solamente orientar. Lo que nos interesa de todo lo dicho hasta aquí es que todas las evidencias experimentales apuntan a que existen alelos letales cuyo afloramiento en homozigosis provoca la muerte de los individuos y, por tanto, el empobrecimiento demográfico y genético de la población. A tenor de este resultado, parece claro que para comprender el impacto de la depresión por consanguinidad en las poblaciones naturales hemos de preguntarnos cómo influye el tamaño poblacional en la frecuencia de homocigosis de los alelos letales.

### LA FRECUENCIA DE LOS ALELOS LETALES EN UNA POBLACIÓN NATURAL

Wright fue el primero en percibir la problemática del tamaño poblacional y, para intentar esclarecerla, estudió la distribución esperada de alelos letales en una población (Wright 1937) planteando un modelo matemático sencillo en el cual asumió:

- Un locus con dos alelos,  $A_1$  y  $A_2$  de frecuencias relativas  $p$  y  $q$



- Adaptabilidad (fitness) relativa de los genotipos  $A_1A_1$ ,  $A_1A_2$  y  $A_2A_2$  de 1,  $1-hs$  y  $1-s$  (respectivamente), donde  $h$  mide el nivel de dominancia de  $A_1$  y  $s$  la desventaja selectiva ( $s = 1$  indica que el alelo es letal).
- Una tamaño poblacional efectivo de  $N$ , una tasa de mutación ' $u$ ' desde  $A_1$  a  $A_2$  y completa recesividad del alelo deletéreo ( $h = 0$ ).

Según Wright (1937), si  $2Nu > 1$ , entonces la frecuencia de  $A_2$  en el equilibrio será de

$$\bar{q}_\infty \approx \left(\frac{u}{s}\right)^{1/2}$$

Por el contrario, si  $2Nu \ll 1$  para una población de tamaño  $N$ , entonces la frecuencia de  $A_2$  en el equilibrio vendrá dada por

$$\bar{q}_N \approx u \left(\frac{2\pi N}{s}\right)^{1/2}$$

Según esta expresión matemática, cuando el tamaño poblacional es pequeño, la frecuencia media de un alelo recesivo puede ser también muy pequeña. Para percibir mejor este efecto, es más ilustrativo expresar el cociente entre las expresiones en el caso finito e infinito

$$\frac{\bar{q}_N}{\bar{q}_\infty} \approx (2\pi Nu)^{1/2}$$

Así pues, si tomamos  $Nu = 0.001$ , tendremos que  $\frac{\bar{q}_N}{\bar{q}_\infty} = 0.025$ .

Fijémonos que este sencillo ejemplo cuantitativo nos dice que la frecuencia de un letal completamente recesivo en una población de tamaño finito equivaldría solamente al 2.5% de la frecuencia que presentaría en una población de tamaño infinito.

Más adelante, Nei (1968, 1969) utiliza la fórmula general de Wright para obtener una aproximación para  $\bar{q}$  que es independiente del tamaño poblacional efectivo, pero que apenas varía el alcance de las conclusiones ya comentadas y que, por tanto, no vamos a desarrollar.

Basándose en las formulaciones de Wright y Nei, Hedrick et al. (1998) calculan que el número esperado de letales equivalentes en un individuo es, aproximadamente

$$n_{le} \approx 2\bar{q}n$$

donde  $n$  es el número de loci de un organismo que pueden llevar un alelo letal.

## CONSECUENCIAS DE LA DEPRESIÓN POR CONSANGUINIDAD

Quizás la mayor importancia de estas formulaciones teóricas reside en mostrar que, en las condiciones asumidas, el número esperado de equivalentes de letales es una función directa de la frecuencia alélica en el equilibrio. Esta relación general es importantísima para



comprender la relación de la consanguinidad con la viabilidad poblacional, y presenta implicaciones en diferentes áreas de la biología. Probablemente las dos más importantes sean las siguientes:

- 1. Los alelos letales serán más frecuentes cuanto mayor tamaño tengan las poblaciones.** Comprender esta conclusión es en extremo importante, no solamente porque es la base de las demás implicaciones, sino también porque es contraintuitiva. La relación existente entre el tamaño poblacional y la incidencia de la consanguinidad podría llevarnos a pensar (erróneamente!) que las poblaciones más pequeñas serán también más propensas a manifestar depresión por consanguinidad. Sin embargo, la relación entre las frecuencias de los letales en una población finita e infinita (Wright 1937), habría de ayudarnos a entender que los caminos de la consanguinidad y de la depresión por consanguinidad no tienen porqué ser paralelos.
2. Puesto que el número esperado de equivalentes de letales es función directa de la frecuencia alélica en el equilibrio, **el impacto esperado de la consanguinidad en la viabilidad solamente estará en función del tamaño poblacional en el pasado.** Bataillon y Kirkpatrick (2000) muestran que cuando el tamaño poblacional se mantiene pequeño durante mucho tiempo, el nivel esperado de depresión por consanguinidad causada por alelos letales puede reducirse mucho. Además, Hedrick (2002) añade que la frecuencia de los letales con cierto nivel de dominancia también se ve muy reducida por un tamaño poblacional pequeño.

Notemos, pues, que para pronosticar la posible incidencia de la depresión por consanguinidad en poblaciones pequeñas, es imprescindible conocer si han sido pequeñas históricamente o si, por el contrario, su tamaño ha sufrido una reducción reciente. Este aspecto pone de manifiesto que ciertamente el tamaño importa, pero no de manera tan determinante ni tan directa como podría pensarse. Ya en 1991, Barrett y Kohn argüían que es probable que los niveles y distribución de variabilidad genética en plantas dependan de si la especie ha sido siempre rara o de si ha devenido rara súbitamente (por ejemplo, a consecuencia de la influencia humana). Las especies que viven en sistemas poblacionales pequeños y poco distribuidos en el espacio pueden poseer adaptaciones que reduzcan los efectos potencialmente deletéreos de la consanguinidad y de la rareza en general que les permitan distribuirse bien en el tiempo. Un ejemplo conocido lo constituye la especie *Oenothera organensis* que, a pesar de poseer una sola población de tamaño muy pequeño, tiene más de 100 alelos de incompatibilidad diferentes. Por el contrario, las especies que han experimentado reducciones súbitas de su tamaño poblacional pueden ser más susceptibles al stress genético impuesto por un tamaño poblacional inhabitualmente pequeño. **Estas evidencias empíricas y teóricas indican que las poblaciones que han sido pequeñas durante toda su historia pueden haber desarrollado mecanismos para compensar las desventajas inherentes a su pequeñez.**

Lo que estos modelos ponen de manifiesto es que si los tamaños poblacionales son crónicamente bajos, entonces esperaríamos que una menor proporción de la variación genética estuviera influyendo la depresión por consanguinidad. Para el caso particular de las especies amenazadas, este corolario implica que muchos planes para purgar la variación que influye en la depresión por consanguinidad probablemente sean fútiles, puesto que la mayor parte de los letales habrán sido ya eliminados por medios "naturales". Esto es tanto más así cuanto que la reducción de la frecuencia de letales puede ocurrir en unas pocas decenas de generaciones (Hedrick 1994).



## IMPLICACIONES PARA LA CONSERVACIÓN

Las mayores amenazas para el mantenimiento de la viabilidad de poblaciones de especies raras son la pérdida de variabilidad por efecto de la deriva genética y los efectos deletéreos de la consanguinidad. Los contrastes entre la distribución espacial y el tamaño poblacional probablemente den lugar a patrones diferenciados de estructura genética. No obstante, aunque hace algún tiempo se asumía que las plantas raras tendían a presentar niveles depauperados de variabilidad genética, se admite ya desde hace algún tiempo que la relación entre rareza y variabilidad genética no es ni mucho menos directa (Stebbins 1980, Nickrent and Wiens 1989), y que, por tanto las consecuencias genéticas del pequeño tamaño poblacional dependen en gran medida de las características históricas y reproductivas de las especies y, en especial, de su linaje evolutivo. En un trabajo relativamente reciente, Gitzendanner y Soltis (2000) muestran que **cuando las comparaciones de niveles de variabilidad genética en términos de características biológicas (sistema reproductivo, mecanismos de dispersión, hábito de crecimiento) se realizan fuera de un contexto filogenético, su poder de predicción queda atenuado.**

Hay linajes que son más propensos a perder o amplificar su variabilidad genética, ya sea por su antigüedad, por su sistema reproductivo, por su facilidad para la hibridación o por otras características intrínsecas o extrínsecas. En otras palabras, no vale decir que una población pequeña presenta niveles depauperados (o elevados) de variación genética si para establecer la comparación solamente nos basamos en especies o poblaciones que solamente comparten su pequeñez con la que es objeto de nuestro estudio. Las comparaciones con sentido son las que se realizan con poblaciones o especies filogenéticamente cercanas, porque solo en ellas las variables biológicas intrínsecas tendrán un comportamiento comparable. Por ello, a efectos de una evaluación comparativa de los niveles de variabilidad en especies de distribución restringida con poblaciones pequeñas, es indispensable incluir poblaciones congenéricas (o conespecíficas) ampliamente distribuidas. Es también de gran importancia comprender el concepto de deriva genética (ver el apartado correspondiente de esta brújula) y sus implicaciones.

Las consecuencias genéticas del pequeño tamaño poblacional solamente pueden comprenderse desde una perspectiva histórica que tenga en cuenta (entre otras variables) la evolución demográfica de las poblaciones; por lo tanto, existe una estrecha relación entre la demografía y la genética de poblaciones en el ámbito de la biología de la conservación. Fijarse solamente en el tamaño actual de las poblaciones puede tener graves consecuencias para la viabilidad de los planes de conservación sugeridos.

En biología de conservación se hace indispensable averiguar el grado de consanguinidad dentro de las poblaciones, ya que en caso de ser elevado, podría ser necesario adoptar medidas para compensar una posible depresión por consanguinidad. En muchas ocasiones, los organismos amenazados presentan pocas poblaciones de tamaño pequeño, lo cual, según hemos visto, puede acentuar las consecuencias negativas de la consanguinidad.

Es también importante averiguar si la reducción del tamaño poblacional reducido ha sido reciente o se ha mantenido durante generaciones. Cuando la depresión por consanguinidad tiene lugar por causa de tamaños poblacionales pequeños, es probable que los alelos deletéreos se fijen, de manera que el peligro real para estas poblaciones es que la depresión por consanguinidad se fije por deriva genética (Templeton et al. 1991). Cuando



esto ocurre, es posible que se observe muy baja eficiencia reproductiva en las poblaciones afectadas.

Notemos que en muchos organismos una gran proporción de apareamientos consanguíneos no conlleva necesariamente una depresión por consanguinidad. Es más, en numerosas ocasiones se da el fenómeno inverso, es decir una depresión por reproducción cruzada (outbreeding depression), definida por Templeton (1986) como la disminución en la adaptabilidad de los descendientes de reproducción cruzada entre individuos no emparentados por la disruptión (mediante recombinación) de complejos génicos co-adaptados.

Puesto que la acción génica no es aditiva, los mismos genes pueden tener efectos muy diferentes en sustratos genéticos diferentes (de ahí que los complejos génicos co-adaptados puedan poseer valor selectivo). Si la depresión por outbreeding es severa, entonces las probabilidades de extinción de las poblaciones afectadas pueden incrementarse de manera muy dramática. Por lo tanto, comprobar si existe depresión por reproducción cruzada es fundamental antes de diseñar estrategias de reintroducción en poblaciones naturales de organismos amenazados. En caso de existir evidencia de este fenómeno, se hace desaconsejable repoblar una población con individuos provenientes de otra localidad.

Aunque las plantas auto-incompatibles y dioicas no pueden autofecundarse, aún pueden generar consanguinidad a través de apareamientos entre parientes. Cualquier forma de consanguinidad aumenta la homozigosidad por encima de los niveles que esperaríamos si hubiera apareamiento aleatorio. No obstante, el exceso de homocigotos no es exclusivo de la consanguinidad, y puede deberse a muy diversas causas, como por ejemplo una alta incidencia de la reproducción vegetativa (es decir, clonal), o a efectos fundadores (ver el correspondiente apartado de la brújula), por lo cual hay que ser extremadamente cautos en las interpretaciones biológicas de estos parámetros.

**EL TAMAÑO DE LA MÍNIMA POBLACIÓN VIABLE** es el número de individuos necesario para reducir a un nivel aceptable la pérdida de variación genética y de heterozigosidad causadas por la consanguinidad a través de la deriva genética

Puesto que una reducción significativa de la viabilidad es un efecto indeseable para las poblaciones naturales necesitamos tener una medida del tamaño poblacional a partir del cual podemos estar (casi) seguros de que la consanguinidad por deriva no afectará ostensiblemente a nuestra población. Los criadores de animales encontraron por experimentación que un incremento de la consanguinidad del 2-3% por generación puede ser absorbido por las poblaciones sin efectos colaterales graves. Por tanto, un tamaño poblacional que limite la consanguinidad al 1% por generación sería más que adecuado para que la población perdure, y definiría el tamaño de la población mínima viable. Aunque, por razones obvias, esta estimación no debería hacerse extensiva a plantas, el uso del porcentaje calculado para animales de crianza se ha generalizado más allá de su ámbito original de aplicación.

Puesto que, como vimos anteriormente, la heterozigosidad se reduce generación a generación según la expresión

$$H_t = \left(1 - \frac{1}{2N}\right) H_{t-1}$$



resulta que, si aceptamos un decremento máximo de la heterozigosidad de  $\frac{1}{2N} = 0.01$  por generación, entonces el tamaño de la mínima población viable será de 50 individuos.

Estos cálculos ignoran deliberadamente la mutación, porque no tiene sentido considerarla en un intervalo de una sola generación y para un solo gen (su tasa, que se asume del orden de  $10^{-5}$  a  $10^{-6}$ , es muy baja).

Si aceptamos este razonamiento, entonces la tasa de mutación para un carácter cuantitativo podría elevarse a  $10^{-3}$  por generación de forma que, sustituyendo en la ecuación anterior,

$$\frac{1}{2N} = 0.001$$

con lo cual se obtiene que un MVP de 500 individuos sería suficiente para que la variación que surja por mutación reemplace a la que se pierde por deriva genética. El resultado de estos cálculos se conoce en Biología de la Conservación como la "regla del 50/500".

Antes de considerar esta regla, tengamos en cuenta sus limitaciones y el hecho de que estos cálculos no son válidos si trabajamos con caracteres cuantitativos. Estos caracteres están controlados por multitud de genes, de manera que la tasa de mutación espontánea en todos los loci que controlan un carácter cuantitativo podría ser 100 veces más grande que la de un solo gen.

**EL DESEQUILIBRIO EN EL LIGAMIENTO** es la asociación no aleatoria de alelos en dos o más loci (no necesariamente en el mismo cromosoma)

En poblaciones grandes con apareamiento aleatorio, se espera que los alelos de diferentes loci estén asociados al azar (esto es, que exhiban equilibrio en el ligamiento). No obstante, a veces se observa una desviación de las combinaciones aleatorias de las frecuencias génicas, a la que nos referimos como desequilibrio en el ligamiento y que puede estar causado por una variedad de fenómenos (azar, cuellos de botella, consanguinidad, estructura poblacional, selección,...). La desviación de las frecuencias gaméticas observadas respecto de las esperadas en equilibrio se cuantifica mediante el parámetro de desequilibrio en el ligamiento ( $D$ ), denominado así por primera vez por Lewontin y Kojima (1960) [aunque fue introducido por Robbins (1918)] y que, para dos loci, obedece a la fórmula:

$$D = x_{11} - p_1 q_1$$

, donde  $p_1$  y  $q_1$  son las frecuencias de los alelos en los dos loci y  $x_{11}$  es la frecuencia observada del haplotipo que contiene a los dos alelos.

Si no hay equilibrio en el ligamiento,  $D = 0$ .

Si hay desequilibrio en el ligamiento, entonces  $D \neq 0$ .

Si se consideran células diploides, debe tenerse también en cuenta la tasa de recombinación ( $c$ ) entre los dos loci evaluados. Según Falconer y Mackay (1996), después de  $t$  generaciones el desequilibrio remanente es

$$D_t = D_0 (1-c)^t$$



Este parámetro tiene la desventaja de que depende de la frecuencia de los alelos evaluados. Si son 0 o 1 en un locus, entonces  $D$  no puede ser evaluado, y  $D$  es máximo cuando las frecuencias son 0.5. Para paliar este problema, Lewontin (1964) sugirió normalizar  $D$  dividiéndolo por el máximo teórico para las frecuencias alélicas observadas según la expresión

$$D' = D / D_{\max}$$



**EL TEST DE CUELLO DE BOTELLA** determina, en función de un modelo, si una población ha sufrido una disminución brusca de tamaño recientemente a partir de los valores de la heterozigosidad esperada por locus.

La identificación de poblaciones que han experimentado cuellos de botella es importante para el estudio de la diversificación y para la Biología de la Conservación, porque los cuellos de botella pueden incrementar la estocasticidad demográfica, la tasa de consanguinidad, la pérdida de variación genética, la fijación de alelos deletéreos y (cuando dos o más de estos factores convergen) pueden también reducir la probabilidad de supervivencia poblacional (Frankel y Soulé 1981, Luikart y Cornuet 1998).

Cornuet y Luikart (1996) desarrollan un test estadístico para detectar cuellos de botella recientes a través del exceso de heterocigotos. Este test tiene la ventaja de que solamente requiere frecuencias génicas para entre 5 y 20 loci polimórficos en una muestra de aproximadamente 20 a 30 individuos por población.

El test del cuello de botella se basa en que los cuellos de botella generan un exceso de heterocigotos temporal porque al disminuir bruscamente el tamaño poblacional, los alelos raros se pierden a una tasa mucho mayor que la heterozigosidad, como se comentó en el ejemplo de efecto cuello de botella del capitulo de "Conceptos básicos". Así, en una población que haya sufrido un cuello de botella reciente, la heterozigosidad esperada calculada ( $H_e$ ) es mayor que la esperada en el equilibrio ( $H_{eq}$ ), que se obtiene a partir del número de alelos observado ( $k$ ) asumiendo que la población mantiene un tamaño constante (Luikart et al. 1998).

Según Luikart y Cornuet (1998), en poblaciones que no han sufrido cuellos de botella esperaríamos que aproximadamente el 50% de los loci tuvieran un ligero exceso de heterocigotos ( $H_e > H_{eq}$ ), y el otro 50% manifestara un ligero defecto de heterocigotos ( $H_e < H_{eq}$ ) a consecuencia de la deriva genética y del error de muestreo. En cambio, la mayoría de los loci polimórficos en las poblaciones que han experimentado un cuello de botella reciente presentarán con un exceso sustancial de heterozigosidad, lo cual constituye la base del test.

Luikart y Cornuet proponen tres tests para determinar si una población exhibe un número significativo de loci con exceso de heterozigosidad: un test de los signos, un test estandarizado de diferencias y un test de signo de Wilcoxon (Cornuet y Luikart 1996, Luikart y Cornuet 1997). Según estos autores, el primer test tiene bajo poder estadístico y el segundo requiere al menos 20 loci polimórficos. En cambio, el test de Wilcoxon puede usarse a partir de cuatro loci polimórficos y con cualquier número de individuos, aunque se recomiendan 15-40 individuos y 10-15 loci polimórficos para conseguir poder estadístico. En su programa, estos autores también implementan un descriptor cualitativo de distribución de la frecuencia alélica ("mode shift" indicator) que discrimina poblaciones con cuello de botella y poblaciones estables (Luikart et al., 1997). No obstante este método cualitativo no es un test estadístico apropiado, ya que el error de tipo I varía con el tamaño de la muestra y no es apropiado para menos de 30 individuos.

El programa Bottleneck (Piry et al. 1999) calcula por cada muestra poblacional y por cada locus la distribución de la diversidad génica esperada a partir del número observado de alelos ( $k$ ) dado el tamaño muestral ( $n$ ) bajo la asunción de equilibrio mutación-deriva. Esta distribución se obtiene mediante procesos de coalescencia para tres modelos de mutación: alelos independientes (IAM), en dos fases (TPM) y en pasadera (SMM).

La heterozigosidad esperada según cada uno de estos modelos ( $H_{exp}$ ) se compara con la heterozigosidad calculada según las proporciones Hardy-Weinberg ( $H_e$ ) para establecer si existe un exceso o defecto de heterozigosidad en los loci incluidos en los análisis. La diferencia entre los dos valores se estandariza dividiendo la resta de estos valores por la desviación típica (SD) de la distribución de la heterozigosidad según el modelo de mutación-



deriva  $[(He-Hexp)/SD]$ . La distribución obtenida por simulación también permite el cálculo de un valor de probabilidad (P) para la heterozigosidad calculada en el contexto de cada modelo de mutación empleado. Una vez procesados todos los loci polimórficos incluidos para una muestra, se ejecutan los tres tests y se establece si la distribución de la frecuencia alélica tiene forma de L (como se espera en un modelo de equilibrio mutación deriva) o no (como se espera en poblaciones que han sufrido un cuello de botella reciente).

No todos los cuellos de botella pueden detectarse mediante este test, dado que el exceso de heterozigosidad inducida es transitorio y las poblaciones se recuperan de sus consecuencias genéticas en un tiempo relativamente corto que depende de su tamaño y que Luikart y Cornuet (1998) sitúan entre  $0.2N_e$  y  $4.0N_e$  generaciones. A parte de esta limitación "cronológica", el test tiene algunas limitaciones formales importantes. Por ejemplo, los loci que no están en equilibrio Hardy-Weinberg pueden sesgar los resultados del test y han de usarse con precaución (Luikart y Cornuet 1998). El test puede ser desorientador en ocasiones porque puede dar falsos negativos o falsos positivos. En los casos que utilizaron estos autores, el test solamente identificó el 50-75% de las poblaciones naturales que habían sufrido un cuello de botella reciente (normalmente aquellas en que los cuellos de botella habían sido severos y habían inducido un considerable exceso de heterocigotos). Por otro lado, unos pocos datos de poblaciones que no habían sufrido cuellos de botella daban valores significativos en el test (falsos positivos), especialmente bajo el modelo de alelos infinitos. En este aspecto, Luikart y Cornuet (1998) aconsejan utilizar tanto el IAM como el SMM porque argumentan que el verdadero modelo de mutación para la mayoría de los loci debe de ser intermedio entre estos dos. También recomiendan el modelo de mutación en dos fases (TPM) para su test, ya que es intermedio al IAM y SMM (Di Rienzo et al. 1994) y consiste mayormente en mutaciones de un solo paso, pero incluye un porcentaje (5-10%) de cambios en pasadera (Luikart et al. 1998). También aconsejan un mínimo de 10 loci polimórficos en aras de la resolución estadística. La Tabla 2.1 ejemplifica los resultados del test de cuello de botella para los endemismos de Gran Canaria *Lotus genistoides*, *L. holosericeus* y *L. spartoides*.

El exceso de heterozigosidad ( $He > Heq$ ) en poblaciones que han sufrido un cuello de botella se ha demostrado solamente para loci que evolucionan según el modelo de alelos infinitos (IAM) según Maruyama and Fuerst (1985). Bajo el modelo de mutación por pasadera (SMM), pueden darse situaciones de cuello de botella en las que el exceso de heterozigosidad no se observa (Cornuet and Luikart 1996).

Tabla 2.1. Resultados del test de cuello de botella para diferentes poblaciones de las especies *Lotus genistoides*, *L. holosericeus* y *L. spartoides* (a partir de Oliva et al. 2004). PL: número de loci polimórficos usados en los tests. Hd/He: número de loci con defecto de heterocigotos y exceso de heterocigotos según el modelo de alelos independientes (IAE, independent allele model) y de mutación por pasadera (SMM, stepwise mutation model).

Taxón/ población	Tests de cuello de botella				
	PL	IAM Hd/He	IAM P	SMM Hd/He	SMM P
<i>Lotus genistoides</i>					
1. LGCJ	11	9/2	0.021	9/2	0.005
<i>Lotus holosericeus</i>					
2. LHAR	15	15/0	0.000	15/0	0.000
3. LHCU	13	11/2	0.002	11/2	0.001
4. LHPI	15	12/3	0.010	12/3	0.001
5. LHPP	8	5/3	0.641	6/2	0.074
6. LHRO	12	11/1	0.008	11/1	0.002
7. LHTE	8	5/3	0.250	5/3	0.250
<i>Lotus spartoides</i>					
8. LSBE	11	8/3	0.831	8/3	0.365
9. LSTA	11	8/3	0.123	11/0	0.000



**EL TEST DE NEUTRALIDAD DE EWENS-WATTERSON** permite decidir, en función de un modelo, si un locus se desvía de la homozigosidad esperada bajo la hipótesis de neutralidad

Existe una controversia entre los que defienden que la mayoría de variabilidad genética se mantiene por el efecto combinado de la presión de mutación y la deriva genética (neutralismo) y los que defienden que la mayoría de variación se mantiene por selección natural (seleccionismo), de manera que se han desarrollado muchos tests estadísticos para comprobar si se cumple la teoría de la mutación neutral (Kimura 1968), que postula que todos los alelos son selectivamente neutrales. La mayoría de estos tests se basan en la teoría muestral de alelos neutrales desarrollada por Ewens (1972).

La teoría muestral de Ewens asume una población panmíctica de individuos diploides con tamaño efectivo  $N$  con una tasa de mutación  $\mu$  que da lugar a mutantes diferentes de los alelos ya existentes (este es el modelo de alelos infinitos de Kimura y Crow 1964). Ewens (1972) demostró que si cogemos una muestra de  $n$  alelos de una población de este tipo, el número de alelos diferentes ( $k$ ) de la muestra es un estadístico suficiente para estimar el parámetro

$$\theta = 4N\mu$$

Ewens (1972) también mostró que, para un valor fijo de  $\theta$ , la distribución de las frecuencias alélicas en una muestra de tamaño  $n$  viene determinada completamente por el valor de  $k$ . Por lo tanto, si usamos  $k$  para estimar  $\theta$ , entonces la distribución de frecuencias puede usarse para construir un test y determinar si la muestra de frecuencias alélicas es consistente con el modelo de alelos infinitos neutrales. Posteriormente, Watterson (1977, 1978) demostró que la homozigosidad

$$F = \sum_{i=1}^k x_i^2$$

(donde  $x_i$  es la frecuencia alélica del alelo  $i$  en la muestra) puede usarse como estadístico del test de neutralidad.

Para cada locus, el test de Ewens-Watterson calcula la homozigosidad esperada ( $F$ ) asumiendo equilibrio de Hardy-Weinberg y la compara con el valor de la homozigosidad esperada bajo la hipótesis de neutralidad para el mismo tamaño muestral ( $2n$ ) y el mismo numero de alelos ( $k$ ) [el valor de la homozigosidad esperada se obtiene por simulación]. La diferencia entre la homozigosidad observada y esperada se denomina "desviación normal de la homozigosidad", y también se obtiene mediante simulaciones (Salamon et al. 1999). Si su valor es negativo, la homozigosidad observada es más baja que la esperada y la desviación de la neutralidad se interpreta como selección balanceadora. Si el valor de la diferencia es positivo, se atribuye a la selección direccional. En su forma más estándar, el test usa una implementación de Monte-Carlo escrita por Slatkin (1994, 1996), y un método basado en cadenas de Markov para obtener la probabilidad de la homozigosidad bajo la hipótesis de neutralidad. El valor  $p$  es la probabilidad de la homozigosidad observada bajo la hipótesis nula de neutralidad y se puede interpretar como un test de dos colas.

En el ejemplo de la Tabla 2.2, los valores del test de neutralidad se han obtenido mediante el programa Popgen versión 1.32 (1997). Este programa calcula un intervalo de confianza al 95% del valor medio de la homozigosidad esperada bajo la hipótesis de neutralidad. Si el valor promedio de la homozigosidad esperada según las proporciones Hardy-Weinberg (columna "Promedio") entra dentro del intervalo, el test no es significativo y el locus en cuestión puede considerarse neutral. Si el promedio queda fuera del intervalo de confianza, entonces hay motivos para pensar que la selección podría estar actuando en el locus implicado.



Tabla 2.2. Valores del test de neutralidad (obtenidos con el programa Popgene versión 1.32 (1997) para 14 loci polimórficos del endemismo canario *Lotus spartoides* (según Oliva-Tejera et al. 2004). SE es el error estándar, y L95 y U95 son los extremos superior e inferior del intervalo de confianza al 95% de la homozigosidad esperada asumiendo neutralidad. En este caso, todos los loci analizados podrían ser considerados neutrales.

Locus	Test de neutralidad de Ewens-Watterson			
	Promedio	SE	L95	U95
EST-3	0.680	0.034	0.367	0.969
IDH-1	0.815	0.028	0.502	0.984
MDH-2	0.816	0.028	0.502	0.984
MDH-4	0.813	0.028	0.503	0.984
ME-1	0.811	0.028	0.502	0.984
6PGD-1	0.690	0.036	0.361	0.969
6PGD-2	0.825	0.026	0.505	0.984
6PGD-3	0.678	0.034	0.363	0.969
6PGD-4	0.675	0.034	0.370	0.969
PGI-1	0.498	0.027	0.265	0.850
PGM-1	0.575	0.032	0.308	0.923
PGM-2	0.683	0.035	0.358	0.969
PGM-3	0.684	0.034	0.370	0.969
SKD-2	0.820	0.027	0.502	0.984

**LA AUTOCORRELACIÓN ESPACIAL** es una metodología estadística no paramétrica que permite detectar y testar la asociación de los valores de un conjunto de variables en el espacio

La autocorrelación espacial puede definirse como la relación entre los valores de una variable en función de la distribución de los puntos del espacio donde esa variable puede expresarse (Griffith 1987). En el caso de datos genéticos poblacionales, las variables son cada uno de los genotipos detectados en los individuos analizados, y las unidades espaciales son las coordenadas que definen la posición relativa de esos individuos en las poblaciones donde fueron muestreados.

A partir de esta definición, se desprende que para llevar a cabo los análisis de autocorrelación espacial se necesitan en primer lugar matrices de conectividad entre los individuos de una misma población. Estas matrices de conectividad son en realidad matrices de distancia inter-individual que calcula el programa SAAP (Wartenberg 1983) a partir de las coordenadas cartesianas poblacionales que habremos asignado a cada individuo muestreado usando el esquema de la Figura I.2. Para su implementación en SAAP, estas coordenadas deben consignarse (en un archivo de texto) en dos columnas separadas por al menos un espacio, donde la primera columna represente el valor del eje X y la segunda columna el valor del eje Y.

Se necesitan asimismo los valores de las variables a utilizar para cada uno de los individuos incluidos en el análisis, en el mismo orden en el que aparecen en la matriz de coordenadas. Para ello, los genotipos individuales se re-codifican de manera que, para cada alelo considerado, los individuos homozigotos reciban un valor de frecuencia de 1.0, los heterozigotos que contienen el alelo una frecuencia de 0.5 y los individuos que carezcan del alelo una frecuencia de 0.0 (Heywood 1991). Esta forma de codificación no es gratuita, sino que es la única que nos permitirá asimilar el coeficiente de autocorrelación al coeficiente de parentesco de Wright (1922).

Hay una serie de precauciones elementales que deben tomarse antes de generar las matrices de variables para autocorrelación espacial. En primer lugar, los loci monomórficos en una población no deben de ser considerados porque obviamente no contienen ninguna información relevante para este tipo de análisis. De la misma manera, solamente debemos utilizar aquellos alelos de los loci polimórficos con frecuencia poblacional mayor que 0.05 y



menor que 0.95, ya que no es probable que los alelos que se salen de esos umbrales de frecuencia aporten información relevante. En los loci dialélicos que cumplen este requisito, hay que considerar solamente uno de los alelos, ya que el otro provee información redundante (su vector de genotipos codificados podría ser deducido simplemente por comparación con el del otro alelo). En los loci con tres o más alelos, se considerarán todos menos uno, ya que su vector de genotipos también podría ser deducido a partir de los vectores de los demás alelos y aportaría también información redundante.

Una vez culminado este proceso de criba de variables, crearemos un archivo conteniendo los nombres de las variables seleccionadas, separados por un espacio que también ha de introducirse en SAAP.

### El estadístico I de Moran

Por las razones notadas en el párrafo anterior, es conveniente usar el estadístico denominado 'I de Moran' (Moran 1948) como coeficiente de autocorrelación espacial. Este índice se calcula en base a la co-variación en el mapa de muestreo poblacional de los valores de cada una de las variables utilizadas y varía de  $-1$  a  $+1$ , con un valor esperado de  $E(IM) = -1/(n-1)$  [Oden 1984], donde 'n' es el tamaño de la muestra correspondiente.

Cuando  $IM > 0$  para una variable, se dice que ésta exhibe autocorrelación espacial positiva, lo que significa que los puntos que presentan valores similares para esa variable tienden ser cercanos en el espacio poblacional. Cuando  $IM < 0$ , se dice que la variable presenta autocorrelación espacial negativa. En este caso, los puntos con valores similares para la variable tienden a estar separados en el espacio. Con la codificación de genotipos utilizada,  $IM$  es idéntica al coeficiente de relación  $\rho$  de Wright (1922), según Cockerham (1969) y Barbujani (1987). Todos estos cálculos son también llevados a cabo por el programa SAAP (Wartenberg 1983).

Como ilustra la Figura 2.2, las representaciones de la I de Moran para los alelos cuyos correlogramas resultan significativos acostumbra a ser diferente del correlograma medio para todos los alelos incluidos en el análisis (raramente todos los alelos resultan en correlogramas significativos). Por ello, es bueno disponer de ambas representaciones.

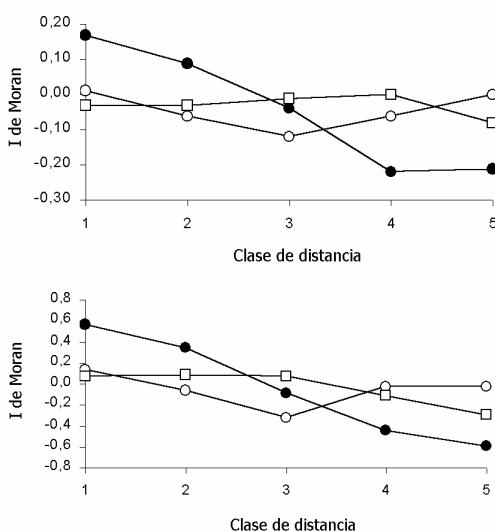


Figura 2.2. Correlogramas medios para todos los alelos (arriba) y para los alelos que dieron correlogramas significativos (abajo) en las poblaciones LL (círculos negros), MP (círculos blancos) y EC (cuadrados blancos). Modificado a partir de Vilches et al. (2004).



**EL TEST DE MANTEL** es una metodología no paramétrica que permite evaluar la significación de la correlación entre matrices de distancia

La correlación entre matrices de distancia no puede testarse de la forma paramétrica, ya que la mayoría de distancias convencionales violan el requisito de independencia exigido (en general, cuando un elemento de una matriz de distancia está muy alejado de un segundo, un tercer elemento muy próximo al primero tenderá también a estar alejado de aquel segundo). Mantel (1967) desarrolló un test para evaluar el grado de correlación entre matrices de distancia referidas a las mismas poblaciones pero basadas en diferentes variables que empezó a aplicarse en geografía y medicina y que luego encontró aplicación en biología poblacional gracias a Sokal (1979).

La hipótesis nula es que la relación observada entre las dos matrices de distancia consideradas podría deberse al azar (es decir, que no hay relación entre las dos variables a las que se refieren las dos matrices de distancia). Para testarla, el test de Mantel compara el valor del determinante resultante del producto de dos matrices de distancia en su orden real ( $Z$ ) con los valores de la distribución de valores generada permutando aleatoriamente el orden de elementos en una de estas matrices.

Si las dos matrices  $A$  y  $B$  están significativamente correlacionadas, entonces el valor observado de l estadístico  $Z$  tendrá que ser grande en comparación con lo que se espera por azar. Si no están correlacionadas, entonces el valor de  $Z$  se confundirá con un porcentaje elevado de los valores que se han generados permutando el orden de una de las matrices al azar. El estadístico  $Z$  tiene la expresión

$$Z = \sum_{i=1}^n \sum_{j=1}^n A_{ij} B_{ij}$$

, donde  $A_{ij}$  y  $B_{ij}$  son cada uno de los elementos de las matrices de distancia A y B.

El estadístico  $Z$  puede normalizarse mediante una transformación estándar, donde la media de la matriz se resta de cada uno de sus elementos y luego cada elemento se divide por la desviación típica de la media. La ventaja de este método es que las unidades estandarizadas son más comparables y fáciles de comprender (Smouse et al. 1986). Este procedimiento da lugar a un estadístico  $r$  que tiene la forma:

$$r = \frac{\sum \sum stdA_{ij} stdB_{ij}}{n - 1}$$

La significación del estadístico  $Z$  (ver Hubert 1987 para los valores aproximados de su media y varianza) puede testarse mediante un test de permutaciones (recomendado para pequeños tamaños muestrales ( $n < 20$ ) o con una aproximación asintótica mediante un estadístico  $t$ ). Aunque pueden calcularse tests de significación de normalidad y éstos se ofrecen como opción en muchos de los programas que realizan el test de Mantel, el mayor poder del test se obtiene comparando el valor observado de  $Z$  con su distribución permutacional. Resultados significativos indican relación entre las dos variables a las que se refieren las matrices de distancia. Un ejemplo de su aplicación en biología puede verse en Malhotra y Torpe (1997).



Según los comentarios del programa NTSYS-pc (Rohlf 1988), el grado de ajuste puede interpretarse subjetivamente como muy bueno ( $0.9 < r$ ), bueno ( $0.8 < r < 0.9$ ), bajo ( $0.7 < r < 0.8$ ) o muy bajo ( $r < 0.7$ ).

Lapointe y Legendre (1992) ofrecen tablas de valores críticos para valores de correlación matricial cuando las dos matrices corresponden a dos cladogramas independientes (distancia patrística), a dos fenogramas (matrices de distancias ultramétricas), o a un fenograma y un cladograma.

### El test parcial de Mantel

Mientras el test de Mantel convencional solamente puede comparar dos variables, el test parcial de Mantel puede comparar tres o más variables (Fortín y Gurevitch 1993). En esencia, el test parcial de Mantel crea una tercera matriz (C) que puede ser una matriz de distancia referida a otra variable o una matriz de distancia referida al diseño experimental. En este caso el test se calcula construyendo una matriz de residuos ( $A'$ ) de la regresión entre A y C, y otra matriz de residuos de la regresión entre B y C ( $B'$ ). Estas dos matrices de residuos se comparan mediante el test de Mantel estándar.



## 2.2.2. DESCRIPTORES DE LA VARIABILIDAD GENÉTICA ENTRE LAS POBLACIONES

---

Las poblaciones naturales de organismos muestran casi siempre diferencias de frecuencias génicas y genotípicas en el espacio geográfico. Si esas diferencias llegan al extremo de ser significativas y las subpoblaciones no intercambian genes (en forma de migrantes, o de granos de polen que fecundan a individuos de otras subpoblaciones), entonces cada una de ellas puede evolucionar independientemente y, con tiempo, dar lugar a nuevas especies. Verificar la existencia de estructura genética poblacional es importante porque en ese caso las poblaciones pueden diversificar y diferenciarse genéticamente.

Aunque la estructuración espacial de la variabilidad genética puede tener consecuencias evolutivas muy importantes, hay que tener en cuenta el impacto de nuestra ignorancia en su estimación antes de entrar en materia. Cuando no somos conscientes de que la entidad colectiva a la que nos referimos como “una única población” puede estar estructurada o subdividida, podemos estar interpretando evolutivamente un fenómeno que no tiene nada que ver con la evolución y que encuentra explicación en nuestra incapacidad de definir con precisión los aspectos más básicos de la naturaleza. Hacer esta consideración tiene sentido, ya que (como hemos visto en el capítulo de conceptos básicos), todos los tests estadísticos sobre predicciones teóricas en genética de poblaciones (por ejemplo, la bondad de ajuste a las proporciones de Hardy-Weinberg) asumen que estamos describiendo adecuadamente las poblaciones. Además, según Slatkin (1982), ignorar la subdivisión poblacional puede sesgar muy fuertemente los resultados de los tests de neutralidad de Ewens-Watterson (ver la sección correspondiente).

El análisis de un número grande de loci incrementará nuestro poder de detección de estructura poblacional porque cada locus contendrá información, en principio independiente, sobre los niveles de deriva, mutación y migración que han afectado a nuestro sistema de estudio. Si los loci analizados están ligados genéticamente, entonces esta información no es independiente y nuestros resultados se referirán más al grupo de ligamiento que a la población de individuos. Por lo tanto, es también importante testar si existe desequilibrio de ligamiento mediante el análisis correspondiente.

Una vez definidos nuestros sistemas poblacionales (un proceso que, en la mayoría de los casos, se realiza *a priori*), existen dos metodologías para estimar la distribución de la variabilidad genética en el espacio. Los estadísticos  $F$  fueron desarrollados originalmente por Wright (1931) para un locus con solo dos alelos y extendidos al caso de alelos múltiples por Nei (1973).



**LOS ESTADÍSTICOS DE ESTRUCTURA POBLACIONAL DE WRIGHT** estiman el grado de diferenciación genética basándose en la comparación entre las heterozigosidades a diferentes niveles espaciales del sistema poblacional considerado.

Los estadísticos de Wright (1931) miden la estructuración genética a diferentes niveles espaciales. Ante un conjunto de poblaciones de una misma especie (o también en una población dividida en varias subpoblaciones) podemos distinguir tres niveles de complejidad: el de los organismos individuales (que notaremos  $I$ ), el de las diferentes poblaciones o subpoblaciones ( $S$ ), y el de la población total ( $T$ ). Atendiendo a esta jerarquización, podemos definir  $H_I$  como el promedio de la heterozigosidad observada para todas las subpoblaciones.

Si notamos  $(H_O)_j$  como la heterozigosidad observada en la subpoblación " $j$ " ( $j = 1, \dots, k$ ), entonces este parámetro se calcularía simplemente como

$$H_I = \frac{1}{k} \sum_{i=1}^k (H_O)_j$$

Por otra parte, si  $p_{ij}$  es la frecuencia del alelo " $i$ " ( $i = 1, \dots, s$ ) en la subpoblación " $j$ " ( $j = 1, \dots, k$ ) y  $(H_e)_j$  es la heterozigosidad esperada según las proporciones Hardy Weinberg en la población " $j$ ", entonces

$$(H_e)_j = 1 - \sum_{i=1}^s p_{ij}^2 \quad \text{y, por consiguiente,} \quad \bar{H}_S = \frac{1}{k} \sum_{j=1}^k (H_e)_j$$

$H_S$  es, por tanto, el promedio de las heterozigosidades esperadas  $(H_e)_j$  por subpoblación según el equilibrio Hardy Weinberg

Finalmente, si  $\bar{p}_i$  es la frecuencia del alelo " $i$ " ( $i = 1, \dots, k$ ) promediada para todas las subpoblaciones, entonces:

$$H_T = 1 - \sum_{i=1}^k \bar{p}_i^2$$

de manera que  $H_T$  es la heterozigosidad esperada considerando las poblaciones analizadas como un todo virtual.

Esta jerarquización de la heterozigosidad es necesaria para calcular los tres indicadores básicos del grado de subdivisión poblacional que se describen a continuación.

**i) El índice de fijación ( $F_{IS}$ ) cuantifica la consanguinidad de los individuos en relación a las subpoblaciones a las que pertenecen**

Según la formulación de Wright (1931) y las definiciones anteriores,

$$F_{IS} = \frac{\bar{H}_S - H_I}{\bar{H}_S} = 1 - \frac{H_I}{\bar{H}_S}$$



Este parámetro también se conoce como coeficiente de consanguinidad, porque ofrece una medida de la reducción de heterozigosidad de un individuo debido al apareamiento consanguíneo no aleatorio dentro de una población. Aunque  $F_{IS}$  se calcula para cada locus analizado, su valor operativo es la media a través de los loci.

- Si  $F_{IS}$  es positivo, las poblaciones tienen un defecto de heterocigotos, que puede ser debido a la consanguinidad o a la selección.
- Si las poblaciones están en equilibrio Hardy-Weinberg,  $\bar{H}_S = H_I$  y entonces  $F_{IS} = 0$ .
- Si  $F_{IS}$  es negativo, las poblaciones presentan un exceso de heterocigotos, que puede haber sido provocado por una inmigración importante conducente a la heterogamia.

El índice de fijación  $F_{IS}$  es el único parámetro de estructuración poblacional que interviene en el cálculo de la tasa de reproducción cruzada (ver el capítulo correspondiente).

## ii) $F_{ST}$ cuantifica la proporción de variación atribuible a la diferenciación inter-poblacional

El parámetro  $F_{ST}$  mide la subdivisión poblacional, que se interpreta normalmente como el efecto combinado de un flujo génico reducido y de deriva genética. Su expresión matemática (Wright 1951) obedece a la fórmula

$$F_{ST} = \frac{H_T - \bar{H}_S}{H_T} = 1 - \frac{\bar{H}_S}{H_T}$$

$F_{ST}$  cuantifica la consanguinidad de las subpoblaciones en relación a la población total de la cual forman parte, y se interpreta como la proporción de la variación total que es atribuible a la diferenciación entre poblaciones (el concepto es análogo al que define el parámetro  $G_{ST}$  de Nei, que se expone en el siguiente apartado).  $F_{ST}$  corresponde a la reducción de la heterozigosidad dentro de las poblaciones y está ligado a las diferencias de las frecuencias alélicas medias entre poblaciones.

Como  $H_S$  corresponde a una parte de la variabilidad que representa  $H_T$ , tendremos que siempre  $H_T \geq H_S$  y por tanto  $F_{ST} \geq 0$  en cualquier caso. Si todas las poblaciones tienen la misma frecuencia alélica y están en equilibrio, entonces  $F_{ST} = 0$  porque en ese caso  $H_T = H_S$ .

$F_{ST}$  refleja la acción conjunta de la deriva (que tiende a diversificar las poblaciones) y la migración (que tiende a homogeneizarlas). Por ello es un indicador de la cohesión genética de las poblaciones (valores altos implican poca cohesión y viceversa). Como la diferencia entre  $H_S$  y  $H_T$  incrementa a medida que las frecuencias génicas se desvían,  $F_{ST}$  se usa también como una medida de distancia genética entre poblaciones.



### iii) $F_{IT}$ cuantifica la proporción de variación atribuible a la diferenciación inter-individual

La cantidad  $F_{IT}$  mide la reducción de heterozigosidad de un individuo en relación al total de poblaciones. Su expresión matemática es

$$F_{IT} = \frac{H_T - H_I}{H_T} = 1 - \frac{H_I}{H_T}$$

Este parámetro cuantifica la consanguinidad de los individuos en relación a la población total de la que forman parte. Puede considerarse la medida más global de consanguinidad, puesto que tiene en cuenta los efectos del apareamiento no-aleatorio dentro de las sub-poblaciones y el efecto de la subdivisión poblacional.

Tabla 2.3. Ejemplificación del cálculo de la heterozigosidad observada ( $H_O$ )<sub>i</sub>, de las frecuencias génicas ( $p_A$ ,  $p_a$ ), de la heterozigosidad esperada, ( $H_e$ )<sub>i</sub> y de los parámetros  $H_I$ ,  $H_T$ , y  $H_S$  en el locus dialélico del sistema poblacional representado en la Figura 2.1.

Poblaciones	$(H_O)_i$	$p_A$	$p_a$	$(H_e)_i$
s1	$4/23 = 0.174$	$14/46 = 0.30$	$32/46 = 0.70 \rightarrow$	$1-(0.30^2+0.70^2) = 0.420$
s2	$1/6 = 0.167$	$9/12 = 0.75$	$3/12 = 0.25 \rightarrow$	$1-(0.75^2+0.25^2) = 0.375$
s3	$3/12 = 0.250$	$11/24 = 0.46$	$13/24 = 0.54 \rightarrow$	$1-(0.46^2+0.54^2) = 0.497$
s4	$3/20 = 0.150$	$23/40 = 0.57$	$17/40 = 0.43 \rightarrow$	$1-(0.57^2+0.43^2) = 0.490$
	↓	↓	↓	↓
	↓	$\bar{p}_A = 0.520$	$\bar{p}_a = 0.480$	↓
	↓		↓	↓
	$H_I = \frac{1}{n} \sum_{i=1}^n (H_O)_i = 0.185$	$H_T = 1 - \sum_{i=1}^n \bar{p}_i^2 = 0.501$		$H_S = \frac{1}{n} \sum_{i=1}^n (H_e)_i = 0.446$

La Tabla 2.3 ejemplifica de forma sencilla los cálculos implicados en la obtención del valor de los parámetros de subdivisión poblacional para el sistema de cuatro poblaciones considerado en la Figura 2.1. En el caso de disponer de varios loci con más de dos alelos por población (como suele ser habitual), se hacen los cálculos para cada locus, pero se usan como referencia para la discusión los valores promedio de estos parámetros. Sustituyendo los cálculos de la Tabla 1 en las ecuaciones de cada uno de los parámetros, los valores de los índices de Wright para cuantificar la subdivisión poblacional serían:

$$F_{IS} = (H_S - H_I)/H_S = (0.446 - 0.185)/0.446 = 0.585 \text{ (corresponde a la reducción global de heterozigosidad observada del individuo respecto a lo esperado en su población)}$$

$$F_{ST} = (H_T - H_S)/H_T = (0.501 - 0.446)/0.501 = 0.110 \text{ (corresponde a la reducción global de heterozigosidad de la población respecto a lo esperado en el total de poblaciones)}$$

$$F_{IT} = (H_T - H_I)/H_T = (0.501 - 0.185)/0.501 = 0.631 \text{ (corresponde a la reducción global de heterozigosidad del individuo respecto a lo esperado en el total de poblaciones)}$$

El bajo valor de  $F_{ST}$  obtenido en este sistema de cuatro poblaciones delataría una considerable cohesión genética, que podría ser efecto de la existencia de niveles moderado-



altos de flujo génico o de una fragmentación reciente. Cualitativamente, esta homogeneidad se manifiesta en el hecho de que los tres alelos considerados se detectan en las cuatro poblaciones y, además, ninguna de ellas posee alelos exclusivos.

### Relación entre los estadísticos $F$

Aunque todos los estadísticos  $F$  cuantifican la consanguinidad, su cálculo y su interpretación biológica difieren dependiendo del nivel de subdivisión espacial que consideremos. Por lo tanto, su estimación es imprescindible para comprender la diferenciación genética espacial en sistemas multi-poblacionales.

Puesto que, como hemos visto,  $F_{IT} = \frac{H_T - H_I}{H_T} = 1 - \frac{H_I}{H_T}$ , tendremos que

$$1 - F_{IT} = \frac{H_I}{H_T} = \left( \frac{H_I}{H_S} \right) \left( \frac{H_S}{H_T} \right) = (1 - F_{IS})(1 - F_{ST})$$

Por lo tanto, los valores de los tres parámetros se relacionan según la fórmula

$$(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$$

Si todas las poblaciones están en equilibrio Hardy-Weinberg, entonces  $F_{IS} = 0$  y  $F_{ST} = F_{IT}$ . Si todas las poblaciones están en equilibrio Hardy-Weinberg y tienen las mismas frecuencias alélicas,  $F_{IS} = F_{ST} = F_{IT} = 0$ , por lo cual la subdivisión poblacional es nula. Cuando los sistemas estudiados tienen múltiples loci y alelos, entonces se utiliza la aproximación multilocus para la estimación de la subdivisión poblacional, debida a Nei (1973).

Los estadísticos  $F$  pueden estimarse a partir de las frecuencias génicas observadas, aunque este no parece ser el mejor método, especialmente cuando los tamaños muestrales son pequeños. Las estimaciones deberían corregirse para los efectos de muestrear un número limitado de subpoblaciones y un número limitado de individuos por subpoblación siguiendo el desarrollo de Weir y Cockerham (1984), pero estas correcciones solamente están disponibles para los estadísticos de Nei (1973) en el programa Genstat PC (Lewis 1993). Por ello, si queremos comparar los valores de los estadísticos  $F$  y de los estadísticos de Nei, es mejor ceñirse a los valores sin modificar que calcula el programa Genstat.

Frecuentemente se utiliza el valor de  $F_{ST}$  en el equilibrio para estimar el número medio de migrantes por generación ( $N/m$ ) entre poblaciones, según las consideraciones del apartado correspondiente de esta brújula.

El tamaño poblacional efectivo de una población subdividida depende también del grado de diferenciación entre los demos (o subunidades de relación genética) en una relación cuantificada por la fórmula (Whitlock et Barton 1996)

$$N_e = \frac{N_d}{(1 - F_{ST})}$$

donde  $N_d$  es el tamaño de cada demo, que se asume fijo.



El parámetro  $F_{ST}$  ha sido durante mucho tiempo el método tradicional para medir el grado de variación genética atribuible al componente inter-poblacional. No obstante, existen fuertes controversias acerca de cómo deberíamos estimar este parámetro. Además, la estimación de los estadísticos  $F$  también se hace más complicada con múltiples alelos y múltiples genes. Weir y Cockerham (1985) sugieren el uso de técnicas de re-muestreo en la matriz de datos de alelos y genes para obtener una distribución empírica de los estadísticos  $F$  usando muchas sub-muestras de la base de datos completa.



**LOS ESTADÍSTICOS DE ESTRUCTURA POBLACIONAL DE NEI** estiman el grado de diferenciación genética a diferentes niveles jerárquicos dentro y entre las poblaciones a través de una descomposición de la variación genética basada en una aproximación para múltiples loci.

Nei (1973, 1987) desarrolló las formulaciones de Wright de forma que  $F_{ST}$  pudiera ser calculado a partir de datos de varios loci. El estimador multilocus de Nei (1973) análogo a  $F_{ST}$  se conoce como  $G_{ST}$  y puede calcularse a partir de la heterozigosidad promedio en la población global ( $H_T$ ) y la heterozigosidad media en cada población ( $H_S$ ) como

$$G_{ST} = \frac{H_T - H_S}{H_T} = \frac{D_{ST}}{H_T}$$

o, alternativamente, como

$$G_{ST} = 1 - \frac{H_S}{H_T},$$

donde  $H_S$  y  $H_T$  se calculan como en el apartado anterior para los estadísticos  $F$  de Wright, y  $D_{ST}$  es la heterozigosidad total distribuida entre las poblaciones, que se calcula mediante

$$D_{ST} = d_s^2 \sum_i \sum_{j > j'} (p_{jj} - p_{jj'})^2, \text{ con}$$

$d_s$  = número de poblaciones consideradas, y  
 $p_{jj}$  = frecuencia del alelo  $i$  en la población  $j$ .

Como  $H_S$  representa un subconjunto de la heterozigosidad total que se encuentra en  $H_T$ , nunca puede superar a  $H_T$  y, por tanto,  $G_{ST}$  siempre será positivo.

$$G_{ST} = \frac{H_T - H_S}{H_T} = 1 - \frac{H_S}{H_T} \Leftrightarrow G_{ST} = \frac{D_{ST}}{H_T} \quad (0 \leq G_{ST} \leq 1)$$

Tomando los valores calculados en la Tabla 4, la estimación de  $G_{ST}$  en el sistema poblacional del ejemplo sería

$$G_{ST} = \frac{0.501 - 0.446}{0.501} = 0.110$$

Según la discusión de este parámetro en la sección correspondiente, el bajo valor delataría una considerable homogeneidad genética en el sistema de cuatro poblaciones considerado (solamente el 11% de la variabilidad total detectada es atribuible a la diferenciación entre poblaciones). Cualitativamente, este resultado se manifiesta en el hecho de que los tres alelos considerados se detectan en las cuatro poblaciones y ninguna de ellas posee alelos exclusivos.

En la práctica habitual,  $D_{ST}$  y  $H_T$  se promedian para todos los loci considerados y se usan para estimar  $G_{ST}$  (según el método de Nei), de forma que



$$G_{st}(Nei) = 1 - \frac{\sum_{i=1}^N H_{si}}{\sum_{i=1}^N H_{ti}}$$

$G_{ST}$  es equivalente a  $F_{ST}$  cuando hay solamente dos alelos en un locus. En caso de múltiples alelos,  $G_{ST}$  es el promedio ponderado de  $F_{ST}$  para todos los alelos (Nei, 1973).  $G_{ST}$  es también similar al parámetro  $\theta$  de Weir y Cockerham (1984), con la excepción de que  $\theta$  tiene en cuenta tamaños de muestra desiguales y número de poblaciones muestreadas. Además,  $\theta$  puede adquirir valores negativos (Weir 1996).

No obstante, Hamrick y Godt (HG) calculan primero los valores para cada locus y después promedian estos valores teniendo en cuenta solamente los loci polimórficos (Berg y Hamrick 1997), de manera que

$$G_{st}(HG) = 1 - \frac{\sum_{i=1}^{N_p} \left( \frac{H_{si}}{H_{ti}} \right)}{N_p}$$

Como es evidente,  $G_{st}(Nei) \neq G_{st}(HG)$ , por lo cual los valores de  $G_{ST}$  dependerán también del método utilizado para calcularlos. Culley et al. (2002) notan este fenómeno y resaltan que estos dos métodos dan el mismo valor solamente en dos casos que se dan muy raramente en la realidad.

Primero, si todas las poblaciones están completamente diferenciadas genéticamente en todos los loci o los valores de  $H_t$  son idénticos en todos los loci, toda la diversidad genética residirá entre poblaciones ( $H_s = 0$ ). En este caso, ambos métodos de estimación darán valores de  $G_{ST} = 1$ . En segundo lugar, si la diversidad genética total se halla dentro de cada población para cada locus analizado ( $H_s = H_t$ ), tendremos que  $G_{ST} = 0$  en ambos métodos de estimación. Por último ambas estimaciones darán el mismo valor si el cociente  $H_s/H_t$  es el mismo para todos los loci.

### Relaciones con otros aspectos de la dinámica poblacional

Frecuentemente los estimadores de  $G_{ST}$  o de  $F_{ST}$  se usan para predecir otros fenómenos genéticos como el flujo génico ( $N_e m = (1 - G_{ST})/4G_{ST}$ ; Wright, 1951) o el número de poblaciones necesario para muestrear un cierto nivel de variación genética ( $P = 1 - G_{ST}$ ; Hamrick et al. 1991). Aunque la aplicabilidad de estas aproximaciones es objeto de debate (Bossart y Prowell 1998, Whitlock y McCauley 1999 sobre la estimación del flujo génico), they representan aplicaciones de  $G_{ST}$  importantes y relativamente comunes. La ecuación de Hamrick et al (1991), por ejemplo, se usa para muestrear y preservar germoplasma de especies amenazadas.

### Implicaciones para la conservación

Hamrick et al. (1983) recomiendan que la consistencia ideal para conservación se consigue cuando al menos el 99% de la variación genética se preserva en bancos de semillas. La estimación del número de poblaciones a muestrear ( $n$ ) para representar una



proporción  $P$  de la variación genética detectada viene dada por la relación  $P = 1 - G_{ST}^{-n}$  (Hamrick et al. 1991). Esta fórmula mostró su robustez frente a diferentes métodos de cálculo de  $G_{ST}$  en las simulaciones llevadas a cabo por Culley et al. (2002). No obstante, un inconveniente potencial a la aplicación de esta ecuación es que asume niveles parecidos de variación genética en las poblaciones consideradas. Esta asunción no presenta implicaciones en el caso de taxones sin subdivisiones significativas, donde estimaríamos  $n$  utilizando el valor de  $G_{ST}$  calculado juntando todas las poblaciones. Pero si la subdivisión poblacional se sustenta a nivel genético, debemos estimar  $n$  para todas las poblaciones y para cada una de las subdivisiones. En este caso, es muy probable que la suma de las  $n$  estimadas para cada subdivisión sea ostensiblemente diferente del valor de  $n$  estimado para las poblaciones agrupadas. Una consecuencia práctica de esta heterogeneidad genética es que necesitaremos muestrear más poblaciones para asegurar la captura de la proporción de diversidad genética sugerida. Esto significa que, en poblaciones subdivididas, es aconsejable optar por la estimación de  $n$  que implica un mayor número de poblaciones, en aras de una buena representatividad genética (ver Caujapé-Castells 2004 para más detalles y un ejemplo práctico).

### **El número de poblaciones a muestrear para capturar una proporción determinada de la variación genética detectada**

Hamrick et al. (1983) recomiendan que la consistencia ideal para conservación se consigue cuando al menos el 99% de la variación genética se preserva en bancos de semillas. La estimación del número de poblaciones a muestrear ( $n$ ) para representar una proporción  $P$  de la variación genética detectada viene dada por la relación  $P = 1 - G_{ST}^{-n}$  (Hamrick et al. 1991). Esta fórmula mostró su robustez frente a diferentes métodos de cálculo de  $G_{ST}$  en las simulaciones llevadas a cabo por Culley et al. (2002). No obstante, un inconveniente potencial a la aplicación de esta ecuación es que asume niveles parecidos de variación genética en las poblaciones consideradas.

Esta asunción no presenta implicaciones en el caso de taxones sin subdivisiones significativas, donde estimaríamos  $n$  utilizando el valor de  $G_{ST}$  calculado juntando todas las poblaciones. Pero si la subdivisión poblacional se sustenta a nivel genético, debemos estimar  $n$  solamente dentro de los grupos genéticamente homogéneos. Este parámetro se aplica eminentemente a efectos de conservación genética y es tratado en mayor profundidad en el apartado 3 de esta brújula.



**LA IDENTIDAD DE NEI ( $I_{NEI}$ ) Y DISTANCIA DE NEI ( $D_{NEI}$ )** miden el grado de proximidad genética entre pares de poblaciones de organismos.

Imaginemos dos poblaciones X e Y, donde disponemos de frecuencias alélicas para un locus isoenzimático con  $i$  alelos ( $i = 1, \dots, n$ ). Sea

$x_i$  = frecuencia del alelo "i" en la población X

$y_i$  = frecuencia del alelo "i" en la población Y.

Nei (1972) define la noción de identidad genética ( $I_{NEI}$ ) entre las poblaciones X e Y mediante la expresión matemática

$$I_{NEI} = \frac{J_{XY}}{\sqrt{J_X J_Y}} \quad 0 \leq I_{NEI} \leq 1$$

donde

$$J_X = \sum_{i=1}^n x_i^2; \quad J_Y = \sum_{i=1}^n y_i^2; \quad J_{XY} = \sum_{i=1}^n x_i y_i$$

y  $i = 1, \dots, n$  es el número de alelos detectado en el locus considerado.

La identidad de Nei (1972) se interpreta como la probabilidad de que, al extraer un alelo al azar de la población X y otro alelo al azar de la población Y, estos alelos sean iguales.

En caso de disponer de datos un número de loci  $L \gg 1$  (como normalmente sucede) la identidad de Nei (1972) se calcula mediante:

$$I_{NEI} = \frac{1}{L} \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}}$$

Si extrapolamos al total de loci analizados,  $I_{NEI}$  representa la probabilidad de que dos poblaciones X e Y sean genéticamente idénticas por descendencia, y es una medida del grado de afinidad genética entre dos conjuntos de organismos.

A partir de la formulación de identidad, Nei (1972) define la distancia genética como

$$D_{NEI} = -\ln I_{NEI} \quad 0 \leq D_{NEI} \leq \infty$$

Posteriormente, Nei (1978) propone una estimación no sesgada de su distancia genética de 1972 que tiene la misma forma general, pero en la cual el cálculo de  $J_X$  y  $J_Y$  se realiza mediante las expresiones

$$J_X = 2n_X \sum_{i=1}^n x_i^2 - 1 \quad J_Y = 2n_Y \sum_{i=1}^n y_i^2 - 1$$

donde  $n_X$  y  $n_Y$  son los tamaños poblacionales en la población X e Y, respectivamente.



El uso de la distancia de Nei (1978) se prefiere comúnmente al de otras distancias porque las simulaciones de Nei et al. (1983) mostraron que este estimador da mejores resultados que otros métodos de cálculo de distancia.

A pesar de esta idoneidad, Hillis (1984) detectó una inconsistencia en la aplicación de la distancia de Nei (1978) a ciertos casos extremos de distribución de frecuencias alélicas entre poblaciones y sugiere una corrección estadística que puede expresarse de la siguiente manera:

$$D_{HIL} = -\ln \frac{1}{L} \sum \left( \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}} \right)$$

Hillis (1984) muestra que el valor de  $D_{HIL}$  no se ve tan afectado por posibles casos extremos de distribución de frecuencias alélicas como  $D_{NEI}$ . Hasta hace poco, uno de los mayores inconvenientes de la distancia de Hillis era que no se hallaba a disposición en los programas de genética de poblaciones al uso. La versión más reciente de NTSYS PC (Rohlf 1988) incorpora una opción para calcularla. En caso de disponer de este programa, es aconsejable utilizar la corrección de Hillis a la distancia de Nei.

Aunque aquí nos hemos restringido la discusión a dos distancias genéticas, existen muchas otras formulaciones, como por ejemplo Wright (1969), Cavalli-Sforza y Bodmer (1971), Moran (1975), Dubois y Prado (1980), Reynolds et al. (1983), Goldstein et al (1995a, b), Tomiuk y Loesche (1991), Weber y Wong (1993), Bowcock et al. (1994), Slatkin (1995), Shriver et al. (1995) o Bruvo et al. (2004).

El término "distancia genética" también se usa en estudios cromosómicos para designar a la distancia entre dos marcadores genéticos en un cromosoma, y se define como el promedio de recombinaciones que tienen lugar entre ellos medida en "unidades de mapa". No hay que confundir ambas nociones.



## ESTIMACIÓN DEL INTERCAMBIO GENÉTICO ENTRE POBLACIONES

La ley de Hardy Weinberg se aplica a sistemas poblacionales cerrados, que no reciben ni exportan migrantes y donde los padres de todos los individuos de la generación siguiente se hallan ya en la población. Estas condiciones raramente se cumplen en la realidad ya que, salvo excepciones, las poblaciones naturales acostumbran a ser sistemas abiertos que intercambian genes.

Existen notables diferencias entre dispersión y flujo génico que debemos tener claras para no malinterpretar los resultados de las formulaciones que se discuten en este apartado. Esencialmente, para que un evento de dispersión afecte al flujo génico entre poblaciones, los migrantes (polen o semillas en el caso de plantas) han de dar lugar a descendencia y al menos algunos individuos de esa descendencia han de alcanzar el estado adulto y a su vez reproducirse.

Aunque tener una estimación numérica del flujo génico es importante, en este apartado veremos que sabemos más acerca de las implicaciones teóricas de la existencia de niveles elevados o bajos de flujo génico que sobre cómo estimarlo correctamente.

El movimiento de genes fuera del espacio poblacional como consecuencia de la dispersión de gametos (polen), zigotos (semillas) o individuos afecta a muchas propiedades ecológicas y evolutivas de los organismos. Entre ellas, quizás la más importante sea la integridad de la especie, ya que el flujo génico contrarresta la divergencia que puede llevar al aislamiento reproductivo que da lugar a la especiación (Whitlock y Mccauley 1999).

Desde que Wright propuso su modelo de estructura poblacional en isla, todos los métodos indirectos para estimar el flujo génico se basan en que, si se cumplen una serie de asunciones, la variancia de las frecuencias génicas entre poblaciones está relacionada con el número de migrantes que llegan a la población cada generación ( $Nm$ ).

Las medidas de flujo génico entre poblaciones basadas en frecuencias génicas usan marcadores genéticos para estimar el producto  $Nm$ , y son una manera de caracterizar el aislamiento reproductivo y la diferenciación genética de las poblaciones. Recientemente, también se están usando datos moleculares sobre sitios de restricción y secuencias de DNA, pero este tipo de aproximaciones son aún minoritarias.

Antes de proseguir es necesario advertir que los patrones de frecuencias génicas poblacionales no detectan solo diferencias atribuibles a flujo génico, sino también aquellas que son debidas a otras fuerzas evolutivas como la selección natural o la deriva genética. No existe en el presente ningún método que permita estimar solamente la contribución del flujo génico a la diferenciación inter-poblacional. Por lo tanto, es indispensable un mínimo conocimiento biológico del organismo para interpretar el valor de flujo génico inter-poblacional.

La estimación del flujo génico interpoblacional puede llevarse a cabo mediante dos métodos: el estadístico  $F_{ST}$  de Wright y el método de Slatkin (1985a, b), basado en las frecuencias de los alelos raros.



**EL NÚMERO MEDIO DE MIGRANTES POR GENERACIÓN** es el promedio de individuos que intercambian dos o más poblaciones en una generación

### Estimación del flujo génico basada en $F_{ST}$

Si disponemos de datos para varios loci multialélicos, Wright (1931) mostró que, asumiendo que los alelos son neutrales

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

Donde  $N$  es el tamaño poblacional local y  $m$  es la tasa media de inmigración en un modelo de estructura poblacional de isla. Despejando de esta expresión podemos estimar indirectamente el número medio de inmigrantes por generación ( $Nm$ ).

La evaluación del flujo génico poblacional según el método de Wright solamente puede realizarse después de determinar la estructura genética poblacional mediante los estadísticos de estimación de la subdivisión poblacional (ver el capítulo correspondiente).

Como norma, se establece que valores de  $Nm > 1$  son típicos de especies con elevado flujo génico (a mayores valores de  $Nm$ , mayores niveles de flujo génico) que, en ausencia de fuerzas opositoras (como ciertas formas de selección natural), tenderán a volverse genéticamente homogéneas. La acción de esas posibles fuerzas opositoras debe evaluarse concienzudamente, ya que pueden mantener la diferenciación poblacional pese a la existencia de niveles elevados de flujo génico.

Aunque el valor de  $Nm$  nos da una estimación cuantitativa, su utilidad es más bien comparativa, ya que con frecuencia las asunciones del modelo de islas en que se basa (especialmente la inexistencia de selección, la ausencia de mutación y la igual contribución de todas las poblaciones al conjunto de migrantes) son violadas por las poblaciones naturales con demasiada frecuencia.

La utilidad de este parámetro para estimar el flujo génico inter-poblacional ha sido criticada duramente por Whitlock y McCaulay (1999). Según estos autores, la mutación no presenta muchos problemas para la estimación del flujo génico a través de  $Nm$  a menos que se produzca a una tasa mucho mayor que la de migración (lo cual es ciertamente improbable en condiciones naturales). No obstante, las demás asunciones del modelo de estructura poblacional en isla son muy poco realistas. Por ejemplo, Whitlock y McCaulay (1999) arguyen que si la tasa de migración no es demasiado elevada, el valor esperado de  $F_{ST}$  dependerá solo del producto  $Nm$  y no de los valores de  $N$  y  $m$  por separado. En consecuencia, aunque  $N$  y  $m$  varíen, no se producirá ningún efecto en el valor de  $F_{ST}$  mientras el valor del producto  $Nm$  se mantenga constante.

Otro punto importante que consideran Whitlock y McCaulay (1999) es que una forma bastante frecuente de variación en el tamaño poblacional es la recurrencia de la extinción local y la colonización de poblaciones. En muchas especies, la tasa de reemplazo de poblaciones es lo suficientemente alta como para afectar al valor de  $F_{ST}$  (Slatkin 1977, Giles y Goudet 1997).

Finalmente, la asunción del modelo de isla de que no existe estructura espacial es violada en la mayoría de ocasiones, puesto que con frecuencia las tasas de migración están correlacionadas con la distancia entre poblaciones (Wright 1943, 1946; Neigel 1997): en ausencia de barreras geográficas, las poblaciones más alejadas entre sí tienden a intercambiar menos migrantes. Por estos problemas asociados al modelo de isla de Wright, Kimura y Weiss (1964) introdujeron el modelo de pasadera (stepping stone), donde el intercambio de migrantes es más probable en las poblaciones adyacentes (ver el correspondiente apartado de la brújula).



### Estimación del flujo génico basada en los alelos privados

A diferencia del método de Wright (1931), que tiene en cuenta a todos los alelos para el cálculo de  $Nm$ , el método de Slatkin (Slatkin 1985a, b) se basa solamente en la frecuencia de aquellos alelos que se encuentran solamente en una subpoblación. El modelo asume que estos “alelos privados” han surgido por mutación y que su frecuencia está modulada exclusivamente por la deriva genética. Los modelos de Slatkin (1985a) establecen que la frecuencia media de los alelos privados es una función simple de  $Nm$ :

$$\ln[\rho(1)] \approx a \ln(Nm) + b$$

, donde

$\rho(1)$  es la frecuencia media de los alelos encontrados en una sola población, y  $a$  y  $b$  son constantes que dependen del número de individuos muestreados por población.

En general, la relación entre  $\rho(1)$  y  $Nm$  está poco afectada por la geometría poblacional, por el número de poblaciones o por el número de individuos por población (Slatkin y Barton 1989). Según Slatkin (1985a), pueden obtenerse resultados fiables muestreando tan solo entre tres y cinco poblaciones.

Aunque esta aproximación asume que los alelos privados tienen un solo origen, Whitkus y Crawford (1987) apuntan que si los alelos ancestrales están restringidos a una sola población, su frecuencia puede no estar relacionada con  $Nm$  de la misma forma que los que se derivan independientemente. Por esta razón, estos autores desarrollan un método de “alelos privados únicos” en el que los alelos privados detectados en taxones relacionados no se tienen en cuenta. Esta aproximación se ha usado en algunos trabajos sobre flujo génico (Wolf y Soltis 1992).

### Reflexiones sobre el movimiento de genes en el espacio

Cualquier nivel de flujo génico entre poblaciones impedirá una diferenciación completa entre ellas, pero el flujo génico ha de sobrepasar ciertos niveles para evitar la diferenciación debida a la deriva genética (Wright 1931). Según Slatkin (1987), la estimación de  $Nm$  es ventajosa precisamente porque nos da una idea de la fuerza relativa del flujo génico y la deriva genética, de manera que (atendiendo a los resultados de las simulaciones llevadas a cabo por este autor), la deriva resultará en diferenciación genética local si  $Nm < 1$ , pero no si  $Nm > 1$ .

Aunque la validez de estos umbrales es solamente comparativa y no absoluta, han servido para hacer inferencias funcionales sobre la diferenciación poblacional. Si el flujo génico es infrecuente (para Slatkin 1987 esta situación se produce cuando  $Nm < 1$ ), entonces las poblaciones tenderán a perder la cohesión genética y a evolucionar de manera independiente. No obstante, la lógica no siempre es aplicable a las poblaciones de organismos, y el hecho de que haya poco flujo génico no debe inducirnos a inferir directamente una divergencia evolutiva rápida. Los niveles de flujo génico requeridos para prevenir la especiación dependerán siempre de la intensidad de actuación de las otras fuerzas evolutivas. Tengamos también en cuenta que la selección natural puede producir adaptaciones a condiciones locales y diferenciación evolutiva incluso en presencia de abundante flujo génico. Lo que parece claro es que un flujo génico abundante puede inhibir la diferenciación genética impidiendo que la selección natural y la deriva establezcan y amplifiquen diferencias locales y que, a igualdad de otros factores, esta inhibición será tanto más efectiva cuando mayores sean los niveles de flujo génico.



## PROGRAMAS MÁS USADOS PARA LA ESTIMACIÓN DE LA DIVERSIDAD GENÉTICA POBLACIONAL

Tabla 2.4. Algunos de los programas de análisis de datos más utilizados en genética de poblaciones, localizaciones de sus sitios web y autores. Los asteriscos señalan las rutinas de transformación de datos que son operativas en Transformer-3 (Caujapé-Castells y Baccarani-Rosas 2005). Los archivos de entrada de todos estos programas podrán ser generados por Transformer-4, actualmente en desarrollo a través de la colaboración entre EXEGEN software y el Departamento de Biodiversidad Molecular del Jardín Botánico Canario "Viera y Clavijo" (Caujapé-Castells y Baccarani-Rosas, in prep.).

PROGRAMA	Sitio web	Referencia bibliográfica
AFLPsurv*	<a href="http://www.ulb.ac.be/sciences/lagev/aflp-surv.html">http://www.ulb.ac.be/sciences/lagev/aflp-surv.html</a>	Vekemans (2002)
Arlequin*	<a href="http://lgb.unige.ch/arlequin/">http://lgb.unige.ch/arlequin/</a>	Schneider et al. (1997)
BAPS	<a href="http://www.rni.helsinki.fi/~jic/bapspage.html">http://www.rni.helsinki.fi/~jic/bapspage.html</a>	Corander et al. (2003)
BAYESASS	<a href="http://www.rannala.org/labpages/software.html">http://www.rannala.org/labpages/software.html</a>	Wilson y Rannala (2003)
BIOSYS*	Pedir información a <a href="mailto:swofford@csit.fsu.edu">swofford@csit.fsu.edu</a>	Swofford y Selander (1989)
Bottleneck*	<a href="http://www.montpellier.inra.fr/CBGP/softwares/">http://www.montpellier.inra.fr/CBGP/softwares/</a>	Piry et al. (1999)
FSTAT	<a href="http://www.unil.ch/izea/softwares/fstat.html">http://www.unil.ch/izea/softwares/fstat.html</a>	Goudet (1995)
GDA*	<a href="http://alleyn.eeb.uconn.edu/gda/">http://alleyn.eeb.uconn.edu/gda/</a>	Lewis y Zaykin (1999)
Genalex*	<a href="http://www.anu.edu.au/BoZo/GenALE/">http://www.anu.edu.au/BoZo/GenALE/</a>	Peakall y Smouse (2005)
GENECLASS-2	<a href="http://www.montpellier.inra.fr/URLB/geneclasse/geneclasse.html">http://www.montpellier.inra.fr/URLB/geneclasse/geneclasse.html</a>	Piry et al. (2004)
GenePop*	<a href="http://wbiomed.curtin.edu.au/genepop/">http://wbiomed.curtin.edu.au/genepop/</a>	Raymond y Rousset (1995)
GeneStat*	<a href="http://lewis.eeb.uconn.edu/lewishome/software.html">http://lewis.eeb.uconn.edu/lewishome/software.html</a>	Lewis (1993)
GEODIS	<a href="http://darwin.uvigo.es/software/geodis.html">http://darwin.uvigo.es/software/geodis.html</a>	Posada et al. (2000)
IMMANC	<a href="http://www.rannala.org/labpages/software.html">http://www.rannala.org/labpages/software.html</a>	Rannala y Mountain (1997)
MIGRATE	<a href="http://popgen.csit.fsu.edu/">http://popgen.csit.fsu.edu/</a>	Beerli y Felsenstein (2001)
NEWHYBRIDS	<a href="http://ib.berkeley.edu/labs/slatkin/eriq/software/software.htm#NewHybs">http://ib.berkeley.edu/labs/slatkin/eriq/software/software.htm#NewHybs</a>	Anderson y Thompson (2002)
NTSys-pc*	Comprar en <a href="http://www.exetersoftware.com/">http://www.exetersoftware.com/</a>	Rohlf (1998)
PARTITION	<a href="http://www.genetix.univ-montp2.fr/partition/partition.htm">http://www.genetix.univ-montp2.fr/partition/partition.htm</a>	Dawson y Belkhir (2001)
PASSAGE*	<a href="http://www.passagesoftware.net/">http://www.passagesoftware.net/</a>	Rosenberg (2002)
PopGene*	<a href="http://www.ualberta.ca/~fyeh/">http://www.ualberta.ca/~fyeh/</a>	Yeh, Yang y Boyle (1997)
SPAGeDi*	<a href="http://www.ulb.ac.be/sciences/ecoevol/spagedi.html">http://www.ulb.ac.be/sciences/ecoevol/spagedi.html</a>	Hardy y Vekemans (2002)
STRUCTURE	<a href="http://pritch.bsd.uchicago.edu/software.html">http://pritch.bsd.uchicago.edu/software.html</a>	Pritchard et al. (2000)
TFPGA	<a href="http://www.marksgeneticsoftware.net/tfpga.htm">http://www.marksgeneticsoftware.net/tfpga.htm</a>	Miller (1997)
WHICHRUN	<a href="http://www-bml.ucdavis.edu/whichrun.htm">http://www-bml.ucdavis.edu/whichrun.htm</a>	Banks y Eichert (2000)



### 3. UNA APLICACIÓN PARA MAXIMIZAR EL MUESTREO DE VARIABILIDAD GENÉTICA EN COLECCIONES *EX SITU*

---



## INTRODUCCIÓN

El conocimiento de los niveles y distribución de la variación genética en las poblaciones naturales de especies de interés es un requisito importante para diseñar estrategias de muestreo que garanticen la captura de una proporción representativa de la variación genética (Hamrick et al. 1991; Frankel et al. 1995; Chamberlain 1998; Francisco-Ortega et al. 2000; Batista et al. 2001, Fay 2003), ya sea con destino a su almacenamiento en bancos de semillas o de genes. Ninguna estrategia de muestreo de material biológico que pretenda ser representativa puede dejar de tener en cuenta las indicaciones de variabilidad genética, ya que ésta es crucial para restaurar hábitats destruidos, reforzar poblaciones en declive o establecer micro-reservas para un monitoreo adecuado.

Hasta hace bien poco, la práctica habitual de muchos bancos de germoplasma consistía en acumular grandes volúmenes de semillas sin disponer siquiera de una estimación básica de la variación genética de las especies diana. Este proceder tiene al menos una consecuencia indeseable, puesto que el gran esfuerzo humano y económico que representa la recolección de germoplasma se muestra de escaso valor efectivo para la conservación: al desconocer cuánta variación genética esperamos recolectar, tampoco podemos disponer de una estimación de la representatividad de nuestro muestreo.

Ciertamente, la estrategia general de basar la conservación *ex-situ* en el análisis espacial de la variación genética demora la recolección de material. No obstante, esta metodología es mucho más ventajosa al menos por tres razones. En primer lugar, porque nos ofrece una mayor garantía de que nuestras recolecciones representan la variación genética de la especie. En segundo lugar, porque minimiza la probabilidad de muestrear individuos relacionados. Y, en tercer lugar, porque reduce el esfuerzo de recolección.

Además, con los avances técnicos actuales, la demora temporal desde que empezamos a generar datos genéticos hasta que adoptamos una decisión de muestreo no tiene porque ser grande. La consecución del programa de análisis de datos *Transformer-3* (Caujapé-Castells y Baccarani-Rosas 2005) ha representado recientemente una gran ayuda, ya que permite la implementación efectiva de urgencia y la reducción del error humano en el creciente número de aplicaciones de los marcadores moleculares poblacionales polimórficos.

La colaboración del Banco de germoplasma del Jardín Botánico Canario "Viera y Clavijo" (hoy en día el recurso de conservación vegetal *ex situ* más importante de Macaronesia) ha sido crucial para esta segunda parte de la brújula. Puesto que la misión de un departamento de este tipo no consiste meramente en "acumular semillas", sino en garantizar la representación de la variabilidad genética de las poblaciones naturales de plantas, parece claro que la estimación de los niveles de variación genética debe ser un cometido inherente a su actividad. En consecuencia, se pensó que complementar la experiencia de este Banco de Germoplasma con las aproximaciones del Departamento de Biodiversidad Molecular podría enriquecer aún más la aportación de estas dos áreas de investigación a la conservación *ex situ* de la biodiversidad vegetal Canaria.

Mi línea de investigación molecular se había concentrado hasta entonces solamente en el estudio del origen filogenético, la diversificación poblacional y la conservación *in situ* de diversos grupos de plantas vasculares. Por ello, la consideración de un enfoque totalmente práctico suponía también un reto para mostrar que los resultados de las técnicas moleculares poblacionales (isoenzimas, microsatélites, RAPDs, AFLPs) pueden también ser utilizados para representar la variación genética en bancos de genes. Si esta oportunidad se ha aprovechado, su grado de éxito solamente podrá evaluarse en el futuro. No obstante, la publicación de nuestras primeras aplicaciones en revistas científicas de considerable impacto internacional solo puede interpretarse como una buena señal.

La estrategia de muestreo resultante (esquematizada en la Figura III.1) propone una evaluación sistemática de los resultados de los análisis genéticos poblacionales con marcadores moleculares que contribuye a acelerar el proceso de toma de decisiones



concretas y a minimizar el error humano en el muestreo *ex situ*. La consecución de este método es un hito importante, ya que la mayoría de investigaciones moleculares con aplicación a la conservación daban siempre orientaciones muy generales en cuanto a las preguntas básicas del muestreo genético para conservación *ex situ*. No obstante, pienso también que esta es solamente la primera etapa de una nueva vía de acción, que pronto nos habrá de permitir incorporar otros tipos de información biológica o histórica a las indicaciones que obtenemos a través de las técnicas moleculares.

En esencia, la mayor novedad del método resultante reside en su manera de relacionar los resultados de algunos desarrollos teóricos independientes (Bengtsson et al. 1995, Hamrick et al. 1991, Sokal 1979) y las indicaciones de algunos estimadores básicos de polimorfismo genético poblacional para dar respuestas concretas a tres preguntas cruciales en el muestreo genético para conservación *ex situ*:

- 1) ¿Cuántas poblaciones habrían de muestrearse intensivamente para capturar un porcentaje significativo de la variación genética de una especie?
- 2) ¿Qué poblaciones habrían de considerarse para cumplir este objetivo con un mínimo esfuerzo?
- 3) ¿Cómo tendría que llevarse a cabo el muestreo en las poblaciones diana para minimizar la probabilidad de muestrear individuos emparentados?

Partiendo de un muestreo de individuos que haya seguido las directrices del apartado I de esta brújula, el método está vertebrado en torno a la respuesta secuencial a las tres preguntas antes formuladas, según las consideraciones que siguen, y que están esquematizadas en el diagrama de flujos de la Figura 3.1.



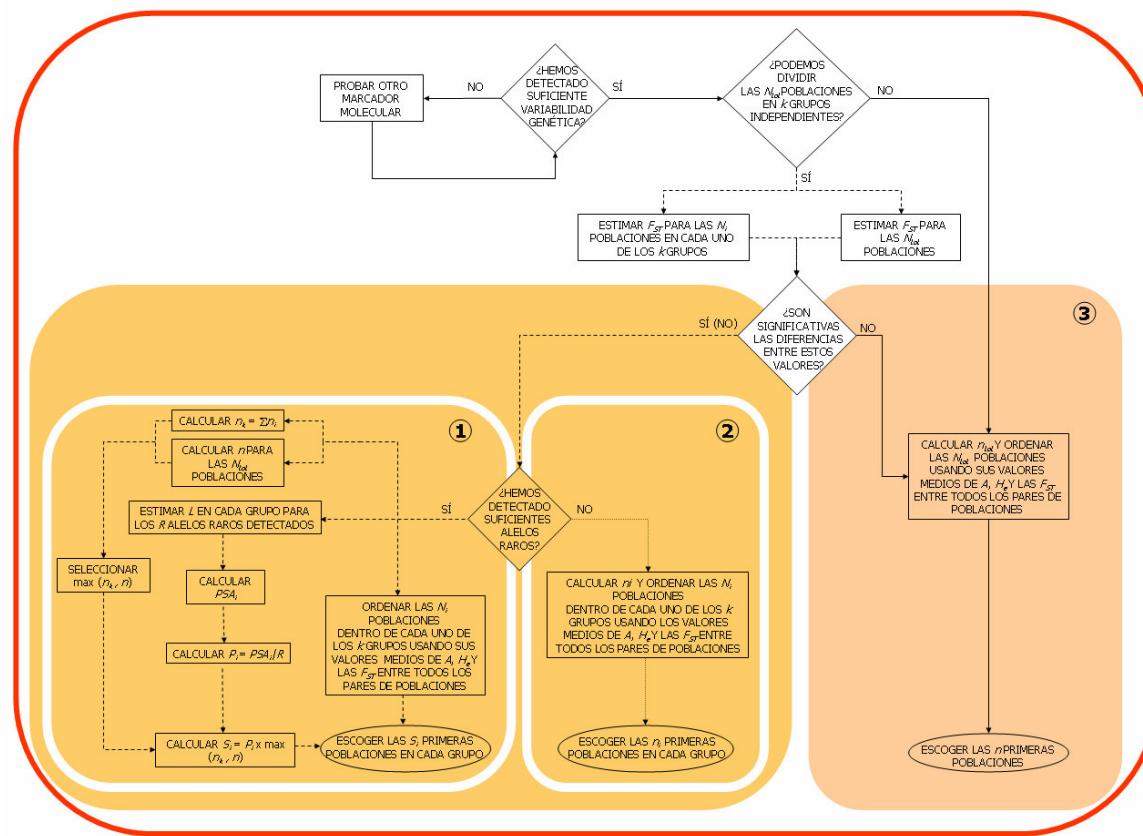


Figura 31. Diagrama de flujos representativo de las etapas del método de muestreo genético desarrollado por el autor de esta brújula (a partir de Caujapé-Castells, in mss.). Las definiciones de los parámetros implicados se dan en la Tabla 3.0. Los números 1, 2 y 3 designan las tres posibles rutas de selección de poblaciones diana para muestreo bajo este método, dependiendo de si hemos detectado suficientes alelos raros (1) o no (2) en grupos con diferencias significativas entre sus valores de  $F_{ST}$ , o de si no hemos detectado diferencias significativas entre los valores de  $F_{ST}$  de los grupos (3).

Tabla 3.0. Definiciones básicas de los parámetros que aparecen en la Figura 3.1 (a partir de Caujapé-Castells, in mss.).

Parámetro	Definición genérica
$N_{tot}$	Número total de poblaciones
$k$	Número de grupos considerados
$N_i$	Número de poblaciones incluidas en la subdivisión "i"
$G_{ST}$	Proporción de variación explicada por el componente inter-poblacional (Nei 1973)
$n_{tot}$	Número de poblaciones a muestrear en el grupo compuesto por todas las poblaciones
$n_i$	Número de poblaciones a muestrear en la subdivisión "i"
$L$	Probabilidad de pérdida alélica para cada alelo que satisface las condiciones de rareza
$R$	Número de alelos que satisfacen las condiciones de rareza
$PSA_i$	Número de alelos raros para los que la subdivisión "i" es el área de muestreo preferida
$P_i$	Proporción de poblaciones a muestrear en la subdivisión "i"
$S_i$	Número de poblaciones a muestrear en la subdivisión "i"
$A$	Número medio de alelos por locus
$H_e$	Heterozigosidad media por locus
$F_{ST}$	Proporción de variación explicada por el componente inter-poblacional (Wright 1951)



## ¿Cuántas poblaciones muestrear para maximizar la variabilidad genética capturada?

---

### Generalidades

A pesar del consenso unánime sobre la importancia del conocimiento de los niveles y distribución de la variación genética en las poblaciones naturales (Hamrick et al. 1991; O'Brien 1994, Frankel et al. 1995; Chamberlain 1998; Pritchard 1999, Shrader-Frechette y McCoy 1999, Francisco-Ortega et al. 2000; Batista et al. 2001, Caujapé-Castells y Pedrola-Monfort 2004, Pearse y Crandall 2004), muchas estrategias de conservación *ex situ* todavía no incorporan consideraciones genéticas. Esta situación es triplemente paradójica: primero, porque contradice el mencionado consenso; segundo, porque ignora el énfasis del primer artículo del Convenio sobre la Diversidad Biológica acerca de la importancia de preservar los recursos genéticos de los organismos vivos (<http://www.biodiv.org/convention>); y tercero, porque existe una gran cantidad de parámetros y tests genéticos que pueden utilizarse para orientar decisiones de muestreo. Acumular semillas sin disponer de una estimación básica de la variación genética de las especies de interés tiene al menos una consecuencia indeseable, puesto que el gran esfuerzo humano y económico que representa la recolección de germoplasma se muestra de escaso valor efectivo para la conservación: al desconocer los niveles de variación genética en el taxón de interés, tampoco podemos saber dónde recolectar, ni disponer de una estimación de la representatividad de nuestra colección.

Ciertamente, la estrategia general de basar la conservación *ex-situ* en el análisis genético molecular es inviable a gran escala y, cuando es viable, demora la recolección de material. No obstante, sí es factible y conveniente aplicar los métodos de la genética a las especies más amenazadas, que son precisamente las que requieren mayor esfuerzo de conservación *ex situ* (Fay 2003). Este proceder presenta al menos tres puntos de interés. En primer lugar, nos puede ofrecer una mayor garantía de que nuestras recolecciones representan la variación genética del organismo a conservar. En segundo lugar, minimiza la probabilidad de muestrear individuos emparentados. Y, en tercer lugar, reduce el esfuerzo humano y económico que implica la recolección. Además, la demora temporal desde que empezamos a generar datos genéticos hasta que adoptamos una estrategia de muestreo no tiene porque ser grande hoy en día.

Como la misión de un banco de germoplasma no consiste meramente en "acumular semillas", sino en garantizar la representación y conservación a largo plazo de la diversidad de las poblaciones naturales de plantas, parece que la estimación de los niveles de variación genética debe ser un cometido inherente a su actividad. El objetivo de este capítulo es mostrar que la información sobre los niveles y distribución de la variabilidad genética puede usarse para diseñar estrategias de muestreo de semillas que representen un máximo de variación genética minimizando el esfuerzo de muestreo, los gastos de recolección y el espacio del banco dedicado al almacenamiento de esas muestras.

Puesto que los alelos más comunes se capturan casi sin querer, cualquier estrategia de muestreo genético debería procurar incluir el máximo número de alelos raros. A efectos de conservación, los alelos raros son productos evolutivos únicos que podrían conferir propiedades ventajosas a la población que los posee frente a eventuales cambios ambientales (Schonewald-Cox et al. 1983; Richter et al. 1994, Bengtsson et al. 1995). Por lo tanto, los diseños de recolección orientados al muestreo de alelos raros ofrecen una herramienta adecuada para reforzar las poblaciones en declive o ayudar a la supervivencia de las plantas reintroducidas.

Estas consideraciones preliminares dejan claro que, cuando se detectan en número suficiente, los alelos raros son importantes en sí mismos en biología de la conservación. Pero



en el caso de taxones con subdivisiones significativas, los alelos raros también pueden ofrecernos indicaciones sobre la proporción de poblaciones a muestrear en cada una de las subdivisiones (ruta 1 de la Fig. 3.1).

Un primer tema relacionado con este tipo de muestreo consiste en saber si una sola población sería suficiente para representar el total de variación estimada en un taxón. Esta consideración puede obviarse en el caso de taxones no subdivididos, ya que la selección de poblaciones a muestrear siempre perseguirá la captura de la mayoría de alelos, especialmente los raros. No obstante (como se plasma en la Figura 3.1) enfocar el problema de la representatividad genética de las poblaciones en términos de alelos raros es indispensable cuando existen subdivisiones, ya que estas consideraciones pueden permitirnos (entre otras cosas) alcanzar una estimación de la proporción de poblaciones de cada subdivisión que deben incluirse en el muestreo total.

## Metodología

Como ilustra la Figura 3.1, una consideración preliminar importante es si deberíamos juntar todas las poblaciones o, por el contrario, considerarlas en grupos separados. Aunque la decisión final siempre vendrá orientada por el método de análisis que aquí se expone, es importante establecer subdivisiones de partida, para testar si nuestros criterios para establecerlas tienen un reflejo a nivel genético, que es el que realmente importa cuando se trata de diseñar una estrategia de muestreo de variabilidad.

En la práctica, puede haber múltiples criterios de subdivisión *a priori*, dependiendo del tipo de taxón y de la información de que dispongamos sobre él. Por ejemplo, Soltis & Gitzendanner (1999) muestran que la conservación de la Biodiversidad tiene más sentido si las unidades biológicas se identifican únicamente con grupos monofiléticos. Por lo tanto, si disponemos de filogenias moleculares para el grupo considerado, uno de estos criterios podría basarse en si hay uno o varios grupos monofiléticos diferenciados a este nivel. Testar la asunción de monofilia es un punto importante, ya que frecuentemente las filogenias moleculares revelan múltiples linajes en sistemas de poblaciones que se habían asumido monofiléticos (ver Soltis et al. 1997 para una revisión).

Otro criterio de subdivisión *a priori* podría ser la separación geográfica (siempre y cuando vaya asociada a un accidente geográfico abrupto o a un tiempo de aislamiento prolongado). Por ejemplo, las poblaciones del endemismo Canario *Matthiola bolleana* (ver ejemplo de aplicación) solamente están descritas en las islas de Fuerteventura y Lanzarote. En principio, es conveniente considerarlas como grupos separados bajo un criterio exclusivamente geográfico y ver si los análisis posteriores confirman esta separación.

Para determinar si las subdivisiones realizadas en base a estos criterios (o quizás a otros que se juzguen adecuadamente objetivos para el caso en cuestión) tienen reflejo a nivel genético, una opción práctica es basarse en las diferencias en el valor del parámetro  $GST$  (Nei 1973). Este estadístico varía entre 0 y 1 y cuantifica la proporción de la variabilidad genética total que se explica por la diferenciación entre poblaciones. Valores muy elevados indican una pronunciada subdivisión genética poblacional, mientras que un valor muy bajo indica una considerable cohesión genética entre las poblaciones. Aunque existen otros métodos para estimar la distribución inter-poblacional de la variación genética (ver por ejemplo Hamrick y Godt 1989), el procedimiento de Nei (1973) parece tener más sentido biológico por ser más sensible a la diferenciación poblacional estimada a través de la variación de frecuencias alélicas entre poblaciones (Culley et al., 2002).

El principio lógico sobre el que nos vamos a apoyar es que si existen diferencias significativas entre el valor de  $GST$  calculado para todas las poblaciones de un taxón y el valor de este parámetro calculado para cada una de las subdivisiones que hemos definido,



entonces podemos argüir que la subdivisión realizada tiene un trasfondo genético y que, por tanto, los grupos deberían considerarse separadamente bajo esta perspectiva. Por el contrario, la existencia de diferencias no significativas entre los valores de  $G_{ST}$  nos ofrece un argumento robusto para considerar a los grupos conjuntamente.

En términos prácticos, si las  $G_{ST}$  parciales no difieren de la  $G_{ST}$  total, está justificado agrupar a todas las subdivisiones como una sola unidad de conservación. Por el contrario, si los valores de  $G_{ST}$  parciales sí difieren de la  $G_{ST}$  total, las subdivisiones consideradas existen también a nivel genético y se sustenta su consideración de manera separada.

Es necesario recordar que, a pesar de que estas indicaciones cuantitativas tienen validez a nivel general, son simplemente pautas orientativas de actuación; en muchas situaciones reales, puede darse el caso de que sea aconsejable mantener las subdivisiones a pesar de que las diferencias genéticas entre ellas no sean significativas. Sería el caso, por ejemplo de taxones con grupos de poblaciones en diferentes zonas ecológicas o en zonas geográficas con difícil intercambio genético entre ellas (diferentes islas, islas y continente, etc...). Hemos de tener presente que los métodos de estimación de la variabilidad genética se basan en un muestreo muy poco significativo del acervo genético total de los organismos (solamente unas pocas zonas del genoma) y, por tanto, siempre estiman la variación genética por defecto. Por ello, es posible que taxones con marcadas diferencias ecológicas, o que hayan permanecido aislados mucho tiempo (esto es, sin intercambiar genes), hayan desarrollado también diferenciaciones genéticas a niveles que no detectemos con nuestro examen de variabilidad. Por estos motivos, mientras solamente hay una trayectoria posible en el organigrama de la Figura 3.1 en caso de detectar diferencias significativas entre los valores de  $G_{ST}$ , parece preferible dejar la opción de escoger entre dos trayectorias si no se detectan diferencias a este nivel (ver Figura 3.1).

En taxones con subdivisiones significativas siempre debemos calcular los valores de  $G_{ST}$  para cada una de las subdivisiones y también para el conjunto de las poblaciones (esto es, como si no hubiera subdivisiones). En taxones sin subdivisiones significativas, bastará el valor de  $G_{ST}$  total.

Hamrick et al. (1983) recomiendan que la consistencia de muestra ideal para conservación se consigue cuando al menos el 99% de la variación genética se preserva en bancos de semillas. La estimación del número de poblaciones a muestrear ( $n$ ) para representar una proporción  $P$  de la variación genética detectada viene dada por la relación  $P = 1 - G_{ST}^{-n}$  (Hamrick et al. 1991). Esta fórmula mostró su robustez frente a diferentes métodos de cálculo de  $G_{ST}$  en las simulaciones llevadas a cabo por Culley et al. (2002). No obstante, un inconveniente potencial a su aplicación generalizada es que asume niveles parecidos de variación genética en las poblaciones consideradas.

Esta asunción no presenta implicaciones en el caso de taxones sin subdivisiones significativas, donde estimaríamos  $n$  utilizando el valor de  $G_{ST}$  calculado juntando todas las poblaciones. Pero si la subdivisión poblacional se sustenta a nivel genético, entonces es preferible estimar  $n$  para todas las poblaciones y para cada una de las subdivisiones. En este caso, es muy probable que la suma de las  $n$  estimadas para cada subdivisión sea ostensiblemente diferente del valor de  $n$  estimado para las poblaciones agrupadas. Una consecuencia práctica de esta heterogeneidad genética es que necesitaremos muestrear más poblaciones para asegurar la captura de la proporción de diversidad genética sugerida (ver los ejemplos de *Androcymbium gramineum* y de *Matthiola bolleana* para dos casos prácticos). Esto significa que, en poblaciones subdivididas, es aconsejable optar por la estimación de  $n$  que implica un mayor número de poblaciones, en aras de una buena representatividad genética.



## La probabilidad de pérdida alélica

En los taxones con subdivisiones significativas donde se hayan detectado alelos raros, es conveniente tener a estos alelos en cuenta en el diseño del muestreo. El razonamiento que subyace a este proceder es que las poblaciones donde hayamos detectado más alelos raros neutrales a través de la técnica molecular aplicada probablemente contendrán también más alelos raros con valor selectivo, que son los que realmente interesan para la conservación de la biodiversidad. Es una asunción discutible (el salto lógico que plantea no tiene por qué ser cierto), pero hemos de considerarla operativa dada nuestra ignorancia sobre el número de los alelos con potencial selectivo y sobre su distribución en los genomas y poblaciones.

Para llevar a cabo este objetivo, es imprescindible comprender la noción de probabilidad de pérdida alélica (ver el capítulo correspondiente en el apartado 2 esta brújula). Para taxones sin subdivisiones, el análisis de la probabilidad de pérdida es opcional, ya que, como se verá más adelante, las poblaciones que tengan más alelos (sean estos raros o no) estarán siempre entre las seleccionadas para el muestreo intensivo.

La probabilidad de pérdida es la probabilidad de que una muestra de tamaño  $N$  no contenga un alelo con frecuencia poblacional  $p$ , y se estima mediante la expresión

$$L = (1 - p)^{2N}$$

solamente para los alelos que cumplan un criterio adecuado de rareza (Bengtsson et al. 1995). Este proceder selectivo de partida es lógico, ya que si se incluyen también en los cálculos los alelos más ubicuos, su bajísima probabilidad de pérdida enmascara a las estimaciones correspondientes a los alelos raros, que son los elementos que nos interesan a este nivel de análisis. Dicho de otra forma, los alelos cuya frecuencia es moderadamente alta tienen una probabilidad de pérdida muy baja, por lo que se capturan casi sin querer y no necesitan de tratamientos estadísticos especiales. Por el contrario, para muestrear los alelos raros hay que “perseguirlos”.

Aunque Bengtsson et al. (1995) recurren a las técnicas estadísticas multivariantes para seleccionar los alelos a analizar mediante esta metodología, personalmente creo que es más práctico (e igualmente efectivo) establecer un umbral de distribución y de frecuencia, como por ejemplo seleccionar los alelos que (1) tengan una frecuencia media menor que 0.5, y (2) estén presentes como máximo en el 50% de las poblaciones consideradas (Vilches et al. 2002, Caujapé-Castells y Pedrola-Monfort 2004, Sánchez et al., in prep.). La Tabla 3.1 ilustra este proceso de selección. La rutina de cálculo de la probabilidad de pérdida alélica del programa Transformer-3 (Caujapé-Castells y Baccarani-Rosas 2005) sigue este criterio, aunque pueden implementarse otros fácilmente.



Tabla 3.1. Frecuencias alélicas para los 11 loci isoenzimáticos interpretados en las 8 poblaciones del endemismo canario *Matthiola bolleana*. Los asteriscos marcan los alelos que cumplieron los criterios de rareza expuestos en el texto. Los alelos enmarcados simbolizan los alelos exclusivos de Fuerteventura (relleno gris) y Lanzarote (sin relleno). Los números enteros son los tamaños muestrales por locus por población. El locus PGD-1 fue excluido de la tabla por ser monomórfico.

Locus/alelo	Fuerteventura					Lanzarote		
	FAJA	CRUZ	OVEJ	TALA	PSBR	CAB1	CAB2	FAMA
MDH1-1A	0,833	0,735	0,775	0,769	0,591	1,000	1,000	1,000
MDH1-1B*	0,000	0,029	0,000	0,038	0,000	0,000	0,000	0,000
MDH1-1C	0,167	0,235	0,225	0,192	0,409	0,000	0,000	0,000
MDH-2A	0,033	0,067	0,031	0,038	0,050	0,000	0,000	0,000
MDH-2B	0,233	0,100	0,188	0,135	0,250	0,000	0,000	0,000
MDH-2C*	0,000	0,000	0,188	0,000	0,000	0,000	0,000	0,000
MDH-2D	0,733	0,833	0,594	0,827	0,700	0,875	0,917	1,000
MDH-2E*	0,000	0,000	0,000	0,000	0,000	0,125	0,083	0,000
IDH-1A*	0,000	0,000	0,000	0,769	0,591	0,000	0,000	0,000
IDH-1B	0,933	0,941	0,975	0,038	0,000	0,125	0,000	0,038
IDH-1C	0,033	0,059	0,025	0,192	0,409	0,000	0,000	0,000
IDH-1D*	0,033	0,000	0,000	0,000	0,000	0,688	0,583	0,846
IDH-1E*	0,000	0,000	0,000	0,000	0,000	0,188	0,417	0,115
PGI-2A	0,800	0,912	0,900	0,846	0,955	1,000	1,000	1,000
PGI-2B	0,200	0,088	0,100	0,154	0,045	0,000	0,000	0,000
GOT-1A	0,000	0,567	0,667	0,280	0,455	1,000	1,000	0,900
GOT-1B*	0,000	0,067	0,111	0,520	0,364	0,000	0,000	0,000
GOT-1C	0,250	0,367	0,167	0,200	0,182	0,000	0,000	0,100
GOT-1D*	0,750	0,000	0,056	0,000	0,000	0,000	0,000	0,000
GOT-2A	0,000	0,118	0,100	0,208	0,091	0,000	0,000	0,000
GOT-2B	1,000	0,882	0,900	0,792	0,909	1,000	1,000	1,000
PGM-1A*	0,179	0,214	0,000	0,000	0,000	0,000	0,000	0,000
PGM-1B	0,500	0,536	0,000	0,000	0,100	0,143	0,083	0,000
PGM-1C	0,321	0,179	0,500	0,808	0,550	0,643	0,500	0,538
PGM-1D	0,000	0,071	0,500	0,192	0,350	0,214	0,333	0,462
PGM-1E*	0,000	0,000	0,000	0,000	0,000	0,000	0,083	0,000
PGM-2A*	0,000	0,000	0,025	0,019	0,000	0,000	0,083	0,154
PGM-2B	0,000	0,000	0,625	0,038	0,000	0,563	0,917	0,846
PGM-2C	1,000	1,000	0,350	0,923	1,000	0,438	0,000	0,000
PGM-2D*	0,000	0,000	0,000	0,019	0,000	0,000	0,000	0,000
PGM-3A*	0,233	0,000	0,000	0,000	0,000	0,000	0,000	0,000
PGM-3B	0,100	0,500	0,125	0,039	0,050	0,875	0,833	0,923
PGM-3C	0,667	0,500	0,875	0,961	0,900	0,125	0,167	0,077
PGM-3D*	0,000	0,000	0,000	0,000	0,050	0,000	0,000	0,000
PGD-2A*	0,000	0,000	0,000	0,019	0,050	0,000	0,000	0,000
PGD-2b	1,000	0,882	0,947	0,923	0,800	1,000	1,000	1,000
PGD-2C*	0,000	0,118	0,053	0,058	0,150	0,000	0,000	0,000



## EJEMPLO PRÁCTICO DE APLICACIÓN

### El endemismo bético-rifeño *Androcymbium gramineum* Cav. McBride (Colchicaceae)

La especie *A. gramineum* es un endemismo exclusivo del área Bético-Rifeña cuyas poblaciones conocidas solamente se distribuyen en la provincia de Almería y en el interior y la costa atlántica de Marruecos (Fig. 3.2). Aunque estos dos conjuntos disjuntos de poblaciones habían recibido denominaciones diferentes (*A. europaeum* para las poblaciones almerienses y *A. gramineum* para las marroquíes), una filogenia molecular de fragmentos de restricción (RFLPs) de DNA cloroplástico (Caujapé-Castells et al. 1999) mostró que deben ser considerados bajo la misma denominación a pesar de la discontinuidad geográfica que representa el estrecho de Gibraltar (Fig. 3.2).

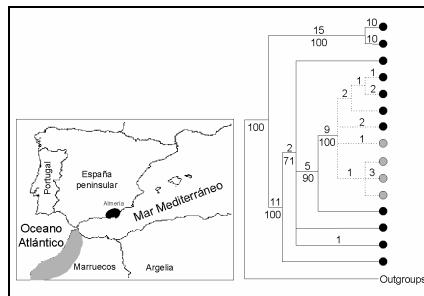


Figura 3.2. Distribución geográfica de *A. gramineum* en Almería (relleno negro) y Marruecos (relleno gris) y filogenia de RFLPs de *A. gramineum* (modificado a partir de Caujapé-Castells y Jansen 2003). Los números encima de las ramas son cambios de sitios de restricción y los números debajo de las ramas, valores bootstrap. Se observa que las poblaciones marroquíes se insieren dentro de un grupo derivado (apoyado por un valor bootstrap del 100%) en el que también hay poblaciones españolas. Esta topología apoya la consideración de las poblaciones españolas y marroquíes bajo la misma denominación taxonómica.

Mas allá de este resultado, Caujapé-Castells y Jansen (2003) apoyan con tiempos de divergencia calculados a partir de los datos moleculares que las poblaciones marroquíes son probablemente más modernas que sus congéneres españolas, y que *A. gramineum* pudo colonizar Marruecos aprovechando la desecación Mesiniense del Mediterráneo (hace unos 5.5 millones de años), que estableció un puente terrestre entre Europa y África a través de lo que hoy es el estrecho de Gibraltar. En consonancia con este hecho, las poblaciones almerienses poseen niveles mucho más altos de variabilidad genética (Caujapé-Castells y Jansen 2003, Caujapé-Castells y Pedrola-Monfort 2004) que las marroquíes.

A pesar de que las evidencias filogenéticas instan a considerar estas poblaciones como un solo grupo, los indicadores de variabilidad genética poblacional (isoenzimas) son muy diferentes en sus dos regiones geográficas de distribución. Como se explica en Caujapé-Castells y Pedrola-Monfort (2004) Almería posee 20 alelos exclusivos, mientras que en Marruecos solamente se detectaron cuatro alelos exclusivos. Además de estas diferencias cualitativas, los valores de  $G_{ST}$  en los dos compartimentos geográficos son muy diferentes entre sí ( $G_{ST} = 0.413$  para Almería y  $G_{ST} = 0.240$  para Marruecos) y respecto del valor de  $G_{ST}$  estimado para *A. gramineum sensu lato* ( $G_{ST} = 0.371$ ). Por si fuera poco, el valor de  $G_{ST}$  en Almería es bastante superior al calculado para la totalidad de la especie, lo cual enfatiza el hecho de que las poblaciones de esta región geográfica son mucho más heterogéneas a nivel genético (presumiblemente como consecuencia de su mayor antigüedad) que las marroquíes.

Al aplicar la fórmula de Hamrick et al. (1991) para estas subdivisiones geográficas (fijando  $P = 0.99$ ), resultan valores de  $n = 5$  para Almería y  $n = 3$  para Marruecos. Para *A. gramineum* sensu lato (es decir, considerando todas las poblaciones como un solo conjunto), resulta un valor de  $n = 5$  (4.63 en números exactos).

Estas cifras nos hacen ver que si tomamos el valor conjunto de  $n$  como referencia cuantitativa para el muestreo, tendríamos que muestrear 3 poblaciones menos que si tomáramos como referencia el valor de la suma de  $n$  para Almería y Marruecos consideradas independientemente. Notemos que si los niveles de variación genética hubieran sido similares en los dos compartimentos geográficos, el valor de  $n$  estimado para *A. gramineum sensu lato* hubiera sido igual a la suma de los valores de  $n$  estimados para cada subdivisión geográfica. En conclusión, la heterogeneidad genética encontrada en cada una de las áreas de distribución de *A. gramineum* hace más adecuado elegir el estimador de  $n$  que implica mayor número de poblaciones. Este número es  $n = 5 + 3 = 8$ .

He seleccionado este ejemplo intencionadamente porque las evidencias filogenética y genética poblacional dan indicaciones antagónicas. Según la filogenia molecular habríamos de considerar conjuntamente las poblaciones almerienses y marroquíes; según las estimaciones de subdivisión genética (y la geografía!) habríamos de tenerlas en cuenta por separado. En caso de conflictos como el ilustrado, parece siempre más práctico resolver a favor del criterio que implique existencia de subdivisión. De esta forma, se preservará el objetivo de esta metodología de muestreo para conservación *ex situ*: la captura de una representación consistente de la variabilidad genética.



Una vez culminado este proceso, deben calcularse las probabilidades de pérdida observada ( $L_o$ ) y esperada ( $L_e$ ) para los alelos seleccionados (deberían ser un mínimo de 10 para garantizar la robustez de los datos).

La probabilidad de pérdida observada,  $L_o$ , se obtiene sustituyendo en la fórmula la frecuencia media de cada alelo en todas las poblaciones consideradas ( $p$ ) y el número de poblaciones donde ha sido detectado ( $N$ ). En este caso, pues, tanto  $p$  como  $N$  son variables para cada uno de los alelos analizados.

La probabilidad de pérdida esperada,  $L_e$ , es la probabilidad de pérdida que tendría un alelo si ostentara su frecuencia media  $p$  en todas y cada una de las poblaciones consideradas. Por lo tanto, para estimarla se fija el valor de  $N$  como constante (igual al número total de poblaciones consideradas), y la única variable a considerar es el valor de  $p$  para cada uno de los alelos incluidos en el análisis.

Los valores de  $L_o$  y  $L_e$  así obtenidos se utilizan posteriormente para hacer una regresión lineal donde las frecuencias medias de cada alelo incluido en el análisis se representan en el eje x y  $-\log L_o$  and  $-\log L_e$  se representan en el eje y (ver figura 3.3). De acuerdo con Bengtsson et al. (1995), puede obtenerse un valor de representatividad ( $R$ ) de una población de ese grupo en relación a los alelos raros detectados dividiendo la pendiente de la línea de regresión observada (basada en los valores de  $L_o$ ) por la pendiente de la línea de regresión esperada (basada en los valores de  $L_e$ ).

En el caso de taxones con grupos de poblaciones claramente heterogéneos, estos cálculos son obligatorios para poder estimar la proporción de poblaciones de cada subdivisión que habrían de incluirse en el muestreo total. En este tipo de casos es necesario hacer estos análisis tanto con el conjunto de las poblaciones (como si no hubiera subdivisión) como para cada una de las subdivisiones (para saber en cual de ellas los alelos raros tienen menos probabilidad de pérdida).

En caso de hallarnos ante un caso de taxones subdivididos, el último paso que hay que llevar a cabo es consignar al lado de cada alelo raro seleccionado cual es la subdivisión en la que sería más probable muestrearlo [esto es, aquella en la que la probabilidad de pérdida observada ( $L_o$ ) sea menor]. Al final de este proceso contamos con una estimación de las proporciones que deben guardar los números de poblaciones de cada subdivisión a incluir en el muestreo (dividiendo el número de veces que la subdivisión correspondiente es seleccionada por el total de alelos incluidos en el análisis). Para una ilustración práctica, ver la Tabla 3.2.

Para estimar cuantas poblaciones debemos muestrear dentro de cada subdivisión basta con multiplicar esta proporción por la  $n$  máxima resultante de aplicar la formulación de Hamrick et al. (1991).



## ¿Qué poblaciones muestrear?

---

Los números resultantes de la aplicación de las formulaciones anteriores no contienen ninguna indicación sobre qué poblaciones debemos muestrear para representar la proporción de variación genética sugerida. La teoría básica de la genética de poblaciones indica que es aconsejable considerar tanto el número medio de alelos por locus ( $A$ ) como el valor de la heterozigosidad esperada ( $H_e$ ) para la selección de los puntos de muestreo. Aunque en muchos casos existe una relación directa entre los valores de estos dos parámetros, éstos ofrecen tipos diferentes de información. Mientras  $A$  da una indicación grosera de la cantidad de variación genética por población,  $H_e$  ofrece una medida indirecta de heterogeneidad genética que está menos afectada por el error de muestreo (es decir, su valor no se ve influenciado por las frecuencias de los alelos raros). Además,  $H_e$  tiene más sentido para la conservación, ya que su relación con caracteres relacionados con la adaptabilidad es comúnmente aceptada (Mitton & Grant 1984; Zouros & Foltz 1987) y se reveló estadísticamente significativa en un meta-análisis (Britten 1996).

En previsión de una falta de resolución en la ordenación suministrada por  $A$  y  $H_e$ , resulta conveniente incluir como tercer criterio un parámetro que nos informe sobre la singularidad genética de cada población, esto es, de su grado de diferenciación genética respecto de las demás. A este nivel, es muy informativo el valor de  $F_{ST}$  (Wright 1951). Por la interpretación biológica de este parámetro (explicada en el apartado correspondiente de esta brújula), mayores valores de  $F_{ST}$  indican mayor diferenciación genética. Por lo tanto, la prioridad a este nivel inter-poblacional será muestrear las poblaciones que ostenten un mayor valor promedio de  $F_{ST}$  respecto a todos los pares de poblaciones incluidas en el análisis.

### Metodología

Ordenar las poblaciones de mayor a menor tomando como referencia los valores de  $A$  y  $H_e$ . En caso de tener varios valores iguales de  $A$ , siempre prevalecerá(n) la(s) población(es) que ostente(n) un mayor valor de  $H_e$  y mayor valor promedio de  $F_{ST}$  con las otras poblaciones. En caso de taxones con subdivisiones significativas, hay que desarrollar este proceso para cada una de ellas. Una vez terminado este proceso, debemos seleccionar las  $n$  primeras poblaciones de cada grupo, según indiquen los resultados de los análisis anteriores. Estas poblaciones serán las candidatas a muestreo intensivo para capturar la variabilidad genética del taxón.

## ¿Cómo muestrear los individuos en las poblaciones?

---

Interpretar los atributos biológicos de los organismos en función de consideraciones espaciales asociadas a su variación genética parece una forma consistente de sugerir diseños de recolección que maximicen la variabilidad capturada y, a la vez, minimicen la probabilidad de incluir en la muestra individuos emparentados.

En base a esta premisa, una vez han sido clarificados los aspectos anteriores del muestreo genético, es también importante saber cómo muestrear los individuos dentro de las poblaciones seleccionadas. En caso de no tener en cuenta este factor, podríamos incluir en las colecciones solamente a unos pocos individuos genéticamente distintos, lo cual podría poner en riesgo el éxito a largo plazo de eventuales reforzamientos poblacionales o reintroducciones. Con frecuencia, la consanguinidad menoscaba la capacidad de



supervivencia de las poblaciones naturales (por ejemplo, por la acción de la depresión por consanguinidad) y debe evitarse en la medida de lo posible. Aunque la comprobación de si la consanguinidad da lugar a una depresión por consanguinidad no puede obtenerse a través de marcadores moleculares, estos sí pueden ofrecernos algunas orientaciones para minimizar la probabilidad de muestrear individuos emparentados.

La autocorrelación espacial (ver Sokal 1979 para sus aplicaciones generales en biología y Caujapé-Castells y Pedrola-Monfort 1997 o Caujapé-Castells et al. 1999 para sus aspectos de genética de poblaciones) es una metodología estadística no paramétrica adecuada para este propósito por dos razones.

En primer lugar, uno de los coeficientes de autocorrelación espacial (la "I de Moran", que notaremos IM a partir de ahora) puede equipararse al coeficiente de relación genética  $\rho$  de Wright (1922). Por lo tanto, la autocorrelación espacial nos proporciona una estimación del grado de parentesco entre los individuos muestreados.

En segundo lugar, cuando se detecta autocorrelación espacial positiva (es decir, una disminución aproximadamente monótona desde valores positivos a valores negativos de IM a medida que la distancia entre individuos aumenta), puede estimarse el diámetro de las áreas genéticamente homogéneas (dentro de las cuales la probabilidad de muestrear individuos emparentados es máxima). En virtud de esta propiedad de la autocorrelación espacial, podemos establecer distancias mínimas de muestreo a partir de las cuales la probabilidad de muestrear individuos emparentados es mínima. Veamos más detalladamente cuales son los requisitos para la aplicación y la interpretación de los resultados de esta técnica.

Si hemos detectado autocorrelación espacial positiva, la primera intercepción del correlograma medio con el eje de abscisas nos da una estimación del tamaño de las áreas genéticamente homogéneas. Esta intersección se calcula por interpolación simple, ya que conocemos las equivalencias en metros para cada clase de distancia.

En caso de no detectar autocorrelación espacial, hemos de asumir que la variación genética se distribuye al azar dentro de la población, con lo cual la separación entre los individuos de los que recolectemos semillas puede ser la que consideremos oportuna. En cualquier caso, es adecuado que quede constancia de las distancias de recolección por si en el futuro observáramos anomalías en las eventuales reintroducciones que se realizaran con estas semillas.



## EJEMPLO PRÁCTICO DE APLICACIÓN 1

### El endemismo canario *Matthiola bolleana* (Cruciferae)

*Matthiola bolleana* (Brassicaceae) es un endemismo canario que hoy en día está distribuido exclusivamente en las islas de Fuerteventura y Lanzarote (Fig. 3.3). Aunque esta especie endémica no está protegida ni amenazada y goza de unos niveles de variación genética moderado-altos (Sánchez et al. 2004), es precisamente en este momento cuando debemos intentar capturar su variación genética, y no esperar a actuar cuando algunos de sus elementos distintivos puedan haber ya desaparecido a causa de algún factor externo.

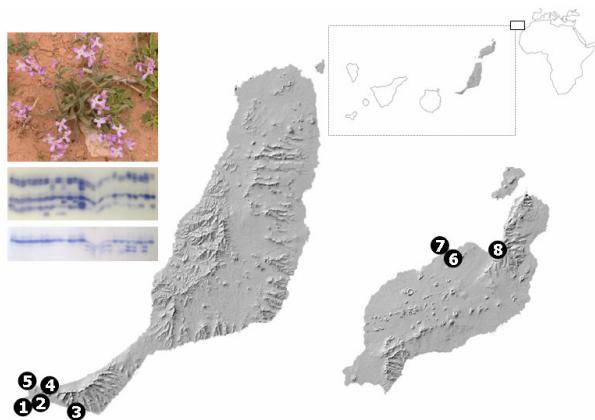


Figura 3.3. Localización espacial de las ocho poblaciones de *Matthiola bolleana* muestreadas en Fuerteventura (izquierda) y Lanzarote (derecha) por Sánchez et al (2004, 2006). Las fotografías muestran, de arriba abajo, una imagen de *Matthiola bolleana* en la naturaleza, la resolución de la enzima PGM (froglucomutasa) y la resolución de la enzima PGI (froglucoisomerasa). (La Fotografía de Matthiola bolleana es de Juli Caujapé-Castells, y las de los enzimas, de Juan Luis Sánchez Doreste)

En ausencia de información filogenética sobre *M. bolleana* (por el momento), su distribución geográfica lleva a considerar a las poblaciones distribuidas en cada isla como grupos independientes de partida. Los datos genéticos poblacionales para esta especie (Sánchez et al. 2004) refuerzan este criterio al estimar que las poblaciones de Lanzarote ( $G_{ST} = 0.074$ ) están genéticamente más cohesionadas que las de Fuerteventura (con una  $G_{ST} = 0.234$ ). La heterogeneidad genética entre las islas se demuestra por el hecho de que, al contrario que en el ejemplo de *A. gramineum*, cuando consideramos todas las poblaciones indistintamente, el valor de  $G_{ST}$  es el más alto ( $G_{ST} = 0.367$ ). Aplicando la fórmula de Hamrick et al. (1991) a estos datos, resultan valores de  $n = 3.17 = 3$  (Fuerteventura),  $n = 1.76 = 2$  (Lanzarote),  $n = 4.79 = 5$  (*M. bolleana sensu lato*). Este caso es pues diferente del de *Androcymbium gramineum* en que la suma de las  $n$  para cada una de las islas coincide con el valor de  $n$  obtenido para toda la especie. Con subdivisiones o no, el número de poblaciones que tendríamos que muestrear intensivamente para conservar la variabilidad genética de *M. bolleana* es de 5.

Para perfilar qué número de poblaciones debería muestrearse en cada área, se seleccionaron 17 alelos que cumplían con los criterios de rareza estipulados en el texto (Tabla 3.1). La representación de las probabilidades de pérdida frente a las frecuencias relativas medias de cada alelo indicó que la representatividad de una sola de las poblaciones de *M. bolleana* en términos de alelos raros es de  $R = 0.504$ , claramente insuficiente para nuestro objetivo de muestreo genético.

La aplicación de la formulación de probabilidades de pérdida (Tabla 3.2) permitió establecer que en 5 de los casos, el área de muestreo preferida sería Lanzarote, y en 12 Fuerteventura. Si observamos con atención esta tabla, nos habremos dado cuenta de que los únicos alelos para los que el cálculo de las probabilidades de pérdida parciales ha sido útil son *Idh1-d* y *Pgm2-a*, que están presentes en poblaciones de Fuerteventura y Lanzarote. El resto son o bien exclusivos de Fuerteventura (12 alelos), o bien de Lanzarote (3 alelos). Notemos que si no hubiera alelos raros compartidos, la proporción de cada isla a incluir en el muestreo vendría dada directamente por el cociente nº de alelos exclusivos de Fuerteventura/17 y nº de alelos raros exclusivos a Lanzarote/17.

Sea como fuere, con estos números es fácil calcular que las proporciones de poblaciones de Fuerteventura y Lanzarote en el muestreo final deben de ser 0,7:0,3 respectivamente (12/17 y 5/17). De manera igualmente sencilla, para saber el número de poblaciones que deberíamos muestrear en cada área basta multiplicar estas proporciones por 5. El resultado es 3,5 poblaciones en Fuerteventura y 1,5 en Lanzarote. Como no podemos muestrear décimas de población, una opción para resolver este problema consiste en recurrir a la Tabla 3.1 y ver qué poblaciones contienen los alelos raros. En Lanzarote, vemos que muestreando en CAB2 es muy probable que capturemos todos los alelos raros detectados en la isla. Por lo tanto, basándonos en este criterio cualitativo, la decisión final de muestreo sería 1 población en Lanzarote y 4 en Fuerteventura.



He seleccionado este ejemplo para poner de relevancia el hecho de que los números de poblaciones a muestrear en cada subdivisión no coinciden necesariamente con los calculados con la fórmula de Hamrick (1991) para cada subdivisión. Por si no ha quedado claro en el apartado de metodología, esta fórmula se utiliza solamente para seleccionar el número total de poblaciones a muestrear intensivamente según la metodología expuesta. El número concreto de poblaciones a muestrear en cada subdivisión se obtiene solamente a partir de las probabilidades de pérdida de los alelos que cumplen el criterio de rareza que hayamos establecido. Cualquier coincidencia numérica es contingente, ya que los valores de  $n$  resultantes de la fórmula de Hamrick solamente serían estimadores directos del número de poblaciones a muestrear en cada subdivisión en caso de no haber alelos raros en la composición genética de nuestra muestra.

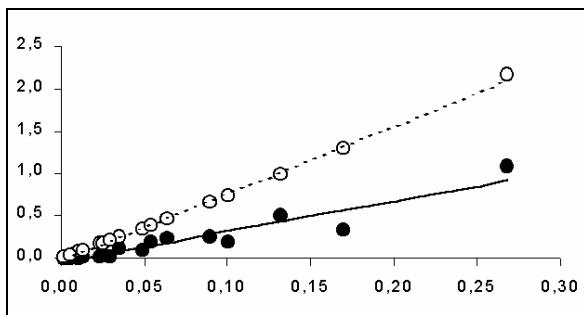


Figura 3.3. Regresión de la frecuencia media de los alelos considerados para la probabilidad de pérdida (eje x) y  $-\log(Lo)$  [puntos negros] y  $-\log(Le)$  [puntos blancos] (eje y). Para esta representación se han considerado los 17 alelos de las Tablas 3.1 para *M. bolleana sencu lato* (ver Tabla 3.2). El valor de representatividad de una sola población fue de  $R = 0,504$  (calculado según se explica en el apartado correspondiente de la brújula).

Tabla 3.2. Probabilidades de pérdida observada ( $L_o$ ) y esperada ( $L_e$ ) para las ocho poblaciones de *Matthiola bolleana*, para las cinco poblaciones de Fuerteventura y las tres de Lanzarote consideradas separadamente. Para cada subdivisión poblacional, N representa el número de poblaciones donde el correspondiente alelo fue detectado. Las probabilidades de pérdida nos permiten decidir en el caso de los alelos *IDH1-a* y *PGM2-a* que la región preferida de muestreo es Lanzarote.

Alelo	Matthiola bolleana s. l.				Fuerteventura				Lanzarote			
	freq	N	$L_o$	$L_e$	freq	N	$L_o$	$L_e$	freq	N	$L_o$	$L_e$
MDH-1b	0,008	2	0,966	0,873	0,014	2	0,947	0,872	-	-	-	-
MDH-2c	0,023	1	0,954	0,684	0,038	1	0,926	0,682	-	-	-	-
MDH-2e	0,026	2	0,900	0,656	-	-	-	-	0,069	2	0,750	0,649
IDH-1a	0,170	2	0,475	0,051	0,272	2	0,281	0,042	-	-	-	-
IDH-1d	0,269	4	0,082	0,007	0,007	1	0,987	0,935	0,706	3	0,001	0,001
IDH-1e	0,090	3	0,568	0,221	-	-	-	-	0,240	3	0,193	0,193
GOT-1b	0,133	4	0,320	0,103	0,212	4	0,148	0,092	-	-	-	-
GOT-1d	0,101	2	0,654	0,183	0,161	2	0,495	0,173	-	-	-	-
GOT-2a	0,065	4	0,586	0,343	0,103	4	0,418	0,336	-	-	-	-
PGM-1a	0,049	2	0,818	0,447	0,079	2	0,721	0,441	-	-	-	-
PGM-1e	0,010	1	0,979	0,846	-	-	-	-	0,028	1	0,945	0,844
PGM-2a	0,035	4	0,751	0,564	0,009	2	1,000	0,915	0,079	2	0,848	0,610
PGM-2d	0,002	1	0,995	0,962	0,004	1	0,992	0,962	-	-	-	-
PGM-3a	0,029	1	0,943	0,623	0,047	1	0,909	0,620	-	-	-	-
PGM-3d	0,006	1	0,988	0,905	0,010	1	0,980	0,904	-	-	-	-
PGD-2a	0,014	2	0,946	0,801	0,035	2	0,869	0,704	-	-	-	-
PGD-2c	0,054	4	0,641	0,410	0,095	4	0,451	0,370	-	-	-	-

La Tabla 3.3 indica los valores de  $A$ ,  $He$  y del promedio de  $F_{ST}$  para cada una de las ocho poblaciones incluidas en el análisis y el ranking resultante de ordenarlas en orden descendente en cada isla de distribución. Según los cálculos anteriores, las cuatro poblaciones a seleccionar para muestreo intensivo en Fuerteventura serían TALA, OVEJ, PSBR y CRUZ. Aunque el ranking para Lanzarote pone a CAB2 en segundo lugar, recordemos que la discusión del ejemplo del apartado anterior optó por incluirla como única representante de Lanzarote, por ser la población cuyo muestreo intensivo permitiría la captura de todos los alelos raros detectados en esta isla. Vemos pues que, como se ha mencionado en varias partes de esta brújula, aunque los criterios cuantitativos explicados son resolutivos en la mayoría de situaciones, en otros casos la decisión final puede depender de factores que estos análisis matemáticos no reflejen o de factores que no estén relacionados con variables moleculares. Por ejemplo, si alguna de estas ocho poblaciones estuviera distribuida en un hábitat ecológico



radicalmente diferente, sería aconsejable muestreala independientemente de lo que resulte de estos análisis cuantitativos, porque podría haber desarrollado diferenciaciones genéticas que no detectemos a nivel molecular.

Tabla 3.3. Ordenación de las poblaciones (por isla) de mayor a menor idoneidad para el muestreo intensivo de variabilidad genética según sus valores de  $A$  y  $H_e$ .

Código	$A$	$H_e$	$F_{ST}$ promedio	Ordenación
Fuerteventura				
TALA	2,5	0,270	0.169	1
OVEJ	2,4	0,294	0.134	2
PSBR	2,3	0,322	0.164	3
CRUZ	2,3	0,287	0.109	4
FAJA	2,0	0,249	0.107	5
Lanzarote				
CAB1	1,6	0,187	0.066	1
CAB2	1,6	0,168	0.092	2
FAMA	1,5	0,128	0.095	3

## EJEMPLO PRÁCTICO DE APLICACIÓN 2

### El endemismo de Gran Canaria *Erysimum albescens* (Brassicaceae)

Vilches et al. (2004) estudian los patrones de variación isoenzimática en ocho poblaciones (Fig. 3.4) que representan el área de distribución del endemismo de Gran Canaria *E. albescens* (las frecuencias de los 26 alelos detectados en los 11 loci resueltos se muestran en la Tabla 3. 4).

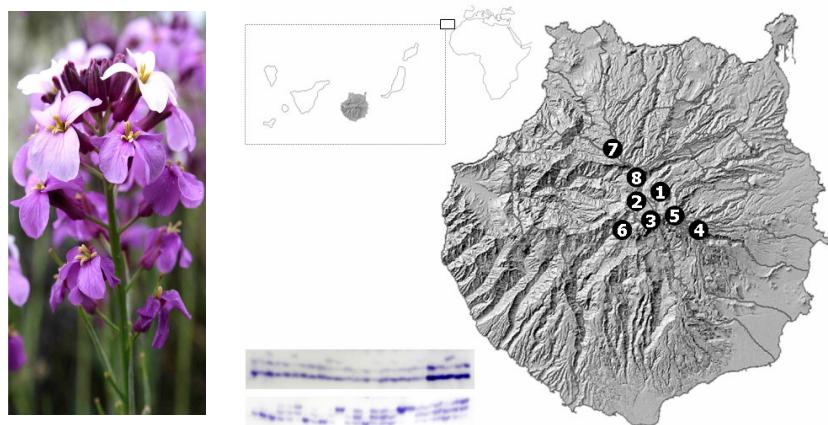


Figura 3.4. Izquierda: detalle de la flor de *Erysimum albescens* (fotografía de Manuel Quevedo, JBCVC). Derecha: localización espacial de las ocho poblaciones de *Erysimum albescens* muestreadas en Gran Canaria por Vilches et al (2004) y fotografías de la resolución de la enzima PGM (orfoglucumotasa, arriba) y de la resolución de la enzima PGI (fosfoglucosomerasa, debajo).

El valor de  $G_{ST}$  (Nei 1973) es muy bajo en las poblaciones de *Erysimum* analizadas ( $G_{ST} = 0.120$ ), lo cual indica una gran homogeneidad genética. La aplicación de la fórmula  $P = (1 - G_{ST})^n$  (Hamrick et al. 1991) para estimar el número de poblaciones ( $n$ ) necesario para representar una proporción del 99% ( $P = 0.99$ ) de la variación isoenzimática detectada entre poblaciones, estimó que tres de las ocho poblaciones analizadas serían suficientes para representar la diversidad genética de este endemismo. A la luz de la aparente consistencia de este resultado, y atendiendo a los indicadores de polimorfismo por población, se seleccionaron las poblaciones 3, 4 y 8 para recolecciones intensivas de semillas. Si revisamos la Tabla 3.4 veremos que entre estas tres poblaciones podríamos capturar el 100% de los alelos detectados.



Tabla 3.4. Frecuencias génicas de los alelos detectados en las 8 poblaciones de *Erysimum albescens* muestreadas por Vilches et al. (2004). Los códigos y frecuencias en negrita y cursiva se señalan a las poblaciones seleccionadas para muestreo intensivo.

Alelo	Poblaciones							
	1	2	3	4	5	6	7	8
PGM-1a	0,433	0,533	<b>0,500</b>	<b>0,500</b>	0,950	0,548	0,455	<b>0,522</b>
PGM-1b	0,567	0,467	<b>0,500</b>	<b>0,500</b>	0,050	0,452	0,545	<b>0,478</b>
PGM-2a	0,000	0,000	<b>0,000</b>	<b>0,014</b>	0,033	0,000	0,000	<b>0,196</b>
PGM-2b	0,733	0,833	<b>0,880</b>	<b>0,943</b>	0,967	1,000	1,000	<b>0,804</b>
PGM-2c	0,267	0,167	<b>0,100</b>	<b>0,043</b>	0,000	0,000	0,000	<b>0,000</b>
PGM-3a	0,000	0,000	<b>0,000</b>	<b>0,000</b>	0,000	0,000	0,000	<b>0,000</b>
PGM-3b	0,933	1,000	<b>1,000</b>	<b>1,000</b>	1,000	1,000	1,000	<b>1,000</b>
PGM-3c	0,067	0,000	<b>0,000</b>	<b>0,000</b>	0,000	0,000	0,000	<b>0,000</b>
IDH-1a	1,000	1,000	<b>1,000</b>	<b>1,000</b>	1,000	1,000	1,000	<b>1,000</b>
PGI-1a	0,400	0,300	<b>0,280</b>	<b>0,129</b>	0,233	0,043	0,104	<b>0,522</b>
PGI-1a b	0,600	0,700	<b>0,720</b>	<b>0,871</b>	0,767	0,957	0,896	<b>0,478</b>
MDH-1a	0,000	0,000	<b>0,000</b>	<b>0,000</b>	0,000	0,000	0,000	<b>0,000</b>
MDH-1b	0,000	0,000	<b>0,040</b>	<b>0,043</b>	0,400	0,239	0,174	<b>0,341</b>
MDH-1c	0,000	0,000	<b>0,000</b>	<b>0,000</b>	0,000	0,000	0,000	<b>0,045</b>
MDH-1d	0,733	0,700	<b>0,640</b>	<b>0,957</b>	0,367	0,500	0,457	<b>0,295</b>
MDH-1e	0,267	0,300	<b>0,320</b>	<b>0,000</b>	0,233	0,261	0,370	<b>0,318</b>
MDH-2a	0,000	0,000	<b>0,000</b>	<b>0,000</b>	0,000	0,000	0,000	<b>0,000</b>
MDH-2b	1,000	1,000	<b>0,880</b>	<b>0,957</b>	0,900	0,826	0,870	<b>0,957</b>
MDH-2c	0,000	0,000	<b>0,120</b>	<b>0,043</b>	0,100	0,174	0,130	<b>0,043</b>
MDH-3a	0,867	0,640	<b>0,980</b>	<b>1,000</b>	1,000	0,978	0,957	<b>0,727</b>
MDH-3b	0,133	0,360	<b>0,020</b>	<b>0,000</b>	0,000	0,022	0,043	<b>0,273</b>
6PGD-1a	1,000	1,000	<b>1,000</b>	<b>1,000</b>	1,000	1,000	1,000	<b>0,905</b>
6PGD-1b	0,000	0,000	<b>0,000</b>	<b>0,000</b>	0,000	0,000	0,000	<b>0,095</b>
6PGD-2a	1,000	1,000	<b>1,000</b>	<b>1,000</b>	1,000	1,000	1,000	<b>0,905</b>
6PGD-2b	0,000	0,000	<b>0,000</b>	<b>0,000</b>	0,000	0,000	0,000	<b>0,095</b>
6PGD-3a	1,000	1,000	<b>1,000</b>	<b>1,000</b>	1,000	1,000	1,000	<b>1,000</b>

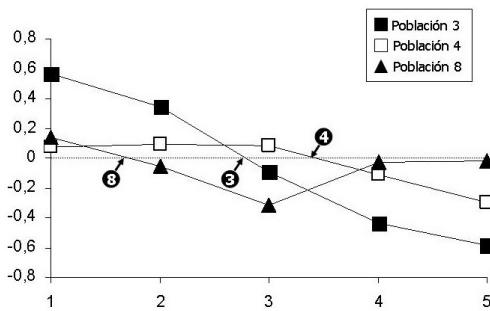


Figura 3.5. Correlogramas de las poblaciones 3, 4 y 8. Las flechas con el código de la población señalan las correspondientes intercepciones de cada correlograma con la abscisa, a partir de las que se estimaron las áreas genéticamente homogéneas (o vecindarios genéticos).

Después de seleccionar los alelos que cumplían los requisitos estipulados en el texto (ver Tablas 3.5, 3.6 y 3.7) se efectuaron análisis de autocorrelación espacial para estas tres poblaciones (Tabla 3.8). Según las estimaciones de los puntos de intercepción de los correlogramas significativos (Fig. 3.5), la probabilidad de muestrear individuos emparentados en estas poblaciones es máxima en un rango de 6,03 m (en 8), 14,76 metros (en 3) y 27,06 m (en 4).

Puesto que el punto de intercepción para los correlogramas significativos es más conservativo (esto es, resulta en distancias más grandes), parece aconsejable ajustarse a ésta estimación. Aunque esta aproximación ha sido criticada por su sujeción a cierta variación estocástica y estadística (Epperson 1990), presenta utilidad para la conservación *ex situ* porque provee un valor operativo del tamaño de los vecindarios genéticos intrapoblacionales (esto es, las regiones donde la probabilidad de encontrar individuos emparentados es máxima). La aplicación práctica de este resultado es que solamente muestreando semillas en individuos separados por estas distancias estaremos en condiciones de asegurar que las muestras recolectadas representan la diversidad genética de las poblaciones. Con ello evitaremos problemas asociados a una eventual depresión por consanguinidad cuando estas semillas sean utilizadas para reintroducciones o reforzamientos poblacionales.



Tabla 3.5. Codificaciones de los alelos incluidos en el análisis de autocorrelación espacial para cada uno de los 23 individuos muestreados en la población 8. X e Y son las coordenadas (en metros) asignadas a cada individuo según el mapa del muestreo. Las codificaciones de los alelos corresponden a 0.5: individuos heterocigotos para el alelo; 1: individuos homocigotos para el alelo; 0: individuos que carecen del alelo. El código -99 designa las observaciones faltantes.

Coordenadas		Alelos incluidos en el análisis									
X	Y	PGM-1a	PGM-2b	PGI-2a	MDH-1a	MDH-1c	MDH-1d	MDH-2b	MDH-3b	6PGD-1b	6PGD-2b
25,0	8,0	0.5	1	0.5	0	0.5	0.5	0	0	-99	-99
22,0	7,8	0.5	1	0.5	0	0.5	0.5	0	0	1	1
20,3	7,5	0.5	1	0.5	0.5	0.5	0	0.5	0.5	0	0
18,1	7,4	0.5	0.5	0.5	0.5	0.5	0	0	0.5	0	0
15,0	7,2	0.5	1	0.5	0.5	0.5	0	0	0.5	0	0
14,2	6,9	0	1	0.5	1	0	0	0	0.5	0	0
12,2	6,6	0.5	1	0.5	0.5	0.5	0	0	0.5	0	0
11,0	6,6	0.5	1	0.5	0.5	0.5	0	0	0	0	0
9,3	5,8	0.5	1	0.5	0	0	1	0	0	0	0
6,8	5,8	0.5	1	0.5	1	0	0	0	0	0	0
5,2	5,8	0.5	0	0	0.5	0.5	0	0	0.5	0	0
3,4	5,2	0.5	0.5	0.5	1	0	0	0	0	-99	-99
2,8	5,0	0.5	0.5	0.5	-99	-99	-99	0	-99	1	1
1,3	4,3	0.5	1	0.5	0	0	0	0	0.5	0	0
0,0	2,1	0.5	1	0.5	0	0	1	0	0.5	0	0
3,2	2,0	0.5	0.5	0.5	0.5	0.5	0	0	0.5	0	0
5,6	1,5	0.5	0.5	1	0	0	1	0	0.5	0	0
7,1	1,2	1	0.5	0.5	0.5	0.5	0	0	0.5	0	0
8,0	0,8	1	1	0.5	0.5	0.5	0	0.5	0	0	0
10,3	0,6	0.5	1	0.5	0	1	0	0	0	0	0
12,0	0,4	0.5	0.5	0.5	0	0	1	0	0.5	0	0
14,0	0,2	0.5	1	0.5	0	0	1	0	0	0	0
15,8	0,0	0.5	1	1	0	0	1	0	0	0	0

Tabla 3.6. Codificaciones de los alelos incluidos en el análisis de autocorrelación espacial para cada uno de los 25 individuos muestreados en la población 3. X e Y son las coordenadas (en metros) asignadas a cada individuo según el mapa del muestreo. Las codificaciones de los alelos corresponden a 0.5: individuos heterocigotos para el alelo; 1: individuos homocigotos para el alelo; 0: individuos que carecen del alelo. El código -99 designa las observaciones faltantes.

Coordenadas		Alelos incluidos en el análisis									
X	Y	PGM-1a	PGM-2b	PGM-2c	PGI-2b	MDH-1a	MDH-1c	MDH-1d	MDH-2b	MDH-3a	
8,60	30,00	0.5	1	0	0.5	0	1	0	0	0	1
7,50	29,20	0.5	1	0	0.5	0	0	1	0	0	1
7,00	28,00	0.5	1	0	0.5	0	0	1	0	0	1
6,50	27,40	0.5	1	0	0.5	0	0	1	0	0	1
6,30	27,00	0	1	0	1	0	0	1	0	0	1
5,00	26,50	0.5	1	0	1	0	0	1	0	0	1
3,80	26,00	0.5	1	0	0.5	0	0	1	0	0	1
2,60	24,00	0.5	1	0	0.5	0	0	1	0	0	1
1,50	23,00	0.5	0	1	0.5	0	0	1	0	0	1
1,00	22,40	1	1	0	0.5	0	1	0	0	0	1
0,00	21,00	0.5	1	0	0.5	0	1	0	0	0	1
0,50	18,00	0.5	1	0	0.5	0.5	0.5	0	0.5	0	1
0,45	15,00	0.5	0.5	0.5	0.5	0	1	0	0	0	1
0,65	12,00	0.5	1	0	0.5	0	1	0	0	0	1
1,40	10,20	0.5	0.5	0.5	1	0	1	0	0.5	0	1
2,00	8,60	0.5	0.5	0.5	1	0	1	0	0	0	1
2,30	7,20	0.5	1	0	0.5	0	1	0	0	0	1
4,00	7,00	0.5	1	0	1	0	1	0	1	0	1
4,40	6,50	0.5	1	0	1	0.5	0.5	0	0	0	1
5,10	5,00	0.5	0.5	0	0.5	0	1	0	0	0	1
5,25	4,00	0.5	1	0	1	0	1	0	0	0	1
6,00	3,00	0.5	1	0	1	0	1	0	0	0	1
6,30	2,40	0.5	1	0	1	0	1	0	0	0	0.5
6,80	1,50	0.5	1	0	1	0	1	0	0.5	0	1
7,00	0,00	0.5	1	0	1	0	1	0	0.5	0	1



Tabla 3.7. Codificaciones de los alelos incluidos en el análisis de autcorrelación espacial para cada uno de los 35 individuos muestreados en la población 2. X e Y son las coordenadas (en metros) asignadas a cada individuo según el mapa del muestreo. Las codificaciones de los alelos corresponden a 0.5: individuos heterocigotos para el alelo; 1: individuos homocigotos para el alelo; 0: individuos que carecen del alelo. El código -99 designa las observaciones faltantes.

Coordenadas		Alelos incluidos en el análisis					
X	Y	PGM-2a	PGM-2b	PGM-2c	PGI-2b	MDH-1c	MDH-2b
6,0	0	0	1	0	0.5	1	0.5
6,0	2	0	1	0	1	1	0
6,0	4	0	1	0	0.5	1	0.5
6,0	6	0	1	0	0.5	1	0
6,0	8	0	1	0	0.5	0.5	0
6,0	10	0	0.5	0.5	1	1	0
6,0	12	0	1	0	1	1	0
6,0	14	0	1	0	1	1	0.5
6,0	16	0	1	0	1	1	0
6,0	18	0	1	0	1	1	0
6,0	20	0	1	0	0.5	1	0
6,0	22	0	1	0	1	1	0
5,6	24	0.5	0.5	0	1	1	0
5,6	26	0	1	0	1	1	0
5,6	28	0	1	0	1	1	0
5,6	30	0	1	0	0.5	1	0
5,6	32	0	1	0	1	1	0
5,6	34	0	1	0	1	1	0
5,6	36	0	0.5	0.5	1	1	0
5,6	38	0	1	0	1	0.5	0
5,6	40	0	1	0	1	0.5	0
5,6	42	0	1	0	1	1	0
5,6	44	0	1	0	0.5	1	0
5,6	50	0	1	0	1	1	0
0,0	35	0	1	0	0.5	1	0
0,0	37	0	1	0	1	1	0
0,0	39	0	1	0	0.5	1	0
0,0	41	0	1	0	1	1	0
0,0	43	0	1	0	1	1	0
0,0	45	0	1	0	1	1	0
0,0	47	0	1	0	1	1	0
0,0	49	0	1	0	1	1	0
0,0	51	0	0.5	0.5	1	1	0
0,0	53	0	1	0	1	1	0
0,0	55	0	1	0	1	1	0

Tabla 3.8. Estadísticos básicos de autocorrelación espacial en las tres poblaciones de *E. albescens* seleccionadas para muestreo intensivo de semillas (a partir de Vilches et al. 2004). Los números entre paréntesis al lado de los códigos de las poblaciones representan el número de alelos sometidos a autocorrelación y el número de alelos que resultó en correlogramas significativos, respectivamente. "Promedio significativo" se refiere al promedio de los valores de la I de Moran para los alelos cuyo correlograma resultó significativo. "Promedio total" se refiere al promedio de los valores de la I de Moran para todos los alelos sometidos a análisis. "Distancia límite" es el extremo superior del intervalo correspondiente a cada una de las cinco clases de distancia consideradas.

Población	Clases de Distancia				
	1	2	3	4	5
<b>Población 8 (9, 1)</b>					
Promedio significativo	0,14	-0,06	-0,32	-0,03	-0,02
Promedio total	0,01	-0,06	-0,12	-0,06	-0,00
Distancia límite (m)	4	7	9	10	26
<b>Población 3 (9, 3)</b>					
Promedio significativo	0,57	0,34	-0,09	-0,44	-0,59
Promedio total	0,17	0,09	-0,04	-0,22	-0,21
Distancia límite (m)	4	9	16	21	30
<b>Población 2 (6, 1)</b>					
Promedio significativo	0,07	0,09	0,08	-0,11	-0,30
Promedio total	-0,03	-0,03	-0,01	-0,00	-0,08
Distancia límite (m)	8	12	20	32	55



#### 4. UNA APLICACIÓN PARA RESOLVER CASOS TAXONÓMICOS CONFLICTIVOS EN AUSENCIA DE CARACTERES MORFOLÓGICOS DISCRIMINANTES

---



## INTRODUCCIÓN

Casi cualquier biólogo, y algún no-biólogo, sabe que la taxonomía es la disciplina científica que se ocupa de clasificar los seres vivos según una jerarquía biológica (phylum, familia, género, ....) y de asignarles nombres que los identifiquen. Es una definición muy general para una ciencia tan compleja, pero que considero válida a los efectos de esta parte de la brújula. Para llevar a cabo su cometido, la taxonomía se ha basado hasta ahora en caracteres morfológicos, normalmente macroscópicos (forma y tamaño de hojas, número y color de los pétalos, posición del ovario...), a veces microscópicos (número de colpos del grano de polen, estructura de los pelos, morfología de los estomas...), pero siempre observables *de visu*.

Ciertamente, los caracteres morfológicos ofrecen variación que se ha mostrado útil para evaluar las relaciones evolutivas o para establecer categorías taxonómicas en muchos linajes vegetales. Pero esta variabilidad es difícil de tratar estadísticamente porque es en su mayor parte atribuible a la acción combinada de muchos genes diferentes (Doebley et al., 1990; Koornneef et al., 2004), cuya herencia y contribución exacta a fenotipo son desconocidas. Uno de los problemas adicionales que frecuentemente encuentra la taxonomía clásica en los niveles inferiores de la jerarquía biológica (especies, subespecies, variedades, poblaciones) es que, unas veces, los organismos se parecen demasiado y, otras veces, son demasiado variables. En ambos casos, los caracteres morfológicos generan confusión en lugar de consenso. En la primera situación, porque no son discriminantes; en la segunda, porque son de interpretación ambigua y dan pábulo a la subjetividad de los taxónomos. Además, la variabilidad inducida por el ambiente (que es adaptativa y no genética) puede confundirse fácilmente con la basada en caracteres heredables (esto es, la sistemáticamente significativa). Esta confusión a menudo menoscaba el valor de las inferencias taxonómicas basadas solamente en características morfológicas. Sin tener en cuenta estas limitaciones intrínsecas, cuando la morfología no provee caracteres diagnósticos sin ambigüedades la clasificación genera controversias entre taxónomos que hacen difícil la delimitación de las Unidades Taxonómicas Operativas.



Prestigioso locutor español

Estas brevísimas consideraciones son solamente manifestaciones superficiales del gran problema de fondo: nuestro desconocimiento de la naturaleza de la realidad. Las cosas raramente son lo que nos parecen, y comparar con objetividad es en realidad un difícil ejercicio. O si no, que se lo digan al prestigioso locutor español que, con sus modulaciones de voz en un noticiero de máxima audiencia, parecía querer interpretar el estupor popular al saber que, según recientes datos moleculares basados en la secuencia completa de los genomas, el hombre se parece mucho más al mono que el ratón a la rata.

A poco que reflexionemos, deberemos admitir que, seamos prestigiosos, locutores, o descalzos peatones de la existencia, cuando opinamos que algo "se parece" a otro algo (o no), decimos lo que nos parece sin saber en realidad de qué estamos hablando. Después de todo, tal vez nuestros sentidos no estén hechos para comprender la realidad, sino solamente para permitirnos interpretarla de una manera miope, pero suficiente para sobrevivir en ella.

Pero entonces... ¿es posible establecer comparaciones relevantes? El ejemplo del ratón y la rata es solamente una de las últimas anotaciones de una lista antiquísima de evidencias que apuntan a que la realidad raramente se corresponde con lo que vemos. Por lo tanto, y por paradójico que parezca, es bien realista considerar la posibilidad de que algunas de las claves para comprender la realidad puedan tener que buscarse en características invisibles.

Este capítulo de la brújula explica cómo en el campo de la taxonomía vegetal, los (invisibles) marcadores moleculares poblacionales pueden ofrecernos datos comparativos válidos en casos donde la morfología, por la razón que sea, no se revela suficientemente discriminatoria. Como ilustración, describo un método para utilizar la información molecular poblacional en la resolución de problemas taxonómicos en los niveles bajos de la jerarquía taxonómica (especies, subespecies, variedades...) que



he desarrollado en el Departamento de Biodiversidad Molecular del Jardín Botánico Canario "Viera y Clavijo". Las consideraciones teóricas (sintetizadas en el organigrama de la Figura IV.1) se ejemplifican con el caso del endemismo de Gran Canaria *Lotus kunkelii* (Fabaceae), que ha representado la primera aplicación del método (Oliva-Tejera et al. 2006).

Aunque esta brújula está dedicada solamente a técnicas moleculares en genética de poblaciones vegetales, cabe mencionar en este apartado que las secuencias de ADN pueden representar también una gran ayuda para clarificar situaciones taxonómicas conflictivas, porque se mantienen invariables a lo largo de la vida de un organismo desde su nacimiento hasta incluso después de su muerte (siempre que el estado de conservación haya sido adecuado) y, por tanto, permiten la identificación en cualquier estadio del desarrollo, o solamente a través de una las partes del organismo. Además, sabemos con exactitud que analizamos las mismas regiones del genoma en todos los organismos considerados y podemos establecer las comparaciones nucleótido a nucleótido después de haber alineado las secuencias (en virtud de lo que se denomina "homología posicional"), con lo cual garantizamos la total validez científica de las comparaciones. Estas ventajas de las secuencias del ADN respecto de las variables morfológicas apoyan la idea de que podría ser posible utilizar la información molecular para identificar a todas las especies de organismos vivos. En esencia, un código de barras de ADN consiste en el uso de una secuencia corta y estandarizada del genoma como herramienta molecular precisa y fiable para ayudar a caracterizar, distinguir o descubrir especies y para asignar individuos (o partes de individuos no identificadas) a sus especies correctas.

Actualmente, el Departamento de Biodiversidad Molecular, el Banco de ADN, el Herbario y el Departamento de Flora Amenazada del JBCV están investigando la utilidad de cinco regiones de la molécula del DNA cloroplástico para elaborar un código de barras para el Parque Nacional de Garajonay, en colaboración con grupos del Jardín de Aclimatación de La Orotava, de la Universidad de La Laguna, y de la Universidad de Murcia. Para más información sobre el tema del código de barras molecular, puede verse el sitio web del Consorcio para el Código de Barras de la Vida (<http://barcoding.si.edu/>) y el Manual del Banco de ADN del Jardín Botánico Canario "Viera y Clavijo" (Caujapé-Castells et al. 2006).



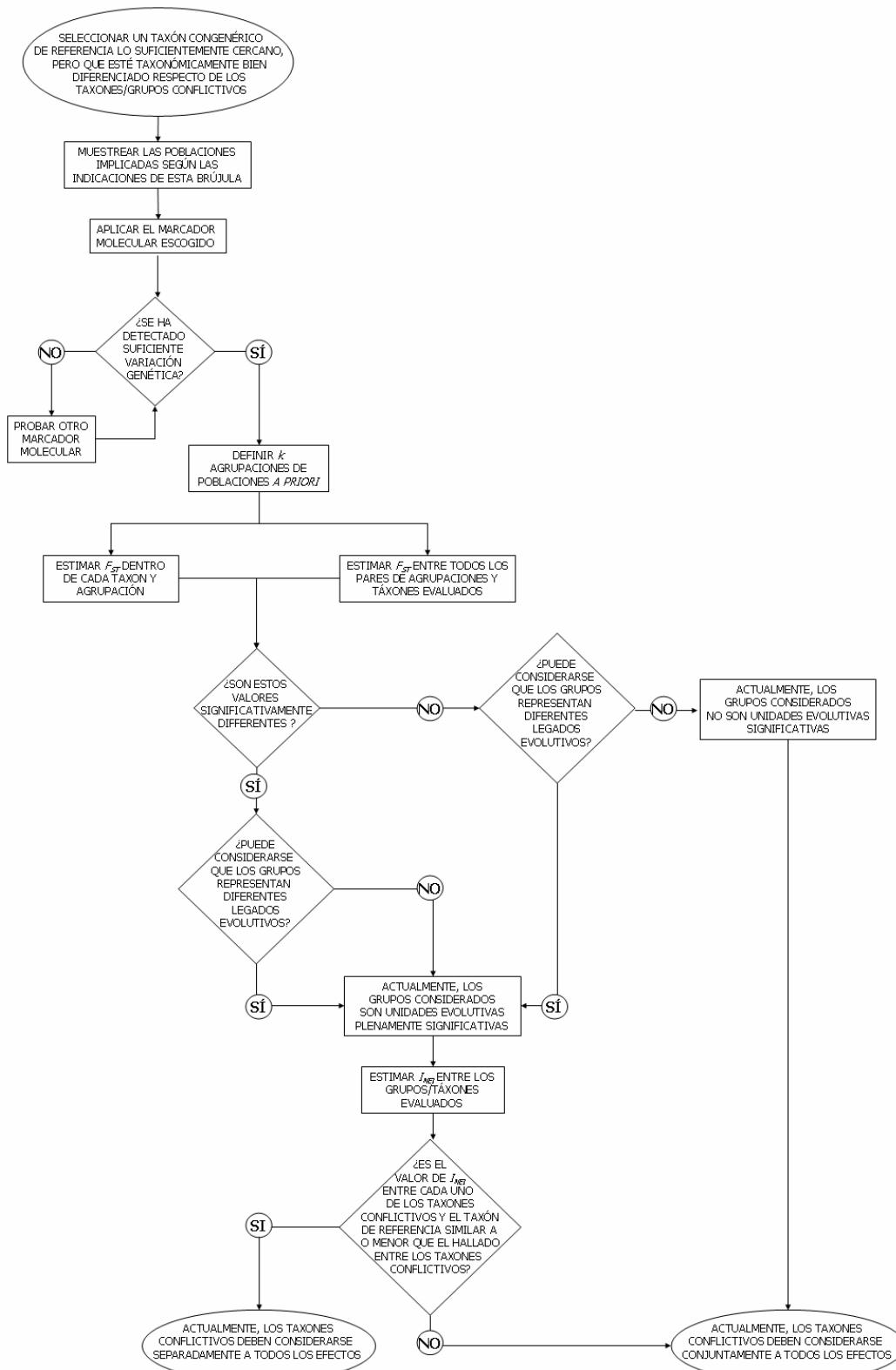


Figura 4.1. Organigrama de decisiones para contribuir a resolver situaciones taxonómicas confusas en niveles bajos de la jerarquía taxonómica mediante algunos de los parámetros usados en genética de poblaciones cuando los caracteres morfológicos no son resolutivos.



## Generalidades

Para aplicar la metodología esquematizada en la Fig. 4.1, necesitamos partir de los siguientes tres elementos:

1. **Dos taxones conflictivos** a los que, por las causas que sean, los caracteres morfológicos no logran discriminar o asociar de manera resolutiva.
2. **Un taxón de referencia** congenérico a los dos cuya taxonomía intentamos resolver, pero que no presente problemas de discriminación con ninguno de ellos.
3. **El concepto de Unidad Evolutivamente Siginificativa (UES)**, definido por Waples (1991, 1995) como una población o grupo de poblaciones que (i) está(n) reproductivamente aislada(s) de otra(s) población(es) con-específicas, y (ii) representa(n) un componente importante del legado evolutivo de una especie. No obstante, para los fines que nos ocupan en este capítulo es mejor redefinir la condición (i) de la siguiente forma: (i') está(n) reproductivamente aislada(s) de otra(s) población(es) que la taxonomía clásica no puede distinguir mediante los caracteres morfológicos.

Puesto que nuestro campo de actuación es la biología comparada, hemos de establecer un calibre para las comparaciones, y de ahí la necesidad de utilizar un taxón de referencia para verificar si las diferencias entre los conflictivos son asimilables a las que cada uno de ellos presenta con aquel. La necesidad de redefinir ligeramente la condición (i) de las UES se debe a que la definición original se aplicaba para seleccionar unidades de conservación dentro de una especie (Moritz, 1994; Karl and Bowen, 1999).

El razonamiento que subyace a la utilización de la noción de UES con fines taxonómicos es que solamente si los dos taxones conflictivos son UES se justifica evaluar si sus diferencias genéticas pueden tener consecuencias taxonómicas. En caso de que los grupos de poblaciones conflictivos puedan asimilarse a UES distintas, el criterio que utilizaremos para evaluar si pueden ser ascritos a diferentes especies es la distancia de Nei (1978), que da una medida de la diferenciación genética y ha sido utilizada con frecuencia como calibre en diferentes niveles de la jerarquía taxonómica. Para calibrar el valor de la identidad de Nei ( $I_{NEI}$  de ahora en adelante) son de especial relevancia las comparaciones de cada uno de los taxones conflictivos con el taxón de referencia, pero también es adecuado recurrir a valores publicados para otros taxones congenéricos con los que son objeto del conflicto.

Para evaluar el grado de aislamiento reproductivo entre los taxones según el esquema de la Figura 4.1, usamos los valores de  $F_{ST}$  (Wright 1951), porque este parámetro da cuenta de la proporción de variabilidad total que se explica por la diferenciación entre poblaciones y se interpreta normalmente como una estimación de los niveles de flujo génico. En efecto, la teoría de genética de poblaciones predice que los grupos de poblaciones con  $F_{ST}$  más elevadas pueden considerarse genéticamente heterogéneos, presumiblemente como consecuencia de la interrupción del intercambio genético entre las poblaciones constituyentes (Slatkin 1985a, 1985b, 1987, 1994). Por el contrario, se asumen altos niveles de flujo génico en grupos de poblaciones que tienen menores valores de  $F_{ST}$  (Slatkin 1985a).

En el procedimiento que se expone, se utiliza el parámetro  $F_{ST}$  para ver si su valor dentro de cada uno de los taxones conflictivos es diferente, por una parte, del valor resultante de unirlos (como si fueran un solo taxón) y, por otra parte, del valor resultante de agrupar a cada uno de ellos con el taxón de referencia. Si solamente existen diferencias significativas entre la  $F_{ST}$  para cada uno de los dos taxones conflictivos y la de su agrupación con el taxón de referencia, este resultado es ya una primera indicación importante de que los taxones conflictivos no exhiben diferencias genéticas sustanciales entre ellos (esto es,



que considerarlos conjuntamente no implicaría una disrupción de su estructura genética). En cambio, si el valor de  $F_{ST}$  para el grupo compuesto por los dos taxones conflictivos es mucho mayor que el correspondiente a cada uno de ellos por separado, esto significa que el conflicto taxonómico tiene un trasfondo genético a nivel del marcador que estamos analizando.

La segunda parte de la definición de UES es más difícil de evaluar cuantitativamente, porque hay muchas maneras posibles de interpretar el concepto de "legado evolutivo". De acuerdo con Waples (1995), este término se refiere a "la variación genética que es producto de acontecimientos evolutivos pasados y que representa el reservorio del que depende el potencial evolutivo futuro". Dado que los alelos raros pueden conferir capacidades de supervivencia a las poblaciones después de una contingencia ambiental (Schonewald-Cox et al., 1983), son componentes importantes del potencial evolutivo de las poblaciones, y serán nuestros indicadores de "legado evolutivo". Hay que resaltar que el hecho de que detectemos alelos raros mediante una de las técnicas moleculares poblacionales que se emplean actualmente no significa que esos alelos tengan un potencial adaptativo, porque la mayoría de estas técnicas detectan solamente varioación genética "neutral" (esto es, que no viene influida por el ambiente). Así pues, cuando utilizamos los alelos raros para evaluar el legado evolutivo estamos asumiendo tácitamente que las poblaciones que poseen más alelos raros neutrales poseerán también más alelos raros con valor adaptativo, que son los que importan para superar las contingencias ambientales.

En la Figura 4.1, solamente consideraremos que los grupos de poblaciones evaluados no son ESUs si no presentan diferencias significativas en los valores de  $F_{ST}$  y tampoco tienen "legado evolutivo" distintivo. Por el contrario, si cualquiera de estas dos condiciones se verifica, entonces ya los consideraremos como UES, y pasaremos a evaluar las identidades de Nei (1978), que nos indicarán si el grado de diferenciación entre los conflictivos es del mismo rango que entre cada uno de ellos y el taxón de referencia. Solamente en ése caso podremos argüir que los taxones de referencia deben ser considerados como especies diferentes.



## EJEMPLO PRÁCTICO DE APLICACIÓN

### El caso de *Lotus kunkelii* y *L. lancerottensis*

Oliva-Tejera et al. (2006) usan datos aloenzimáticos para investigar el caso de *Lotus kunkelii*, un endemismo exclusivo de la isla de Gran Canaria cuya identidad taxonómica basada en caracteres morfológicos era objeto de controversia porque algunos autores lo consideraban como una especie bien delimitada (*Lotus kunkelii* (Esteve) Bramwell & Davis) [Bramwell 1972], mientras que otros lo consideraban como una subespecie del endemismo canario *L. lancerottensis* Webb & Berthel., que está ampliamente distribuido en Fuerteventura y Lanzarote (*L. lancerottensis* subsp. *kunkelii*/Esteve) [Esteve Chueca 1972].

Para calibrar el grado de diferenciación genética entre la única población de *L. kunkelii* (1 en la Fig. 4.2) y las de *L. lancerottensis* (2 a 11 en la Fig. 4.2) según la metodología esquematizada en la Fig. IV.1, se consideró como referencia a *L. arinagensis* Bramwell, otro endemismo de la isla de Gran Canaria cercano a los dos taxones conflictivos (pero que está taxonómicamente bien diferenciado de ambos) y que tiene una distribución moderadamente amplia, con tres poblaciones conocidas (12, 13 y 14 en la Fig. 4.2).

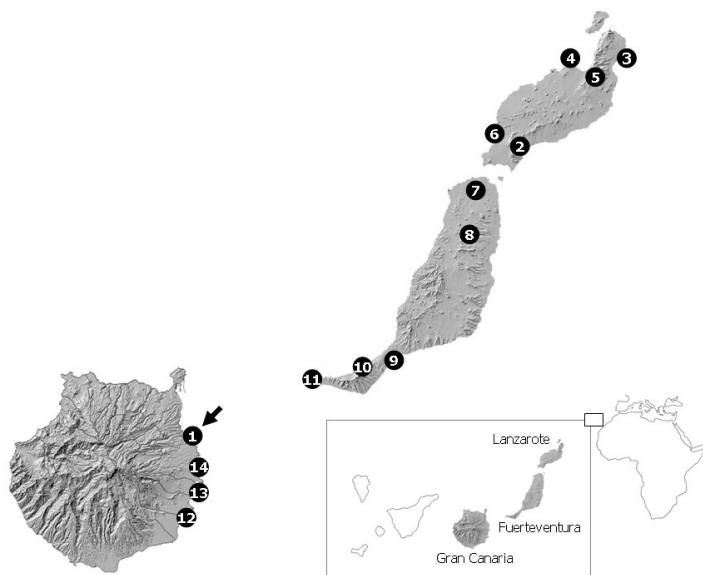


Figura 4.2. Muestreo de *Lotus arinagensis*, *L. kunkelii* y *L. lancerottensis* para la evaluación de las diferencias genéticas entre los dos últimos. Los muestreos representan el área geográfica de distribución de *L. lancerottensis* y a todas las poblaciones conocidas de *L. arinagensis*. La flecha señala la única población conocida de *L. kunkelii*, endemismo amenazado de Gran Canaria.

Siguiendo el protocolo de la Figura 4.1, se establecieron  $k = 3$  agrupaciones artificiales (*L. arinagensis* + *L. kunkelii*, *L. lancerottensis* + *L. kunkelii*, y *L. arinagensis* + *L. lancerottensis*) para evaluar comparativamente cuál era el impacto de la adición de *L. kunkelii* en la estructura genética de los otros dos taxones. Cada una de estas agrupaciones, junto con el valor de  $F_{ST}$  resultante, se considera en la Tabla 4.2.

Examinando la Tabla 4.2 observamos que la consideración de *L. kunkelii* en agrupaciones artificiales con *L. arinagensis* o con *L. lancerottensis* indujo siempre un incremento sustancial en los valores de  $F_{ST}$  ( $F_{ST} = 0.215$  para *L. kunkelii* + *L. lancerottensis*, y  $F_{ST} = 0.162$  para *L. kunkelii* + *L. arinagensis*) respecto de los valores de estos dos taxones sin la adición de *L. kunkelii*. Toda vez que el valor absoluto más elevado de  $F_{ST}$  para estas agrupaciones artificiales correspondió a la unión de *L. kunkelii* con *L. lancerottensis*, el incremento más elevado respecto del valor de  $F_{ST}$  sin *L. kunkelii* correspondió a *L. arinagensis* ( $F_{ST} = 0.067$  para *L. arinagensis* y  $F_{ST} = 0.189$  para *L. lancerottensis*).

Estas evidencias cuantitativas indican que la asimilación taxonómica de *L. kunkelii* con *L. lancerottensis* o *L. arinagensis* provocaría una disrupción de la estructura genética de estos dos taxones. Otra indicación muy importante de estas evaluaciones genéticas es que la diferenciación genética entre *L. kunkelii* y *L. lancerottensis* ( $F_{ST} = 0.215$ ), que protagonizan el conflicto taxonómico, es mayor que la detectada entre *L. lancerottensis* y *L. arinagensis* ( $F_{ST} = 0.205$ ), cuya individualidad taxonómica nunca ha sido objeto de debate.

Aunque no se detectó un "legado evolutivo" sustancial en *L. kunkelii* (solamente un alelo exclusivo, Tabla 4.1) es importante resaltar a este nivel cualitativo que este taxón compartió más alelos con *L. arinagensis* (20 de 21) que con *L. lancerottensis* (15 de 21) (Tablas 4.1 y 4.2). Estos dos resultados genéticos apoyan la consideración de *L. kunkelii* y *L. lancerottensis* como UES independientes.



Tabla 4.1. Valores de las frecuencias alélicas en las 14 poblaciones de *L. kunkelii* (KUN), *L. lacenrottensis* (LAN) y *L. arinagensis* (ARI) investigadas por Oliva-Tejera et al. (2006). Los valores enmarcados corresponden a los alelos exclusivos por taxón. Los números de las poblaciones corresponden a la Fig. 4.2.

Alelos	1	KUN											LAN			ARI		
		2	3	4	5	6	7	8	9	10	11	12	13	14	12	13	14	
<i>Idh1a</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000				
<i>Idh1b</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.050	0.000				
<i>Idh1c</i>	0.286	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.097	0.083	0.111				
<i>Idh1d</i>	0.500	1.000	0.976	1.000	0.972	1.000	1.000	1.000	0.841	0.980		0.857	0.850	0.889				
<i>Idh1e</i>	0.214	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
<i>Idh1f</i>	0.000	0.000	0.024	0.000	0.028	0.000	0.000	0.000	0.000	0.159	0.020	0.020	0.000	0.000				
<i>Est1a</i>	0.000	0.000	0.036	0.039	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
<i>Est1b</i>	0.375	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000				
<i>Est1c</i>	0.250	0.000	0.000	0.000	0.025	0.000	0.000	0.065	0.133	0.000	0.044	0.240	0.000	0.100				
<i>Est1d</i>	0.000	0.000	0.000	0.050	0.075	0.200	0.226	0.100	0.194	0.174		0.206	0.039	0.000				
<i>Est1e</i>	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000				
<i>Est1f</i>	0.313	0.706	0.643	0.000	0.225	0.100	0.233	0.532	0.567	0.722	0.783	0.534	0.769	0.900				
<i>Est1g</i>	0.000	0.294	0.321	0.962	0.700	0.825	0.567	0.177	0.133	0.083	0.000	0.007	0.135	0.000				
<i>Est1h</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.058	0.000				
<i>Pgm1a</i>	0.000	0.025	0.000	0.000	0.022	0.000	0.000	0.000	0.000	0.023	0.040	0.016	0.035	0.100				
<i>Pgm1b</i>	1.000	0.975	1.000	1.000	0.978	1.000	1.000	1.000	0.977	0.960		0.984	0.966	0.900				
<i>Pgm2a</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.041	0.000	0.133				
<i>Pgm2b</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.083		0.000	0.000	0.000	0.000	0.000				
<i>Pgm2c</i>	1.000	0.975	1.000	0.933	1.000	0.929	0.867	0.917	1.000	1.000	0.960	0.954	1.000	0.867				
<i>Pgm2d</i>	0.000	0.025	0.000	0.067	0.000	0.071	0.067	0.000	0.000	0.000	0.040	0.005	0.000	0.000				
<i>Pgm3a</i>	0.000	0.000	0.000	0.000	0.107	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
<i>Pgm3b</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.179				
<i>Pgm3c</i>	0.667	0.923	0.691	0.346	0.179	0.500	0.500	0.473	0.467	0.421	0.458	0.801	0.883	0.714				
<i>Pgm3d</i>	0.250	0.077	0.310	0.654	0.714	0.500	0.500	0.514	0.500	0.579	0.521	0.194	0.017	0.107				
<i>Pgm3e</i>	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.033	0.000	0.021	0.005	0.000	0.000				
<i>6-Pgd1a</i>	1.000	0.816	0.875	1.000	0.833	0.842	0.917	0.944	0.972	0.900	1.000	0.980	0.900	1.000				
<i>6-Pgd1b</i>	0.000	0.184	0.125	0.000	0.167	0.158	0.083	0.056	0.028	0.100	0.000	0.020	0.100	0.000				
<i>6-Pgd2a</i>	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.986	1.000	1.000	1.000	1.000	1.000	1.000				
<i>6-Pgd2b</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.000				
<i>6-Pgd3a</i>	0.667	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.393	0.000				
<i>6-Pgd3b</i>	0.333	0.111	0.231	0.091	0.091	0.133	0.125	0.515	0.000	0.175	0.609	0.378	0.446	0.778				
<i>6-Pgd3c</i>	0.000	0.861	0.692	0.818	0.864	0.800	0.792	0.470	0.912	0.725	0.391	0.555	0.161	0.222				
<i>6-Pgd3d</i>	0.000	0.028	0.077	0.091	0.046	0.067	0.083	0.015	0.088	0.100	0.000	0.043	0.000	0.000				
<i>6-Pgd4a</i>	0.000	0.321	0.563	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.043	0.023	0.000				
<i>6-Pgd4b</i>	1.000	0.679	0.438	0.917	0.211	0.458	0.546	0.530	0.794	0.357	0.820	0.362	0.182	0.667				
<i>6-Pgd4c</i>	0.000	0.000	0.000	0.042	0.447	0.458	0.091	0.000	0.029	0.143	0.020	0.328	0.455	0.333				
<i>6-Pgd4d</i>	0.000	0.000	0.000	0.042	0.342	0.083	0.364	0.470	0.177	0.500	0.160	0.250	0.341	0.000				
<i>6-Pgd4e</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000				
<i>Mdh-1a</i>	0.000	0.632	0.023	0.357	0.152	0.262	0.071	0.324	0.000	0.238	0.020	0.053	0.035	0.000				
<i>Mdh-1b</i>	0.000	0.211	0.364	0.214	0.022	0.429	0.500	0.309	0.222	0.191	0.000	0.181	0.172	0.333				
<i>Mdh-1c</i>	0.750	0.158	0.591	0.357	0.435	0.143	0.429	0.368	0.778	0.571	0.480	0.729	0.552	0.667				
<i>Mdh-1d</i>	0.250	0.000	0.023	0.000	0.065	0.167	0.000	0.000	0.000	0.000	0.020	0.027	0.241	0.000				
<i>Mdh-1e</i>	0.000	0.000	0.000	0.071	0.326	0.000	0.000	0.000	0.000	0.000	0.480	0.011	0.000	0.000				
<i>Mdh-2a</i>	0.357	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.035	0.000				
<i>Mdh-2b</i>	0.000	0.028	0.000	0.067	0.261	0.000	0.094	0.097	0.000	0.068	0.020	0.011	0.000	0.333				
<i>Mdh-2c</i>	0.643	0.944	1.000	0.933	0.739	1.000	0.906	0.903	0.972	0.932	0.900	0.974	0.948	0.667				
<i>Mdh-2d</i>	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.080	0.011	0.017	0.000					



Tabla 4.2. Indicadores de afinidad genética según el método de la Figura 4.1.  $I_{NEI}$  es el promedio de la identidad de Nei (1978). La columna "legado" se refiere al número de alelos exclusivos (cuando se considera un taxón solo) o al número total de alelos detectados en *L. kunkelii* (21) que se comparten con el otro taxón que integra la agrupación artificial. El guión significa que el correspondiente valor no se calculó por presentar poca relevancia para la discusión.

Taxón / Agrupación	$F_{ST}$	legado	$I_{NEI}$
<i>L. arinagensis</i>	0.067	5	0.955
<i>L. arinagensis</i> + <i>L. kunkelii</i>	0.162	20/21	0.886
<i>L. lancerottensis</i>	0.189	5	-
<i>L. lancerottensis</i> + <i>L. kunkelii</i>	0.215	15/21	0.823
<i>L. arinagensis</i> + <i>L. lancerottensis</i>	0.205	-	0.955

Tabla 4.3. Valores de  $I_{NEI}$  (encima de la diagonal) y de las distancias de Nei (1978) (debajo de la diagonal) entre todos los pares posibles de poblaciones (Pobs) de *L. kunkelii* (KUN), *L. lancerottensis* (LAN) y *L. arinagensis* (ARI) investigadas por Oliva-Tejera et al. (2006). Los números de las poblaciones se corresponden con la Figura IV.2. C: Gran Canaria, L: Lanzarote, F: Fuerteventura.

KUN		LAN												ARI			
		L						F									
Pobs	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
	—	0.808	0.845	0.802	0.748	0.770	0.820	0.854	0.868	0.831	0.884	0.885	0.883	0.891			
1	—	0.214	—	0.959	0.896	0.854	0.914	0.923	0.926	0.924	0.920	0.891	0.915	0.882	0.882		
2	0.168	0.042	—	0.901	0.893	0.921	0.956	0.944	0.963	0.952	0.924	0.956	0.918	0.921			
3	0.220	0.109	0.104	—	0.940	0.972	0.970	0.910	0.926	0.890	0.868	0.854	0.785	0.802			
4	0.290	0.157	0.113	0.062	—	0.957	0.956	0.910	0.914	0.935	0.872	0.882	0.833	0.808			
5	0.261	0.090	0.082	0.029	0.044	—	0.981	0.920	0.913	0.910	0.853	0.892	0.854	0.824			
6	0.199	0.080	0.045	0.030	0.045	0.019	—	0.970	0.970	0.964	0.908	0.937	0.879	0.874			
7	0.158	0.077	0.057	0.094	0.094	0.084	0.031	—	0.951	0.981	0.960	0.956	0.918	0.931			
8	0.141	0.079	0.038	0.077	0.090	0.091	0.030	0.050	—	0.975	0.942	0.956	0.880	0.903			
9	0.185	0.083	0.050	0.116	0.068	0.094	0.037	0.019	0.026	—	0.943	0.964	0.918	0.906			
10	0.123	0.115	0.079	0.142	0.137	0.159	0.097	0.041	0.059	0.058	—	0.939	0.901	0.948			
11	0.122	0.088	0.045	0.157	0.126	0.115	0.065	0.045	0.045	0.037	0.063	—	0.963	0.955			
12	0.125	0.125	0.085	0.242	0.182	0.158	0.129	0.086	0.128	0.085	0.104	0.038	—	0.948			
13	0.115	0.126	0.083	0.220	0.213	0.194	0.134	0.072	0.103	0.099	0.053	0.046	0.053	—			

La identidad de Nei ( $I_{NEI}$ ) promedio entre *L. kunkelii* y *L. lancerottensis* ( $I_{NEI} = 0.823$ ) fué mucho menor que la detectada para cualquier otro par de especies Canarias de *Lotus* para las que se dispone de datos aloenzimáticos (por ejemplo,  $I_{NEI} = 0.938$  entre *L. spartoides* y *L. holosericeus*, Oliva-Tejera et al., 2005), e incluso más baja que la  $I_{NEI}$  promedio entre *L. kunkelii* y *L. arinagensis* ( $I_{NEI} = 0.886$ ). Además, los valores mínimos de  $I_{NEI}$  se detectaron siempre entre la población de *L. kunkelii* y algunas de *L. lancerottensis* (Tabla 4.2 y 4.3). Considerados en su globalidad, estos datos indican que el grado de divergencia genética entre *L. lancerottensis* y *L. kunkelii* corresponden a los esperados para diferentes especies Canarias de *Lotus*. Es también relevante el resultado de que el valor de  $I_{NEI}$  entre *L. lancerottensis* y *L. arinagensis* ( $I_{NEI} = 0.955$ ), que son dos especies bien delimitadas, es muy superior al calculado entre *L. lancerottensis* y *L. kunkelii* ( $I_{NEI} = 0.823$ ).

En conclusión, estos resultados indican que el grado de diferenciación evolutiva detectado con aloenzimas entre el taxón amenazado *L. kunkelii* y *L. lancerottensis* es mucho mayor que la detectada entre *L. lancerottensis* y *L. arinagensis* o entre otros *Lotus* canarios de la sección Pedrosia para los que se dispone de datos aloenzimáticos (Oliva-Tejera et al. 2005). Además, *L. kunkelii* es genéticamente más próximo a *L. arinagensis* que a *L. lancerottensis*. Estas conclusiones apoyan la consideración de *L. kunkelii* como una especie bien diferenciada de *L. lancerottensis*, cuya única población conocida debe recibir protección según la nueva metodología para el catálogo de especies canarias amenazadas (Martín-Esquível 2004).



## BIBLIOGRAFIA

- Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160: 1217-1229.
- Avise JC (1974) Systematic value of electrophoretic data. *Syst. Zool.* 23: 465-481.
- Banks MA, Eichert W (2000) WHICHRUN: a computer program for population assignment of individuals based on multilocus genotype data. *Journal of Heredity* 91: 87-89.
- Barbujani G (1987) Autocorrelation of gene frequencies under isolation by distance. *Genetics* 117: 777-782.
- Barrett SC, Kohn JR (1991) Genetic and evolutionary consequences of small population size in plants: implications for conservation. In: D. A. Falk and K. E. Holsinger (eds.) *Genetics and conservation of rare plants*, pp. 3-30. Oxford University Press, New York.
- Bataillon T, Kirkpatrick M (2000) Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genetical Research* 75: 75-81.
- Batista F, Bañares A, Caujapé-Castells J, Marrero-Gómez M, Carqué E, Sosa PA (2001) Allozyme diversity in three endemic species of *Cistus* (Cistaceae) from the Canary islands: Intraspecific and interspecific comparisons and implications for genetic conservation. *American Journal of Botany* 88: 1582-1592.
- Beerli P, Felsenstein J (2001) Maximun likelihood estimation of a migration matrix and effective population sizes in subpopulations by using a coalescent approach. *Proc. Nat. Acad. Sci. USA* 98: 4563-4568.
- Bengtsson BO, Weibull P, Ghatnekar L (1995) The loss of alleles by sampling: a study of the common outbreeding grass *Festuca ovina* over three geographical scales. *Hereditas* 122: 221-238.
- Berg EE, Hamrick JL (1997) Quantification of genetic diversity at allozyme loci. *Canadian Journal of Forest Research* 27: 415-415.
- Bossart JL, Prowell DP (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution* 13: 202-202.
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-457.
- Bramwell D, Davis DD (1972) A contribution to the study of *Lotus* L. on Gran Canaria. *Cuadernos de Botánica Canaria* 16: 51-54.
- Britten HB (1996) Meta-analyses of the association between multilocus heterozygosity and fitness. *Evolution* 50: 2158-2164.
- Brown AHD, Briggs JD (1991) Sampling strategies for genetic variation in Ex situ collections of endangered plant species. Pp. 99-122 In Falk DA, Holsinger K (eds.) *Genetics and Conservation of rare plants*. Oxford University Press, New York.
- Bruvo R, Michiels NK, D'Souza TG, Schulenburg H (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol. Ecol.* 13: 2101-2106.
- Byers DL, Waller DM (1999) Do plant populations purge on their genetic load? Effects of population size and mating history on onbreeding depression. *Annual Review of Ecology and Systematics* 30: 479-513.
- Carson HL (1975) The genetics of speciation at the diploid level. *American Naturalist* 109: 73-92.
- Caujapé-Castells J, Pedrola-Monfort J (1997) Space-time patterns of genetic structure within a stand of *Androcymbium gramineum* (Cav.) McBride (Colchicaceae). *Heredity* 79: 341-349.



- Caujapé-Castells J, Pedrola-Monfort J, Membrives N (1999) Contrasting patterns of genetic structure in the South African species *Androcymbium bellum*, *A. guttatum* and *A. pulchrum* (Colchicaceae). *Biochemical Systematics and Ecology* 27: 591 - 605.
- Caujapé-Castells J (2002) La diversidad genética a las duras y a las maduras: la relación entre *Androcymbium gramineum* (Cav.) McBride (Colchicaceae) y las catástrofes naturales. *Resúmenes del 1<sup>er</sup> Congreso de la biología de la conservación de plantas*, Pp. 86. Universitat de València.
- Caujapé-Castells J (in mss) Issues on ESUs: defining *ex situ* conservation targets based on molecular population genetic information. *Plant Systematics and Evolution* (in mss.).
- Caujapé-Castells J, Jansen RK (2003) The influence of the Miocene Mediterranean desiccation on the geographical expansion and genetic variation of *Androcymbium gramineum* (Cav.) McBride (Colchicaceae). *Molecular Ecology* 12: 1515-1525.
- Caujapé-Castells J, Pedrola-Monfort J (2004) A sampling design for the ex-situ genetic conservation of the Ibero-Moroccan endangered endemic *Androcymbium gramineum*: implications for the assessment of a conservation strategy from a survey of genetic diversity for neutral markers. *Conservation Genetics* 5: 131-144.
- Caujapé-Castells J, Baccarani-Rosas M (2005) Transformer-3: a program for the analysis of molecular population genetic data. EXEGEN software & Jardín Botánico Canario "Viera y Clavijo" (<http://www.exegen.org>).
- Caujapé-Castells J, Jaén-Molina R, Cabrera-García N (2006) Manual del banco de ADN del Jardín Botánico Canario "Viera y Clavijo". Cabildo de Gran Canaria, Las Palmas de Gran Canaria (distribuido por los autores).
- Cavalli-Sforza LL, Bodmer WF (1971) *The Genetics of Human Populations*, p. 399, San Francisco, W.H. Freeman and Company.
- Center for Plant Conservation (1991) Genetic sampling guidelines for conservation collections of endangered plants. Pp. 225-238 In In Falk DA, Holsinger K (eds.) *Genetics and Conservation of rare plants*. Oxford University Press, New York.
- Ceska JF, Affolter JM, Hamrick JL (1997) Developing a sampling strategy for *Baptisia arachnifera* based on allozyme diversity. *Conservation Biology* 11: 1133-1133.
- Chakraborty R, Nei M (1977) Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* 31: 347-356.
- Chamberlain JR (1998) Isozyme variation in *Calliandra calothrysus* (Leguminosae): its implications for species delimitation and conservation. *American Journal of Botany* 85: 37-47.
- Charlesworth B, Charlesworth D (1999) The genetic basis of inbreeding depression. *Genetic Research* 74: 329-340.
- Cockerham CC (1969) Variance of gene frequencies. *Evolution* 23: 72-84.
- Cockerham CC, Weir BS (1986) Estimation of inbreeding parameters in stratified populations. *Annals of Human Genetics* 50: 271-271.
- Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* 163: 367-374.
- Cornuet JM, Luikart G (1996) Description and evaluation of two tests for detecting recent bottlenecks. *Genetics* 144: 2001-2014.
- Cornuet JM, Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001-2014.
- Crow JF (1993) Mutation, mean fitness, and genetic load. Pp. 3-42 in D. Futuyma and J. Antonovics, eds. *Oxford surveys in evolutionary biology*. Oxford University Press, Oxford, UK.
- Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York.



- Culley TM, Wallace LE, Gengler-Nowak KM, Crawford DJ (2002) A comparison of two methods of calculating  $G_{ST}$ , a genetic measure of population differentiation. *American Journal of Botany* 89: 460-465.
- Davenport CB (1908) Degeneration, albinism and inbreeding. *Science* 28: 454-455.
- Dawson KJ, Belkhir K (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genet. Res.* 78: 59-77.
- Doebley J, Stec A, Wendel J, and Edwards M (1990) Genetic and morphological analysis of a maize-teosinte F2 population: implications for the origin of maize. *Proceedings of the National Academy of Sciences, USA* 87: 9888-9892.
- Dorken ME, Eckert CG (2001) Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *Journal of Ecology* 89: 339-350.
- Dubois D, Prade H (1980) *Fuzzy Sets and Systems: Theory and Applications*, p. 24, New York, Academic Press.
- East EM (1908) Inbreeding in corn. Report of the Connecticut Agriculture experimental station 1907: 419-429.
- Elseth GD, Baumgardner KD (1981) Population Biology. Van Nostrand, NY
- Epperson BK (1990) Spatial patterns of genetic variation within plant populations. pp 229-253, In *Plant Population Genetics, Breeding and Genetic Resources* (AHD Brown, MT Clegg, AL Kahler and BS Weir, eds.). Sinauer Assoc. Inc. Sunderland, Massachusetts.
- Esteve-Chueca F (1972) Nuevas referencias a la vegetación litoral de Gran Canaria. *Cuadernos de Botánica Canaria* 14-15: 44.
- Ewens W (1972) The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* 3: 87-112.
- Ewens WJ (1979) Mathematical population genetics. Berlin, Springer.
- Falconer DF, Mackay TFC (1996) Introduction to quantitative genetics (4<sup>a</sup> ed.) Longman, Harlow, UK.
- Fay, MF (2003) Using genetic data to help guide decisions about sampling. Pp. 91-96 in Smith RD, Dickie JB, Linington SH, Pritchard HW, Probert RJ (eds.) *Seed Conservation: turning science into practice*. The Royal Botanic Gardens, Kew. The Cromwell Press Ltd, London.
- Fenster CB, Galloway LF (2000) Population differentiation in an annual legume: genetic architecture. *Evolution* 54: 1157-1172.
- Folch R (1998) *Ambiente, emoción y ética. Actitudes ante la cultura de la sostenibilidad*. Editorial Ariel, Barcelona.
- Ford EB (1965) *Genetic polymorphism*. Faber and Faber, London.
- Fortin M, Gurevitch J (1993) Mantel Tests: Spatial Structure in Field Experiments: *in* Design and Analysis of Ecological Experiments (Scheiner SM, Gurevitch J, eds). Chapman & Hall, New York.
- Frankel O, Brown AHD, Burdon JJ (1995) The conservation of plant biodiversity. Cambridge University Press, Cambridge.
- Francisco-Ortega J, Santos-Guerra A, Kim SC, Crawford DJ (2000) Plant genetic diversity in the Canary Islands: a conservation perspective. *American Journal of Botany* 87: 909-919.
- Gale JS (1990) *Theoretical population genetics*. Unwin Hyman Ltd. London.
- Giles BE, Goudet J (1997) A case study of genetic structure in a metapopulation. In Hanski IA, Gilpin ME (eds) *Metapopulation Biology: Ecology, genetics and evolution*. Academic Press, New York.
- Gillespie JH (1991) *The Causes of Molecular Evolution*. New York: Oxford University Press.
- Gillespie JH (2000a) Genetic Drift in an Infinite Population: The Pseudohitchhiking Model. *Genetics* 155: 909-919.
- Gillespie JH (2000b) The Neutral Theory in an Infinite Population *Gene* 261: 11-18.



- Gillespie JH (2001) Is the Population Size of a Species Relevant to its Evolution?. *Evolution* 55: 2161-2169.
- Gillespie JH (2004) *Population Genetics: A Concise Guide*. Baltimore, MD: Johns Hopkins University Press.
- Gitzendanner MA, Soltis PS (2000) Patterns of genetic variation in rare and widespread plant congeners. *Am. J. Botany* 87: 783-792.
- Goldstein DB, Ruiz-Linares A, Feldman M, Cavalli-Sforza LL (1995a) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences USA* 92: 6720-6727.
- Goldstein DB, Ruiz-Linares A, Feldman M, Cavalli-Sforza LL (1995b) An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139: 463-471.
- Good IJ (1962) The Scientist Speculates. Heinemann & Basic Books, New York (General Editor, AJ Mayne, Assoc. Ed., J Maynard Smith, Biol. Ed.).
- Goudet J (1995) FSTAT version 1.2: a computer program to calculate F-statistics. *Journal of Heredity* 86: 485-486.
- Griffiths DA (1987) Spatial autocorrelation: a primer. Washington, dc: Association of American Geographers.
- Hamrick JL (1983) The distribution of genetic variation within and among natural plant populations. Pp. 335-348 en Chambers SM, Schonewald-Cox CM (eds). *Genetics and Wild Population Management*. Addison-Wesley, New York.
- Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. Pp. 43-63 en Brown AHD, Clegg MT, Kahler AL, Weir BS (eds.) *Plant population genetics, breeding and genetic resources*. Sunderland, Massachusetts: Sinauer.
- Hamrick JL, Godt MJW, Murawski DA, Loveless MD (1991) Correlations between species traits and allozyme diversity: implications for conservation biology. Pp. 75-86 en Falk DA, Holsinger K (eds.) *Genetics and Conservation of rare plants*. Oxford University Press, New York.
- Hardy GH (1908) Mendelian proportions in a mixed population. *Science* 28:49-50.
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618-620.
- Harris H, Hopkinson DA, Robson EB (1974) The incidence of rare alleles determining electrophoretic variants: data on 43 enzyme loci in man. *Ann. Hum. Genet. Lond.* 37: 237-253.
- Hedrick PW (1994) Purgling inbreeding depression and the probability of extinction: full sib-mating. *Heredity* 73: 363-372.
- Hedrick PW (2000) Genetics of Populations. Jones and Bartlett. Boston, MA.
- Hedrick PW (2002) Lethals in finite populations. *Evolution* 56: 654-657.
- Hedrick PW, Savolainen O, Karkkainen K (1998) Factors influencing the extent of inbreeding depression: an example from Scots pine. *Heredity* 82: 441-450.
- Hendrick PW, Kumar S (2001). Mutation and linkage disequilibrium in human mtDNA. *Eur. J. Hum. Genet.* 9: 969-972.
- Heywood JS (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics* 22: 335-355.
- Hubert LJ (1987) *Assignment methods in combinatorial data analysis*. New York: Marcel Dekker.
- Hudson RR (1990) Gene genealogies and the coalescent process, pp. 1-42 in Oxford Survey in Evolutionary Biology, Vol. 7, edited by D. Futuyama and J. Antonovics. Oxford University Press, Oxford.
- Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. *Genetics* 466: 465-479.



- Karl SA, Bowen BE (1999) Evolutionarily significant units versus geopolitical taxonomy: molecular systematics of an endangered sea turtle (genus *Chelonia*). *Conservation Biology* 13: 990-999.
- Kimura M (1953) "Stepping stone" model of population. *Ann. Rept. Nat. Inst. Genetics Japan* 3: 62-63.
- Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561-576.
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217:624-626.
- Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49: 725-738.
- Kimura M, Ohta T (1971). Protein polymorphism as a phase of molecular evolution. *Nature* 229: 467-469.
- Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* 55: 141-172.
- Kumar S, Tamura K, Nei M (1993) MEGA: Molecular Evolutionary Genetics Analysis, Version 1.0, Pennsylvania State University, University Park, Penn., 16802.
- Lande R (1988) Genetics and demography in biological conservation. *Science* 241: 1455-1460.
- Lapointe F-J, Legendre P (1992) Statistical significance of the matrix correlation coefficient for comparing independent phylogenetic trees. *Syst. Biol.* 41: 378-384.
- Lewis PO (1993) GeneStat-PC 3.31. North Carolina State University. Raleigh. North Carolina.
- Lewis PO, Zaykin D (1999) Genetic data analysis: Computer program for the analysis of allelic data. Distribuído por los autores.
- Lewontin RC (1964) The interaction of selection and linkage.I. General considerations: heterotic models. *Genetics* 49: 49-67.
- Lewontin RC (1974) *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.
- Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. *Evolution* 14: 458-472.
- Luikart G (1997) Usefulness of molecular markers for detecting population bottlenecks and monitoring genetic change. Ph. D. Thesis. University of Montana, Missoula, USA.
- Luikart G, Cornuet JM (1998) Empirical evaluation of a test identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* 12: 228-237.
- Luikart G, Cornuet JM (1997) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* 12(1): 228-237.
- Malhotra A, Thorpe RS (1997) Size and shape variation in a Lesser Antillean anole, *Anolis oculatus* (Sauria: Iguanidae) in relation to habitat. *Biological Journal of the Linnean Society* 60: 53-72.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209-220.
- Martín-Esquível JL (2004) Propuesta metodológica para la catalogación de especies amenazadas en Canarias. In J. M. Fernández-Palacios and C. Morici. [eds.], Ecología insular/island ecology, 385-412. Asociación Española de Ecología Terrestre (AEET)- Cabildo Insular de La Palma, Santa Cruz de La Palma, Spain.
- Maruyama T, Fuerst PA (1985) Population bottlenecks and non equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* 111: 675-689.



- Matocq M, Villalba FX (2001) Low genetic diversity in an endangered species: recent or historic pattern? *Biological Conservation* 98: 61-68.
- Maynard Smith J, Haigh J (1974) The Hitch-hiking Effect of a Favourable Gene. *Genetical Research* 23: 23-25.
- Mayr E (1954) Change of genetic environment and evolution. Pp 157-180, *In* J Huxley (ed.) *Evolution as a process*. Allen & Unwin, London.
- Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge, Mass.
- McKersie BD, Senaratna T, Walker MA, Kendall EJ, Hetherington PR (1988) Deterioration of the membrane during aging in plants: evidence for free radical mediation. In Noodeon LD, Leopold AC (eds) *Senescence and aging in plants*, pp. 441-464. Academic Press, San Diego.
- Miller MP (1997) Tools for population genetics analysis (TFPGA) 1.3: A Windows program for the analysis of allozyme and molecular population genetic data.
- Mitton JB, Grant MC (1984) Associations among protein heterozygosity, growth rate, and developmental homeostasis. *Annual Review of Ecology and Systematics* 15: 479-499.
- Moran PAP (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society B* 37: 243- 251.
- Moran PAP (1975) Wandering distributions and the electrophoretic profile. *Theoretical Population Biology* 8: 318-330.
- Moritz C (1994) Defining 'evolutionarily significant units' for conservation. *Trends in Ecology and Evolution* 9: 373-375
- Nei M, Murata Y (1966). Effective population size when fertility is inherited. *Genetical Research* 8: 257-260.
- Nei M (1972) Genetic distance between populations. *American Naturalist* 106: 283-292.
- Nei M (1968) The frequency and distribution of lethal chromosomes in finite populations. *Proc. Natl. Acad. Sci. USA* 60: 517-524.
- Nei M (1969) Heterozygous effects and frequency changes of lethal genes in populations. *Genetics* 63: 669-680.
- Nei M (1973) Analysis of genetic diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA* 70: 3321-3323.
- Nei M (1975) Molecular population genetics and evolution. North Holland, Amsterdam.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.
- Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics* 47: 253-253.
- Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *J. Molec. Evol.* 19: 153-170.
- Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press. New York.
- Nei M (1986) Definition and estimation of fixation indices. *Evolution* 40: 643-643.
- Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Ann. Rev. Ecol. Syst.* 28: 105-128.
- Nickrent DL, Wiens D (1989) Genetic diversity in the rare California shrub *Dedeckera eurekaensis* (Polygonaceae). *Systematic Botany* 14: 245-253.
- Richter TS, Soltis PS, Soltis DE (1994) Genetic variation within and among populations of the narrow endemic, *Delphinium viridescens* (Ranunculaceae). *American Journal of Botany* 81: 1070-1076.
- Oden NL (1984) Assessing the significance of a spatial correlogram. *Geographical Analysis* 16: 1-16.
- Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246: 96-98.
- Ohta T (1974) Mutational pressure as the main cause of molecular evolution and polymorphism. *Nature* 252: 351-354.



- Ohta T, Kimura M (1973a) The age of a neutral mutant persisting in a finite population. *Genetics* 75: 199-212.
- Ohta T, Kimura M (1973b) A model for mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetic Research* 22: 201-204.
- Oliva-Tejera F, Caujapé-Castells J, Naranjo-Suárez J, Navarro-Déniz J, Acebes-Ginovés JR & Bramwell D (2005) Population genetic differentiation in taxa of *Lotus* (Fabaceae: Loteae) endemic to the Gran Canarian pine forest. *Heredity* 94: 199-206.
- Oliva-Tejera F, Caujapé-Castells J, Navarro-Déniz J, Reyes-Betancort A, Scholz S, Baccarani-Rosas M, Cabrera-García N (2006) Patterns of genetic divergence of three Canarian endemic *Lotus* (Fabaceae): implications for the conservation of the endangered *L. kunkelii*. *American Journal of Botany* 93: 1116-1124.
- Peakall R, Smouse PE (2005) GenAIEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. The Australian National University, Canberra, Australia.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GeneClass2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95: 536-539.
- Piry S, Luikart G, Cornuet JM (1999) Bottleneck, a program for detecting recent effective population size reductions from allele frequency data. INRA, URLB, Laboratoire de Modélisation et Biologie Evolutive. Montpellier, France.
- Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* 9: 487-488.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proc. Nat. Acad. Sci. USA* 94: 9197-9201.
- Raymond M, Rousset F (1995) GenePop (version 1.2) population genetic software for exact tests and ecumenicism. *Journal of Heredity* 86: 248-249. (<http://www.cefe.cnrs-mop.fr/>).
- Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105: 767-779.
- Robbins RB (1918) Some applications of mathematics to breeding problems III. *Genetics* 3: 375-389.
- Roff DA (1997) Evolutionary quantitative genetics. Chapman and Hall, New York.
- Roff DA (2002) Inbreeding depression: tests of the overdominance and partial dominance hypotheses. *Evolution* 56: 768-775.
- Roff DA, Bentzen P (1989) The statistical analysis of mitochondrial DNA polymorphisms: X2 and the problem of small samples, *Mol. Biol. Evol.*, 6: 539-545.
- Rohlf FJ (1998). NTSYS-pc. Numerical Taxonomy and Multivariate Analysis version 2.02j. Applied Biostatistics Inc.
- Rosenberg MS (2002) PASSAGE: Pattern Analysis, Spatial Statistics, and Geographic Exegesis. Version 1.0. Tempe, AZ: Department of Biology, Arizona State University. (<http://www.public.asu.edu/~mrosenb/PASSAGE/>).
- Salamon H, Klitz W, Easteal S, Gao X, Erlich HA, Fernandez-Vina M, Trachtenberg EA (1999) Evolution of HLA class II molecules: Allelic and amino acid site variability across populations. *Genetics* 152: 393-400.
- Sánchez JL, Reyes-Betancort JA, Scholz S, Caujapé-Castells J (2004) Patrones de variación genética poblacional en el endemismo canario *Matthiola bolleana* Webb ex Christ. *Botánica Macaronésica* 25: 1-11.



- Savolainen V, Cuénoud P, Spichiger R, Martínez MDP, Crèvecoeur M, Manen JF (1995) The use of herbarium specimens in ADN phylogenetics: evaluation and improvement. *Pl. Syst. Evol.* 197: 87-98.
- Schneider S, Roessli D, and Excoffier L (1997) Arlequin: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Schonewald-Cox CM, Chambers SM, MacBride B, Thomas WL, editors (1983) *Genetics and conservation: a reference for managing wild animal and plant populations* Benjamin Cummings, Menlo Park, CA.
- Scott ME, Possingham JV (1983) Changes in chloroplast ADN levels during growth of spinach leaves. *J. Exp. Bot.* 34: 1756-1767.
- Selander RK (1976) Genetic variation in natural populations. Pp. 21-45 *In* Ayala FJ (ed.) *Molecular Evolution*. Sinauer Assoc., Inc., Sunderland MA.
- Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* 12: 914-920.
- Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor. Pop. Biol.* 12: 253-262.
- Slatkin M (1982) Testing neutrality in subdivided populations. *Genetics* 100: 533-645.
- Slatkin M (1985a) Rare Alleles as Indicators of Gene Flow. *Evolution* 39: 53-65.
- Slatkin M (1985b) Gene flow in natural populations. *Ann. Rev. Ecol. Syst.* 16: 393-430.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236: 787-792.
- Slatkin M (1994) An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research* 64: 71-74.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.
- Slatkin M (1996) A correction to the exact test based on the Ewens sampling distribution. *Genetical Research* 68: 259-260.
- Slatkin M, Barton NH (1989) A comparision of three methods for stimating average levels of gene flow. *Evolution* 43: 1349-1368.
- Slatkin M, Maddison WP (1990) Detecting isolation by distance using phylogenies of genes. *Genetics* 126: 249-260.
- Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* 35: 627-632.
- Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution* 206: 353-373.
- Soltis PS, Gitzendanner MA (1999) Molecular systematics and the conservation of rare species. *Conservation Biology* 13: 471-483.
- Sokal RR (1979) Testing statistical significance of geographical variation patterns. *Systematic Zoology* 8: 227-232.
- Stebbins JL (1980) Rarity of plant species: a synthetic viewpoint. *Rhodora* 82: 77-86.
- Swofford DL, Selander RB (1989) BIOSYS-1: a computer program for the analysis of allelic variation in genetics. University of Illinois, Urbana, Ill.
- Templeton AR (1979) The unit of selection in *Drosophila mercatorum* II. Genetic revolutions and the evolution of coadapted genomes in parthenogenetic strains. *Genetics* 92: 1265-1282.
- Templeton AR (1980) The theory of speciation *via* the founder principle. *Genetics* 94: 1011-1038.
- Templeton AR, Shaw K, Routman E, Davis SK (1991) The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden* 77: 13-27.



- Templeton AR (1986) Coadaptation and outbreeding depression. Pp 105-116 In M. Soulé (ed.) *Conservation Biology: Science of scarcity and diversity*. Sinauer, Sunderland, MA.
- Tomiuk J and Loeschke V (1991) A new measure of genetic identity between populations of sexual and asexual species. *Evolution* 45: 1685-1694.
- Vekemans X (2002) AFLPsurv version 1.0: a program for genetic diversity analysis with AFLP (and RAPD) population data. Laboratoire de Génétique et Ecologie végétales, Université Libre de Bruxelles.
- Vilches B, Roca A, Naranjo J, Navarro B, Bramwell D, Caujapé-Castells J (2002) La variación genética de *Erysimum* L. (Cruciferae) en Gran Canaria. *Resúmenes del 1<sup>er</sup> Congreso de la biología de la conservación de plantas* Pp. 100. Universitat de València.
- Vilches B, Roca A, Naranjo J, Navarro B, Bramwell D, Caujapé-Castells J (2004) Estructura espacial de la variación genética de *Erysimum* L. (Cruciferae) en Gran Canaria: implicaciones para la conservación *ex situ* en bancos de germoplasma. *Botánica Macaronésica* 25: 15-30.
- Wagensberg, J (2006) *A más cómo, menos por qué*. Metatemas 92, Tusquets Editores, Barcelona.
- Waples R (1991) Genetic interactions between hatchery and wild salmonids: lessons from the Pacific Northwest. *Can. J. Fish. Aquat. Sci.* 48: 124-133.
- Waples R (1995) Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. *Am. Soc. Fish. Symp.* 17: 8-27.
- Wartenberg D (1989) SAAP version 4.3. Spatial Autocorrelation Analysis Program. Distributed by the author.
- Watterson GA (1977) Heterosis or neutrality? *Genetics* 85: 789-814.
- Watterson GA (1978) The homozygosity test of neutrality. *Genetics* 88: 405-417.
- Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* 2: 1123-1128.
- Weinberg W (1908) Über den Nachweis der Verebung beim Menschen. *Jahresh. Verein f. Vaterl. Naturk* in Württemberg 64:368-82.
- Weir BS (1996) Genetic data analysis II. Sinauer, Sunderland, Massachusetts, USA.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1358.
- Whitkus R, Crawford DJ (1987) Estimates of gene flow. The importance of historical distributions. *Am. J. Bot.* 74: 671-672.
- Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration:  $F_{ST} \neq 1/(4Nm + 1)$ . *Heredity* 82: 117-117.
- Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163: 1177-1191.
- Wolf PG, Soltis PS (1992) Estimates of gene flow among populations, geographic racea and species in the Ipomopsis aggregata complex. *Genetics* 130: 639-647.
- Wright S (1922) Coefficients of inbreeding and relationship. *American Naturalist* 56: 330-338.
- Wright S (1931) Evolution in mendelian populations. *Genetics* 16: 97-159.
- Wright S (1937) The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the USA* 23: 307-320.
- Wright S (1943) Isolation by distance. *Genetics* 28: 114-138.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenetics* 15: 323-323.
- Wright S (1969) Evolution and the genetics of populations, Vol. 2. University of Chicago Press, Chicago.
- Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997) PopGene, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.



Zouros E, Foltz DW (1987) The use of allelic isozyme variation for the study of heterosis. *Isozymes* 13: 1-59.



