

Integrando demografía y genética de poblaciones a través de la genética del paisaje

C. García ^{1,*}

(1) Institute of Integrative Biology, Department of Evolution, Ecology, and Behaviour, University of Liverpool, Crown Street, Liverpool L69 7ZB, United Kingdom.

* Autor de correspondencia: C. García [Cristina.Garcia-Perez@liverpool.ac.uk]

> Recibido el 25 de febrero de 2019 - Aceptado el 15 de abril de 2019

García, C. 2019. Integrando demografía y genética de poblaciones a través de la genética del paisaje. *Ecosistemas* 28(1):75-90. Doi.: 10.7818/ECOS.1694

La genética del paisaje ha sido una de las disciplinas más fructíferas en los últimos años gracias a la aplicación rutinaria de marcadores moleculares hipervariables, y más recientemente de secuenciación masiva, combinadas con la caracterización exhaustiva de los fenotipos y de la heterogeneidad ambiental. Los objetivos principales de estudio en las últimas décadas han sido: (1) inventariar la diversidad genética dentro de individuos, poblaciones o especies; (2) la identificación de patrones genéticos en poblaciones naturales (micro-evolutivos) y de los factores ecológicos actuales o históricos que los determina; (3) la cuantificación de los procesos macro-evolutivos actuales o históricos (e.g., migración, especiación, hibridación); (4) la inferencia de parámetros demográficos como tasas de migración, tamaño efectivo poblacional (N_e) y cuellos de botella; (5) la resolución de relaciones de parentesco que caracterizan los sistemas reproductivos y los factores que los modulan; (6) la cuantificación del efecto de la diversidad genética/genómica en la dinámica poblacional; y (7) la monitorización espacial y/o temporal de la diversidad genética o genómica. En este artículo se revisan algunas de las aproximaciones de estudio más interesantes que han ayudado a avanzar nuestro conocimiento sobre los factores y los procesos que subyacen las dinámicas eco-evolutivas de las especies en ambientes cambiantes.

Palabras clave: adaptación local; conectividad; cuello de botella; deriva genética; estructura genética; flujo génico; genética de la conservación; genómica del paisaje; inferencia demográfica; marcadores hipervariables; procesos demográficos; tamaño efectivo poblacional; variabilidad genética

García, C. 2019. Integrating demography and population genetics through landscape genetics. *Ecosistemas* 28(1):75-90. Doi.: 10.7818/ECOS.1694

Landscape genetics has been a highly productive discipline in the latest decades promoted by the advent of highly variable molecular markers and, more recently of next generation sequencing tools, combined with an exhaustive characterization of the phenotype and the environmental heterogeneity. The most relevant goals addressed in the latest decades are: (1) providing an exhaustive inventory of biodiversity at the genetic, population and species level; (2) identifying genetic patterns across the landscape and the ecological factors that shape them; (3) assessing the contemporary and/or historical evolutionary processes that determine current observed genetic patterns (e.g., migration, speciation or hybridization); (4) inferring demographic parameters such as N_e and demographic bottleneck; (5) resolve genetic relationships among individuals within and among populations and the ecological factors that determine them; (6) quantifying the effect of genetic diversity in shaping population dynamics; and (7) monitoring spatio-temporal changes in biodiversity across individuals, populations, species and communities. Here I provide an overview of the most relevant approaches applied in landscape genetics to advance our knowledge on the ecological and evolutionary processes that underlie the eco-evolutionary dynamics of natural populations in changing landscapes.

Key words: bottleneck; connectivity; conservation genetics; demographic inference; demographic processes; effective population size; gene flow; genetic drift; genetic structure; genetic variation; hyper-variable genetic markers; landscape genetics; landscape genomics; local adaptation; next generation sequencing

Introducción a la genética del paisaje: una aproximación histórica

La genética del paisaje aborda el estudio de los procesos y patrones microevolutivos en un contexto espacialmente explícito e integra conceptos, modelos y herramientas de la genética de poblaciones, de la ecología del paisaje y de la estadística espacial (Manel et al. 2003; Balkenhol et al. 2015). La variabilidad genética es la forma de variabilidad biológica fundamental y determina patrones y procesos evolutivos a nivel de individuo, población y especie (Futuyama 1998).

Durante el desarrollo de la Síntesis Evolutiva Moderna (primera mitad del s. XX) varios grupos de investigación debatieron los mecanismos que generan y mantienen la variabilidad genética de las poblaciones. Entre ellos, el grupo de Dobzhansky, Wright y Haldane publicaron estudios con datos empíricos que testaban el papel del flujo génico y el de la deriva genética, mientras que la contribución de la escuela británica con Ford, padre de la Ecología Genética (Ford 1964) y el estadístico R.A Fisher se centró en demostrar el papel de la selección natural y de la deriva genética en la evolución de las poblaciones naturales (Fisher 1930). El debate, aún vigente (Kern y Hahn 2018; Jensen et al. 2019),

sobre el papel del flujo génico, la deriva y la selección natural en generar y mantener la diversidad genética dentro y entre poblaciones fomentó el interés por la geografía de la genética de poblaciones con contribuciones fundamentales, como el modelo de aislamiento por distancia (Wright 1943; Malécot y Blaringhem 1948). Malécot desarrolló los modelos matemáticos que permiten estudiar la variación espacial de la variabilidad genética con poblaciones discretas y en espacios continuos y propuso los coeficientes de coancestría como medidas de parentesco entre individuos, base de la actual teoría de coalescencia (Nagylaki 1994). Años más tarde, Kimura y Crow propusieron que variabilidad genética de las poblaciones se debe a la acumulación gradual de mutaciones neutrales que se fijan en la población por deriva genética (Kimura y Crow 1964). Además, tanto Dobzhansky como Ford y sus respectivos colaboradores establecieron los protocolos de trabajo para testar los modelos teóricos de la biología evolutiva con experimentos manipulativos y observacionales que producían datos empíricos sobre los cambios evolutivos de las poblaciones naturales en respuesta a factores ambientales (Ford 1964; Lewontin 1974). Por ejemplo, Dobzhansky, Wright y Haldane publicaron entre 1938 y 1976 una serie de 43 trabajos compilados bajo el título de *The Genetics of Natural Populations* donde se exponen la teoría cuantitativa de la evolución basada en datos empíricos y describen distribución geográfica de la variabilidad genética de poblaciones naturales de *D. pseudoobscura* (Lewontin et al. 2003). Estos trabajos testan empíricamente los modelos evolutivos propuestos por S. Wright (Wright 1943), cuantifican los patrones de variación genética a lo largo de gradientes altitudinales y latitudinales (Lewontin et al. 2003 cap. XVI), y se obtienen estimas empíricas de las tasas de flujo génico de las poblaciones naturales a lo largo de dichos gradientes (Lewontin et al. 2003, cap. X). Además, se testan explícitamente el efecto del aislamiento por distancia entre poblaciones (Lewontin et al. 2003 cap. XXXI), y se cuantifica el efecto de factores ecológicos en los procesos de selección natural, como el uso de insectidas (Lewontin et al. 2003, cap. XXXIX) o el impacto de los eventos de sequía (Lewontin et al. 2003, cap. XX). Ford y Fisher desarrollaron y mejoraron diferentes métodos de captura y re-captura para estimar el tamaño poblacional y las tasas de mortalidad de las poblaciones naturales estudiadas (Fisher y Ford 1947). Estos métodos, siguen aplicándose para inferir el valor selectivo de algún carácter fenotípico de estudio (e.g. coloración) o la presión selectiva de algún factor ambiental (e.g. resistencia o susceptibilidad a pesticidas). La disponibilidad de mayor poder computacional y los avances matemáticos desarrollaron enormemente la estadística espacial en los 1970's que fue introducida a la genética de poblaciones por Robert R. Sokal (Sokal y Oden 1978) y posteriormente se ha aplicado para describir patrones de distribución de la variabilidad genética a gran escala y, de esta forma, detectar barreras al flujo génico (Barbujani y Sokal 1990); patrones de migración a escala continental (Cavalli-Sforza y Feldman 2003); identificar patrones de cruzamientos dentro y entre grupos (Sokal et al. 1986). La aplicación de forma rutinaria de los marcadores hipervariables basados en ADN a partir de los 1990's permite extender el estudio de la variabilidad genética más allá de especies modelo y abordar cuestiones más complejas desde un punto de vista espacial y temporal (Epperson 2003). Tras la publicación del trabajo seminal de Manel et al. (2003), la *genética del paisaje* ha cristalizado como disciplina integradora entre la ecología, biología evolutiva y geografía y el número de trabajos científicos relacionados con la genética (y genómica) del paisaje ha aumentado enormemente.

Tres factores principales han posibilitado el desarrollo de la genética del paisaje (Fig. 1). Por un lado, la incorporación de forma rutinaria de marcadores polimórficos hipervariables basados en ADN (mayoritariamente AFLPs, SSRs, SNPs) en los estudios de ecología (Schlotterer 2004). El acceso generalizado a herramientas moleculares ha permitido no sólo trabajar con es-

pecies modelo en biología evolutiva (como *Drosophila melanogaster*, *Arabidopsis thaliana* o *Neurospora crassa*) sino con especies no-modelo pero relevantes en ecología (e.g. *Mimulus* spp., *Salix* spp.), agronomía y ciencias forestales (*Eucalyptus* spp., *Pinus* spp., *Olea europea* var. *sylvestris*) o biología de la conservación (*Medusagyne oppositifolia*). Por otro lado, nuestra capacidad para monitorizar y registrar datos ambientales ha aumentado en las últimas décadas (Asner et al. 2016), lo cual nos permite evaluar el impacto de diferentes factores ecológicos en la generación de los patrones demográficos y genéticos observados en poblaciones naturales a diferentes escalas espaciales y temporales. Además, en las últimas décadas se han desarrollado diferentes herramientas computacionales que permiten manejar una gran cantidad de datos que mejoran nuestra capacidad para inferir procesos y caracterizar patrones de forma robusta (Beaumont y Rannala 2004; Excoffier y Heckel 2006; Allendorf et al. 2010). Este aumento de nuestra capacidad de computación nos permite integrar diferentes bases de datos para cuantificar la relación entre factores fenotípicos, ambientales y patrones genéticos, así como para obtener estimas de procesos demo-genéticos como la migración, el flujo génico y existencia de cuellos de botella. Además, no sólo podemos testar nuestras hipótesis basados en los datos muestreados en el campo o en el laboratorio, sino que podemos plantear diferentes escenarios para predecir el posible efecto del cambio ambiental a través de modelos de simulación (Hoban 2012), altamente útiles en trabajos de conservación y gestión. Esta combinación de técnicas moleculares, monitorización ambiental detallada y nuevas herramientas estadísticas ha permitido incorporar diferentes niveles de estudio (desde genomas a biomas, Whitman et al. 2006) y extender las escalas espaciales (desde una localidad a una escala global) y temporales (de una generación a múltiples generaciones). Por último, la necesidad acuciante de cuantificar el efecto de la actividad antrópica sobre los ecosistemas naturales y manejados, así como de predecir la respuesta de los individuos, poblaciones y especies a futuros cambios climáticos ha acelerado el interés por la genética y la genómica del paisaje, tanto por su valor predictivo como por su utilidad como herramienta de gestión en biología de la conservación (Allendorf 2017).

Esta breve presentación de la disciplina de la *Genética (y genómica) del paisaje* pretende: (i) relatar sucintamente el desarrollo de la disciplina en los últimos años y delinear las principales cuestiones que se abordan en la actualidad; (ii) describir de forma sintética las herramientas actualmente disponibles para abordar los temas fundamentales en ecología del paisaje; (iii) discutir posibles cuestiones relevantes para el avance de esta disciplina en los próximos años. Aunque se trata de una disciplina en expansión y cuyas aplicaciones son numerosas y diferentes, los temas fundamentales que se han abordado a través de la genética y genómica del paisaje se pueden agrupar en (Fig. 2): (1) inventariar la diversidad genética dentro de individuos, poblaciones o especies; (2) identificación de patrones genéticos en poblaciones naturales (micro-evolutivos) y de los factores ecológicos actuales o históricos que los determina; (3) cuantificación de los procesos macro-evolutivos actuales o históricos (e.g., migración, especiación, hibridación); (4) inferencia de parámetros demográficos como tasas de migración, tamaño efectivo poblacional (N_e) y cuellos de botella; (5) resolución de relaciones de parentesco que caracterizan los sistemas reproductivos y los factores que los modulan; y (6) cuantificación del efecto de la diversidad genética/genómica en la dinámica poblacional; y (7) Monitorización espacial y/o temporal de la diversidad genética o genómica. Finalmente, se incluyen protocolos básicos para diseñar un estudio de genética de poblaciones (Cuadro 1), se apuntan futuras líneas de investigación en este campo y se incluyen referencias a los principales paquetes y funciones de R más utilizados en este campo para facilitar su uso.

Figura 1. De la Síntesis Moderna a las dinámicas eco-evolutivas en el Antropoceno

**La Síntesis Moderna de la Evolución
(20's~70's)**

Contexto general: Desarrollo y establecimiento de la Síntesis Evolutiva Moderna que integra los principios evolutivos de Darwin y los mecanismos genéticos descritos por G. Mendel mediante la elaboración de herramientas cuantitativas, estadísticas y modelos experimentales desarrolladas por los genéticos de poblaciones y estadísticos T. Dobzhansky, S. Wright, R. Fisher, J.B.S. Haldane, entre otros. **Marcadores aplicados:** isoenzimas. **Temas principales de estudio:** papel del flujo génico, deriva genética, mutación y selección natural en determinar la diversidad y estructura genética de las poblaciones, así como los mecanismos de especiación¹. Combinación de modelos experimentales de laboratorio y poblaciones naturales. **Modelos de estudio:** especies modelos como *Drosophila melanogaster* y otras utilizadas en agronomía como el maíz (*Zea mays*).

**Síntesis Moderna Extendida
(~70's-2010's)**

Contexto general: integración de la biología evolutiva y la biología del desarrollo (evo-devo) y de la ecología. Demostración del efecto de la diversidad genética sobre la viabilidad poblacional en poblaciones de laboratorio y naturales. **Marcadores aplicados:** isoenzimas, AFLPs, RAPDs, RFLPs, SSRs. **Temas principales de estudio:** depresión por endogamia y declive poblacional (vórtex de extinción²); debate sobre el papel de la estocasticidad ambiental y genética en la extinción de las poblaciones³; estimas de tamaños poblacionales mínimos y el diseño de reservas que garantice el tamaño mínimo viable del habitat (SLOSS) para maximizar la biodiversidad, basados en la teoría de biogeografía de islas⁴; manejo de los paisajes y ecosistemas con el fin de maximizar la biodiversidad de especies y genética. Consolidación de la Genética del Paisaje⁵ con estudios sobre dispersión y flujo génico en paisajes manejados y procesos metapoblacionales en paisajes fragmentados. **Modelos de estudio:** especies emblemáticas en peligro de extinción (eg. lince ibérico, ecosistemas prístinos (eg. bosque amazónico) vs. paisajes fragmentados, especies invasoras.

**Dinámicas eco-evolutivas
(~2010's-)**

Contexto general: reconocimiento del impacto de los procesos evolutivos en los caracteres ecológicos de las especies generando dinámicas eco-evolutivas⁶. **Marcadores aplicados:** SNPs, secuenciación masiva, -omics, edición genética. **Temas principales de estudio:** Respuestas funcionales de los individuos y poblaciones (eg. epigenética, proteómica, metabolómica) a diferentes factores de cambio global; Factores genéticos como determinantes de la resiliencia de las comunidades y ecosistemas; Biología de la conservación tiene como prioridad maximizar la resiliencia de los ecosistemas y la restauración ecosistemas (rewilding) garantizando tanto la biodiversidad como la provisión servicios ecosistémicos (pj. actividad de los polinizadores). **Modelos de estudio:** la secuenciación masiva permite trabajar con especies no modelo y obtener inferencias robustas tanto de procesos históricos (desde el Pleistoceno y a lo largo del Holoceno, con especial énfasis en el Antropoceno) como predicciones robustas frente a diferentes escenarios de cambio global que incluyen la actividad humana⁷. Métodos de edición genética comienzan a considerarse para incrementar las posibilidades de adaptación de las especies a nuevos ambientes.

¹ Lewontin RC, Moore JA, Provine WB, Wallace B (2003) Dobzhansky's genetics of natural populations. Columbia University Press. New York. ² Soulé ME (1985) What is Conservation Biology. BioScience 35, 727-734. ³ Lande R (1988) Genetics and demography in Biological Conservation. Science 241, 1455-1460. ⁴ MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press. Princeton. ⁵ Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. TREE 18,189-197. ⁶ Hendry AP (2017) Eco-evolutionary dynamics. Princeton University Press. Princeton. ⁷ Urban MC (2016) Improving the forecast for biodiversity under climate change. Science 353, aad8466.

Figura 1. Cuadro comparativo de los diferentes temas de estudio y marcadores moleculares aplicados en genética de poblaciones en un contexto ecológico desde la Síntesis Moderna a la actualidad.

Figure 1. Comparative summary of the main research topics, approaches and molecular markers applied in population genetics addressed from an ecological context from the Modern Synthesis to nowadays.

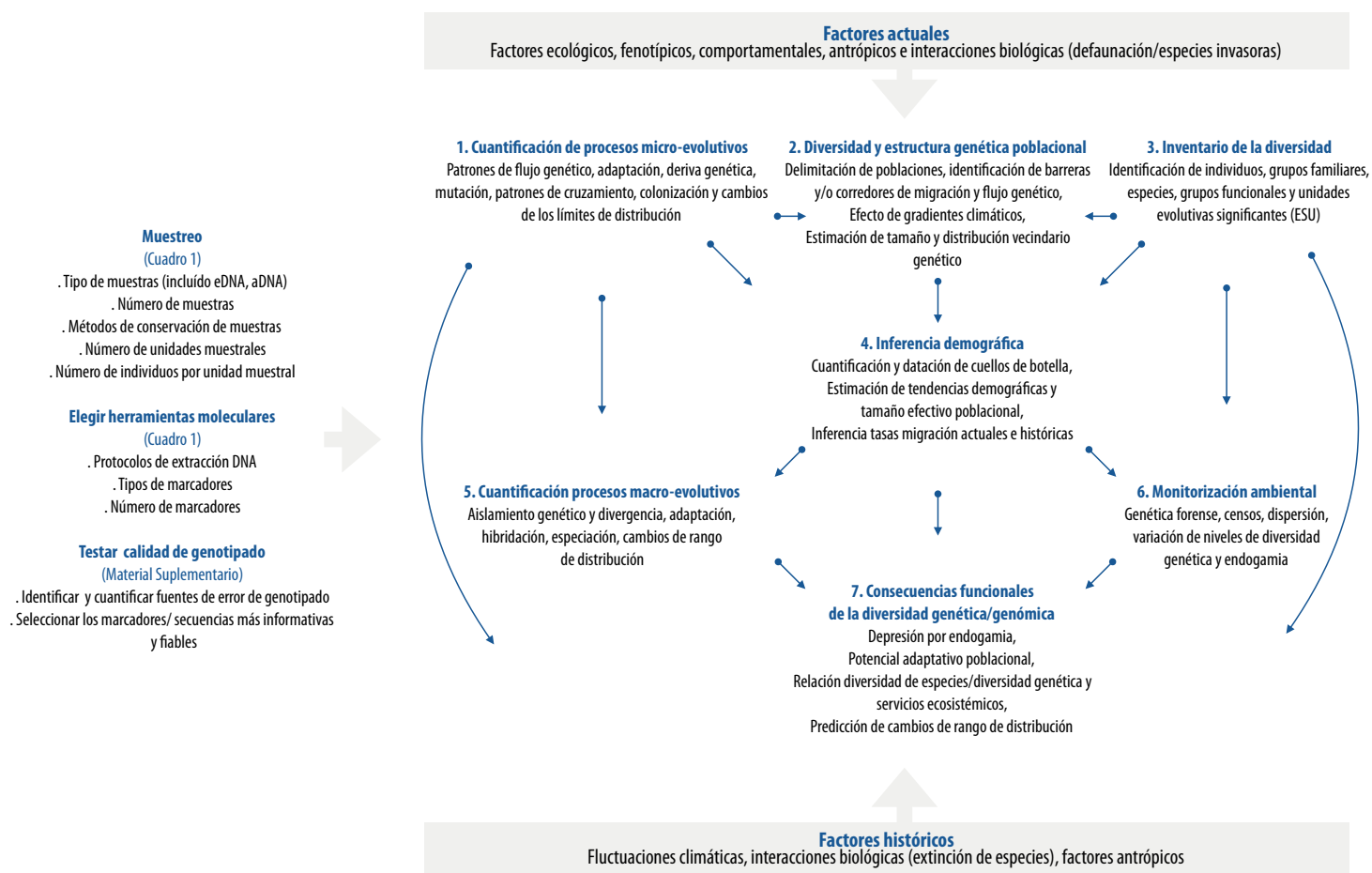


Figura 2. Principales temas de estudio abordados desde la genética del paisaje. Las flechas unen cuestiones que suelen estudiarse de forma conjunta, por ejemplo, la identificación de estructuras genéticas espaciales y la cuantificación de los procesos que los generan.

Figure 2. Main themes addressed by Landscape Genetics. Narrows link issues that typically are jointly addressed, such as the identification of genetic barriers and genetic clusters and the quantification of gene flow patterns across the landscape.

CUADRO 1. Diseño muestral

La decisión correcta sobre el número de poblaciones, muestras por población y número y tipo de marcadores es fundamental para cualquier estudio de genética y genómica del paisaje, aunque no hay una norma general aplicable a la mayoría de los estudios. Factores como el nivel de diversidad genética de la población, el grado de diferenciación genética entre poblaciones, el tipo de cuestiones que se quieren abordar y el presupuesto disponible condiciona el diseño del muestreo óptimo (Ryman et al. 2006). En general, cuestiones que requieren la identificación a nivel de individuo (análisis de paternidad, maternidad y/o parentesco) requieren una batería de marcadores altamente polimórficos e informativos donde el esfuerzo de muestreo se concentra dentro de las poblaciones de estudio (Hoban et al. 2013). Por otro lado, la cuantificación de la estructura genética entre poblaciones a lo largo de rangos de distribución amplios suele necesitar menos poder de resolución a nivel molecular y un extenso muestreo que incluya un gran número de poblaciones y menor número de individuos por población (Avice 2004). Lo más adecuado es basarnos en trabajos preliminares similares o bien diseñar un estudio piloto donde podamos estimar el número mínimo de muestras y marcadores necesarios para identificar/cuantificar el patrón/proceso de interés con un alto nivel de probabilidad para diferentes escenarios posibles. Además, reportar este resultado como parte del material suplementario de un trabajo puede ser de gran utilidad para otros investigadores que trabajen el mismo sistema de estudio.

Si el estudio carece de información genética previa, se puede comenzar estableciendo varios escenarios de muestreo en un estudio piloto para, en función de los resultados observados en este estudio, modificar el número de marcadores y muestras por población de acuerdo con los objetivos principales del trabajo. El software SPOTG (Hoban et al. 2013) simula diferentes escenarios definidos por el número de muestras y marcadores y evalúa el poder para estimar diferentes parámetros (conectividad, cuellos de botella, hibridación y asignación a poblaciones). De esta forma podemos decidir si nos interesa incrementar el número de muestras, de poblaciones o de marcadores en función de los resultados observados para cada escenario simulado. El contenido informativo de cada marcador también se puede evaluar con POWSIM (Ryman y Palm 2006) y HP-RARE (Kalinowski 2005) que aplican análisis de rarefacción para evaluar qué marcadores son más informativos. Cuando el objetivo principal está relacionado con análisis de paternidad, maternidad y/o parentesco es conveniente evaluar el poder de exclusión de cada marcador. Programas como GenALex 6.5 (Peakall y Smouse 2012) y Cervus 3.0.7 (Kalinowski et al. 2007) estiman el contenido informativo de cada marcador (PIC, *Polymorphic Information Content*). Además, estos programas también estiman la probabilidad de exclusión de un parental (PE) cuando el primer parental es conocido (PE1) o cuando ningún parental es conocido (PE2) para un conjunto de marcadores. De esta forma, cuando tenemos los genotipos del embrión de una semilla y de su madre, podemos decidir el conjunto mínimo de marcadores más informativo a la hora de aplicar análisis de paternidad, i.e., identificación del árbol donante de polen, (García et al. 2005) o incluso análisis que nos permitan identificar a la madre y el padre más probables (Hardesty et al. 2006).

Estas simulaciones nos ofrecen punto de partida fiable para comenzar a muestrear y los resultados preliminares además nos servirán para corregir algunos sesgos presentes en nuestro sistema de estudio que las simulaciones preliminares no tienen en cuenta pero que son importantes para obtener resultados correctos. En este sentido es importante resolver el nivel de endogamia dentro de la población y la presencia de estructura espacial dentro de la población para definir el diseño muestral definitivo. Además, conviene tener presente las posibles modificaciones históricas del paisaje (e.g. fragmentación o tala selectiva) ya que la composición y distribución de la variabilidad genética actual podría resultar de procesos históricos en lugar de responder a impactos recientes, particularmente cuando trabajamos con especies de larga vida como especies forestales (Bacles y Jump 2011). Dentro y entre poblaciones conviene tener en cuenta diferencias fenológicas que pueden explicar tamaños efectivos pequeños o patrones de migración asimétrica, como las que se establecen en presencia de gradientes climáticos (Gauzere et al. 2013). Por último, conviene dedicar tiempo a definir la unidad muestral de nuestro estudio porque el término población representa realidades diferentes para genéticos, ecólogos, demógrafos, biólogos evolutivos y gestores (Waples y Gaggiotti 2006).

Muchas especies de interés en ecología carecen de recursos genéticos desarrollados previamente (Neale y Kremer 2011) y en este caso SPOTG también puede utilizarse para decidir qué tipo de marcadores ofrecen una mayor resolución para cada objetivo. Estudios recientes que comparan los resultados obtenidos con SSRs y SNPs en general concluyen que un número moderado de SSRs (entre 10 y 30) consiguen resultados similares a centenares o miles de SNPs cuando el objetivo fundamental es la descripción de la variabilidad genética y estructura a lo largo del paisaje (Hodel et al. 2016). Los SNPs, por otro lado, suelen dar resultados más robustos cuando el objetivo principal es la inferencia demográfica, particularmente para tamaños muestrales pequeños (García et al. 2018), aunque en casos de cuellos de botella muy drásticos, rápidos y relativamente recientes se pueden obtener resultados fiables con una decena de SSRs (Moodley et al. 2017). Aproximaciones más recientes aplican algoritmos basados en *machine learning* para seleccionar el conjunto mínimo de marcadores que consiguen la resolución deseada en análisis de asignación de individuos a poblaciones mejorando de forma significativa el poder de asignación con un número reducido de marcadores (Sylvester et al. 2018).

Cuadro 1. Diseño muestral: cuántas poblaciones, cuántos individuos, cuántos marcadores.

Box 1. Sampling design: how many populations, individuals, and markers.

Marcadores moleculares

Los primeros marcadores moleculares que se utilizaron en biología evolutiva fueron las aloenzimas (i.e. formas alternativas de una enzima codificadas por un mismo locus genético) que permitieron caracterizar de forma rutinaria y a bajo coste la variabilidad genética dentro y entre poblaciones para un espectro amplio de especies de animales y de plantas (Charlesworth et al. 2016). Además, permitieron resolver relaciones de parentesco con cierta fiabilidad y resolver algunas relaciones fenotipo-genotipo, aunque no eran suficientemente resolutivos para mapear caracteres fenotípicos. Por tanto, los marcadores más utilizados hasta los 1990's fueron las aloenzimas que representan una forma indirecta de cuantificar la variabilidad genética ya que son proteínas (fenotipo) y, por tanto, no permiten medir la variabilidad en partes del genoma que no se codifica o la variabilidad debida a mutaciones silenciosas. Por ello, la llegada de los marcadores moleculares basados en ADN en los 1980's supuso un gran avance cualitativo en ecología. Los primeros marcadores basados en ADN que se utilizaron fueron los RFLPs (Restriction Fragment Length Polymorphisms) y posteriormente los minisatélites que proporcionaron un nivel de polimorfismo alto y, por primera vez, suficiente para identificar individuos que

quedaban caracterizados a nivel individual por una huella genética (fingerprinting). Sin embargo, los marcadores moleculares que revolucionaron la ecología molecular y la biología evolutiva fueron los microsátélites (SSRs, Short Sequence Repetitions ó STR, Short Tandem Repeat). Los SSRs son, al igual que los minisatélites, secuencias (motivos) repetidas en tándem de longitud variable y fácilmente amplificables mediante PCRs (Polymerase Chain Reaction). Los SSRs son altamente polimórficos y están repartidos por todo el genoma y, en principio, representan la variabilidad genética neutral de los individuos y poblaciones caracterizadas. Este conjunto de características, junto a su coste asequible, los ha convertido en el marcador más utilizado en ecología y biología evolutiva para abordar un gran número de cuestiones. Además, estos marcadores hipervariables basados en PCR posibilitaron trabajar con tejidos problemáticos hasta entonces por su bajo contenido en ADN y/o su alta degradación como muestras de museos (Mondol et al. 2013), excrementos (Fedriani y Kohn 2001), restos de tejidos dérmicos (Morin y Mccarthy 2007), tejidos leñosos (Godoy y Jordano 2001), o granos de polen (Isagi y Suyama 2011). Además, permitió caracterizar no sólo la variabilidad genética del ADN nuclear sino también del ADN mitocondrial y ADN cloroplástico, lo cual supuso una herramienta fundamental para los estudios de filogeografía.

grafía y biogeografía. Su gran polimorfismo permite resolver relaciones de parentesco y análisis de paternidad con una alta fiabilidad, caracterizar a los individuos de forma única, obtener estimas robustas de tamaños poblacionales, hacer inferencia demográfica histórica fiable y mapear caracteres fenotípicos. Una de las desventajas principales de los SSRs es su complejo patrón de mutaciones que puede generar patrones no siempre fáciles de interpretar a la hora de asignar el tamaño de los alelos (scoring) de forma consistente y robusta (Dewoody et al. 2006) (ver Anexo sobre la identificación y estimas de errores de genotipados). Además, los SSRs requieren desarrollar las secuencias de los cebadores o *primers* para cada especie de estudio, aunque habitualmente los marcadores desarrollados para una especie pueden transferirse a especies cercanas (Avisé 2004). La alternativa al desarrollo de cebadores específicos fue la aplicación de marcadores dominantes como los RAPDs (Randomly Amplified Polymorphic DNAs), los ISSR (Inter-Simple Sequence Repeat) y, principalmente de AFLPs (Amplified Fragment Length Polymorphism) que se basan en la amplificación de fragmentos que resultan de la digestión del ADN por enzimas de restricción. La variabilidad genética queda caracterizada como presencia/ausencia de los fragmentos obtenidos, pero este bajo nivel de polimorfismo por locus queda compensado por un gran número de loci caracterizados. Sin embargo, el coste cada vez más asequible de los SSRs los ha convertido en los marcadores más utilizados en ecología y biología de poblaciones desde finales de los 2000's hasta la llegada de la secuenciación masiva a los laboratorios de ecología molecular. El desarrollo de marcadores hipervariables basados en técnicas de secuenciación masiva ha permitido incrementar en varios órdenes de magnitud el número de loci polimórficos disponible incluso para especies de estudio no modelos. Sin embargo, los SSRs continúan siendo una buena opción para resolver relaciones de parentesco o caracterizar la distribución de la variabilidad genética a través el paisaje, incluso obtener estimas de cuello de botella cuando la señal es muy intensa, por ejemplo, cuando el declive poblacional ha sido rápido y muy intenso los cuellos de botella se pueden datar y cuantificar con unos pocos marcadores SSRs (Moodley et al. 2017). Por tanto, la elección de SSRs vs. SNPs en estudios de ecología molecular y de la conservación varía en función de los objetivos, el tamaño de muestras manejadas en el estudio, la historia demográfica de la población, la disponibilidad de recursos genéticos de la especie y de las herramientas informáticas al alcance del grupo de investigación (Hodel et al. 2016). La llegada de las técnicas de secuenciación masiva permite, además, caracterizar la variabilidad funcional del genoma, lo cual ha supuesto el inicio de la genómica del paisaje (Luikart 2003) que posibilita la detección de patrones de adaptación a ambientes heterogéneos de las poblaciones a lo largo de gradientes ambientales a diferentes escalas espaciales (Eckert et al. 2010).

Aplicaciones de la genética y genómica de poblaciones

Descripción de estructuras espaciales y los factores ambientales que los determinan

Una de las aplicaciones más frecuentes de las herramientas de la genética del paisaje es la identificación de las estructuras genéticas que forman las poblaciones en el paisaje. Estas estructuras se clasifican, grosso modo, como agrupamientos o estructuras genéticas de agregación (*clusters*), clinas, y estructuras jerárquicas donde se pueden identificar grupos estructurados dentro de estructuras mayores (Balkenhol et al. 2015). Los agrupamientos se definen formalmente como un conjunto de individuos en equilibrio Hardy-Weinberg por lo que representan vecindarios genéticos y habitualmente consisten en individuos cercanos, aunque estos grupos pueden ocupar extensiones considerables (Jordano 2017). La presencia de barreras al flujo génico tanto naturales (estrechos marinos o sistemas montañosos) como generados por la acción del hombre (autopistas, embalses o terrenos deforestados) impide el movimiento de individuos y genera agrupamientos genéticos cla-

ramente identificables como consecuencia de la disminución del flujo genético (Hevroy et al. 2017). Además, la disminución de la capacidad dispersiva de las especies, por ejemplo, por pérdida de frugívoros dispersores de semillas, también incrementa la estructuración interna de las poblaciones y crea vecindarios genéticos reducidos (Pérez-Méndez et al. 2016).

La heterogeneidad ambiental también puede moldear la estructura genética en forma de grupos cuando las poblaciones ocupan hábitats contrastados a lo largo de un gradiente ambiental. Este es el caso de las especies alpinas que ocupan un rango altitudinal amplio donde en flujo génico entre poblaciones localizadas en latitudes altas y bajas es muy pequeño o nulo como consecuencia de las diferencias fenológicas (Morente-López et al. 2018). En este caso, la estructura genética a lo largo del paisaje suele aparecer en forma de clina en lugar de grupos perfectamente definidos. Además de procesos demográficos neutrales, estas clinas también pueden indicar procesos de selección local, con lo que es aconsejable identificar variabilidad en sitios a lo largo del genoma que medien respuestas selectivas (temperatura, floración etc.) para distinguir entre el efecto del aislamiento por distancia y la adaptación local en estos casos. Finalmente, los procesos de emparejamiento selectivo producen típicamente estructuras jerárquicas (grupos dentro de grupos) donde es posible identificar un primer nivel de estructuración a gran escala debido o al aislamiento por distancia y un segundo nivel de estructuración dentro de poblaciones, normalmente coincidiendo con vecindarios genéticos (Balkenhol et al. 2015). Las aplicaciones más habituales para identificar estructuras genéticas son STRUCTURE (Pritchard et al. 2000) y GENELAND (Guillot et al. 2005, 2008) que utilizan una aproximación bayesiana para primero estimar el número más probable de agrupamientos (*clusters* o *K*) y luego para estimar la probabilidad de la ancestría de los individuos que permite asignarlos a dichos grupos con una probabilidad (Lawson y Falush 2012). La aplicación de este tipo de modelos ha sido tan generalizada que pueden llegar a aplicarse sin tener en cuenta si el modelo subyacente es relevante para el caso de estudio (Lawson et al. 2018). Esto es importante porque en caso de estructuras tipo clina que responden a un gradiente ambiental (en lugar de agrupamientos definidos por una pérdida gradual de flujo genético con la distancia) estos modelos pueden llegar a identificar poblaciones de origen espurias a lo largo de gradientes. Cuando el caso de estudio no se ajusta a los modelos en los que se basan en STRUCTURE o GENELAND es posible aplicar técnicas de agrupamiento no basadas en modelos, como los análisis discriminantes de los componentes principales a partir de genotipos multilocus (DAPC) (Jombart et al. 2010). Recientemente, se han aplicado nuevas aproximaciones que basadas en algoritmos de *machine learning* (implementadas en *AssignPop*, Chen et al. 2018) que integran datos genéticos y no-genéticos (e.g. fenotípicos o espaciales) para aumentar la robustez de la asignación a poblaciones, particularmente en situaciones con un alto nivel de flujo génico entre poblaciones (Chen et al. 2018).

Además de las estructuras espaciales a escala del paisaje, las estructuras genéticas dentro de población también pueden de interés porque están determinadas por patrones de cruzamiento y heterogeneidad ambiental dentro de población. En este sentido, la presencia de estructuras de genéticas agregación (agregación de genotipos similares) o repulsión (agregación de genotipos distantes) se puede explorar con *adegenet* (Jombart 2008) mientras que la estructura genética a escala espacial fina y su magnitud y extensión se puede describir mediante la obtención de correlogramas espaciales con *SPAGeDi* (Hardy y Vekemans 2002) o GenAlex 6.5 (Peakall y Smouse 2012). Una vez que se han identificado las estructuras genéticas es interesante identificar los factores ecológicos o fenotípicos que los determinan (Tabla 1).

Cuantificación directa de la dispersión de propágulos en poblaciones de plantas

El nivel de flujo génico entre poblaciones es uno de los procesos que más influye en la estructuración espacial de la diversidad genética, pero es difícil obtener estimas directas y robustas que

Tabla 1. Aproximaciones de análisis del impacto de los factores ecológicos y fenotípicos que determinan la diversidad genética de las poblaciones y su distribución espacial en paisajes heterogéneos.

Table 1. Approaches to study the impact of ecological, phenotypic, and geographic factors in determining genetic diversity and their spatial distribution across heterogeneous landscapes.

Método	Basado en	Variables	Paquete R	Referencia
Test de Mantel Parcial	Testar la correlación entre una matriz de distancias genéticas y otra ambiental controlando por una matriz de distancias geográficas	Matrices de distancias genéticas, ecológicas y geográficas	vegan ¹	Legendre et al. (2015) , Guillot y Rousset (2013)
Modelización causal recíproca (RCM)	Múltiples Test de Mantel Parciales entre pares de matrices ambientales, geográficas y genéticas	Matrices de distancias genéticas, ecológicas y geográficas	ade4 ²	Cushman et al. (2013)
Análisis de redundancia parciales (pRDA)	RDA parcial determina la relación entre un conjunto de variables ecológicas y/o fenotípicas y variables genéticas medidas a nivel de individuo (p.j. nivel de homocigosidad), población (p.j. riqueza alélica) o vecindario (p.j. tamaño del vecindario genético, Nb) mientras se controla por otra variable, por ejemplo la localización espacial.	Matrices de distancias genéticas, ecológicas y geográficas	vegan ¹	Pournosrat et al. (2018)
Análisis de Correlación Canónico (CCA)	Los análisis de correlación canónica generan variables compuestas (canónicas) a partir de dos vectores de variables (p.j. medidas de variabilidad genética y medidas de variabilidad ambiental) de forma que se maximice la correlación entre los conjuntos de variables.	Vectores de medidas genéticas y ecológicas	vegan ¹	García et al. (2005)
Matrices Múltiples de Regresión con aleatorización (MMRR)	Regresión de las matrices de distancia genética, ambiental y geográfica teniendo en cuenta la no-independencia entre las matrices	Matrices de distancias genéticas, ecológicas y geográficas	PopGenReport ³	Wang et al. (2013)
Modelos mixtos con factores latentes (LFMM)	Cuantifica la asociación entre la frecuencia de diferentes loci en el genoma y un conjunto de variables ecológicas	Frecuencias alélicas y variables ecológicas	LEA ⁴	Pournosrat et al. (2018)
Modelos generalizados mixtos con autocorrelación espacial	Estimas basadas en máxima probabilidad para los factores fijos y aleatorios de modelos mixtos, incluyendo factores de correlación espacial (Rosset y Ferdy 2014)	Estimas de diversidad genética y/o parámetros demográficos	spaMM ⁵	Morente-López et al (2018) , Muñoz-Pajares et al. (2017)
Redes espaciales dendríticas (SSN)	Modelización conjunta de la covariación espacial y ambiental en estructuras espaciales dendríticas como ríos o sistemas montañosos	Estimas de diversidad genética y/o parámetros demográficos	SSN ⁶	Rodríguez-González et al. (2019)
Modelos de ecuaciones estructurales (SEM)	Conjunto de ecuaciones de regresión que miden el efecto directo, indirecto o dinámicas de retroalimentación que se establecen entre un conjunto de variables ecológicas y fenotípicas y una variable respuesta genética (p.j. nivel de endogamia o diversidad genética)	Estimas de diversidad genética y/o parámetros demográficos	piecewiseSEM ⁷	García et al. (2016)
Aprendizaje automático (ML, Machine Learning)	Identificación de factores ecológicos que determinan la distancia genética entre poblaciones	Estimas de diversidad genética y/o parámetros demográficos	randomForest ⁸	Schrider y Kern (2018) , Hether y Hoffman (2012)

¹ Oksanen, et al 2019; ² Dray y Dufour 2007; ³ Adamack y Gruber 2014; ⁴ Frichot y François 2015; ⁵ Rousset y Ferdy 2014.; ⁶ Ver Hoef et al 2014; ⁷ Lefcheck 2016; ⁸ Liaw y Wiener 2002.

nos permitan evaluar su importancia. Los marcadores moleculares hipervariables, como los microsatélites o SSRs, han permitido obtener estimas directas de la dispersión de semillas dispersadas. En un trabajo seminal, [Godoy y Jordano \(2001\)](#) demostraron que el genotipado de tejidos maternos de las semillas, concretamente del endocarpio, permitía identificar el árbol madre que había producido la semilla por comparación de genotipos de los endocarpios de las semillas dispersadas con todos los genotipos de los árboles que habían producido frutos en la población. Si además, el punto de recolección de la semilla dispersada y el árbol madre están mapeados, se puede obtener la distancia de dispersión de cada una de las semillas de las que se ha genotipado el endocarpio ([Jones y Hubbell 2006](#); [García et al. 2007](#); [García y Grivet 2011](#)). El conjunto de semillas que queda sin asignar a la población o parcela de es-

tudio se asume han sido dispersadas desde árboles de fuera y por tanto representan eventos de dispersión a larga distancia que conectan poblaciones a través del paisaje ([Jordano et al. 2007](#)) y una estima directa del flujo génico actual entre poblaciones de estudio. Esta técnica se puede extender para identificar los parentales (padre y madre) de las plántulas y juveniles establecidas en la población. Los análisis parentesco (en los que se identifican a ambos progenitores) basados en máxima probabilidad a partir de modelos de herencia mendeliana permiten obtener una probabilidad de asignación de cada plántula a cada uno de los padre y madres potenciales. En este caso, las estimas de flujo génico representarían estimas del flujo génico efectivo ya que la semilla dispersada ha llegado a establecerse con éxito en el sitio de deposición ([Hardesty et al. 2006](#)).

De forma similar, a partir del genotipado del embrión de las semillas muestreadas de frutos recolectados de las copas de árboles con frutos se pueden obtener estimas de flujo génico contemporáneo del polen así como sus distancias de dispersión dentro de la población focal y los factores ecológicos y fenotípicos que influyen en el éxito reproductor de cada árbol donante de polen (García et al. 2005; Geng et al. 2008). Al muestrear directamente de la copa del árbol madre, la identidad de uno de los parentales es conocida, y por tanto, aumenta mucho la probabilidad de identificar el genotipo del árbol donante de polen. Otra aproximación fascinante es el genotipado de granos de polen individuales a partir de los cuales se puede identificar el donante de polen, por ejemplo, de granos de polen recolectados directamente del cuerpo de un polinizador o muestreando los granos de polen que han quedado adheridos al estigma de la flor (Isagi y Suyama 2011).

Cuando no es posible medir las distancias de dispersión de forma directa, por ejemplo, en poblaciones forestales continuas y muy extensas donde es inviable genotipar cada árbol reproductor, se pueden aplicar modelos que estiman de forma conjunta los parámetros de las funciones de dispersión de polen así como las estimas de fecundidad y el efecto de factores fenotípicos (e.g. número de flores o tamaño de la copa) (Gauzere et al. 2013). El software MEMM (Klein et al. 2011) combina genotipos multilocus de embriones, madres y padres potenciales y datos fenotípicos de éstos para caracterizar de forma conjunta las funciones de dispersión, así como los patrones de cruzamiento (e.g. proporción de eventos de autopolización vs. polinización cruzada de cada árbol madre) de la población de estudio, lo cual permite evaluar y comparar el efecto de diferentes factores a lo largo de un gradiente ambiental (Gauzere et al. 2013). Las aproximaciones para evaluar el efecto de los factores ecológicos y fenotípicos en la estructuración de la variabilidad genética dentro y entre poblaciones son numerosos (Tabla 1).

Cuantificación del flujo génico histórico y cuantificación del tamaño efectivo de población

La estructuración de la diversidad genética es el resultado de eventos de dispersión, procesos de adaptación local, expansiones o contracciones demográficas (i.e., cuellos de botella), mutaciones y deriva genética que han actuado a lo largo de muchas generaciones. Por tanto, las estimas de dispersión y flujo génico basadas en métodos directos son insuficientes para patrones de migración históricos. De hecho, los patrones de diversidad y estructuración genética en poblaciones de especies de larga vida, como muchas especies forestales, están más relacionados con eventos demográficos históricos que con patrones de dispersión recientes (Lowe et al. 2015). En este contexto, diferentes herramientas de genética de poblaciones se pueden aplicar para inferir patrones de migración histórica entre poblaciones a varias escalas de paisaje a partir de los genotipos multilocus de un número suficiente de individuos de las poblaciones de estudio. Por ejemplo, métodos basados en coalescencia permiten estimar tasas de migración históricas (a lo largo de muchas generaciones) a partir de la información genealógica de los genes muestrados (e.g. MIGRATE, Beerli y Felsenstein 2001). Además, aproximaciones bayesianas como BayesAss (Wilson y Rannala 2003) o GENELAND (Guillot et al. 2008) identifican individuos migrantes recientes (una o dos generaciones) a partir las señales de desequilibrio gamético generada por los individuos migrantes entre poblaciones con un nivel alto o moderado de estructura genética. Una aproximación similar a BayesAss es aplicada por Faubet y Gaggiotti (2008) para cuantificar, además de tasas de migración entre poblaciones en la última generación, el efecto de los factores ecológicos que pueden estar afectando los patrones de migración, como las densidades de población locales, el tamaño de los fragmentos o la distancia entre ellos (Faubet y Gaggiotti 2008). Por ejemplo, Sharma et al. (2013) aplicaron BayesAss para cuantificar la efectividad de los corredores que unen cuatro poblaciones de tigres (*Panthera tigris*) en India central en un paisaje altamente fragmentado. Las estimas de flujo génico reciente comprobaron que la migración entre fragmentos es abundante, par-

ticularmente en corredores forestales. Además, pudieron inferir un alto nivel de tasas de flujo génico históricas y un declive poblacional que comenzó hace 1000 años aplicando MIGRATE con tan sólo siete marcadores moleculares (SSRs). El estudio concluyó que el declive demográfico aceleró su ritmo hace unos 200 años coincidiendo con la urbanización del paisaje a gran escala (Sharma et al. 2012, 2013).

Las estimas del tamaño efectivo de población (Wright 1931) (N_e) así como la magnitud de los cambios que han sufrido a lo largo de un periodo de estudio es otro de los aspectos de estudio clave en la genética del paisaje en los que demografía, genética de poblaciones y biología evolutiva y de la conservación quedan indisolublemente unidos (Lowe et al. 2017). El tamaño efectivo de una población se define como el tamaño que tendría una población hipotética (infinitamente grande y en equilibrio Hardy-Weinberg) donde los cambios de frecuencias alélicas sólo dependen de la deriva genética. Por tanto, el N_e de una población de estudio es el tamaño de una población hipotética con el mismo nivel de endogamia y deriva génica que nuestra población de estudio (Wang 2017). NeEstimator (Do et al. 2014) incluye diferentes métodos de estimación del N_e (exceso de heterocigotos, momentos, desequilibrio de ligamiento y coancestría). El método adecuado variará en función del grado de estructura genética de la población, de su tamaño de censo, y de la variabilidad genética del conjunto de marcadores utilizado. Otras aproximaciones basadas en coalescencia, como *msms* (Ewing y Hermisson 2010) permiten evaluar los cambios temporales del N_e en escenarios complejos donde las poblaciones están fuertemente estructuradas internamente (en grupos familiares, por ejemplo) o están sujetas a selección. La datación de cuellos de botella y cuantificación de los cambios temporales de N_e a partir de datos moleculares es una herramienta muy útil en biología de la conservación ya que permite evaluar el efecto demográfico y genético de acciones antrópicas actuales y pasadas. Por ejemplo, Quéméré et al. (2012) utilizaron datos moleculares para testar si la presencia y extensión de grandes paisajes abiertos en Madagascar se debe un cambio climático histórico en la región que favorecería la expansión de herbáceas en ambientes crecientemente áridos o si es la consecuencia de una deforestación intensa que comienza con la llegada de los europeos a la isla (Quéméré et al. 2012). Las inferencias de los cambios de N_e de uno de los lémures forestales de la isla mostraron que su declive demográfico comenzó antes de la llegada de los primeros humanos a la isla y que, por tanto, la causa de la reducción del N_e está más relacionada con un cambio climático en la región que con la actividad antrópica. La información genética de poblaciones actuales también puede inferir las tasas de crecimiento o declive poblacional por generación cuando la población ha sufrido cambios drásticos de tamaño mediante métodos de coalescencia (LAMARC) (Kuhner 2006) y otras aproximaciones de simulación resultan útiles para simular escenarios demográficos y evolutivos más complejos (Cuadro 2).

Resolución de las relaciones de parentesco y cuantificación de los niveles de endogamia en las poblaciones y su efecto en la dinámica poblacional

La biología de la conservación ha sido uno de los campos que más se ha beneficiado de la aplicación los marcadores moleculares, particularmente los estudios de las causas y consecuencias de la consanguineidad en la viabilidad de las poblaciones naturales (Frankham et al. 2017). En poblaciones pequeñas y en peligro de extinción, la consanguineidad debido al apareamiento entre individuos genéticamente relacionados es el factor genético determinante de su viabilidad, junto con factores ecológicos que pueden limitar su supervivencia. Los marcadores moleculares hipervariables han permitido trazar la genealogía de los individuos de la población, estimar su nivel de parentesco y cuantificar la endogamia de la población y de los individuos a partir de datos de progenies maternas o a partir de genotipos con ascendencia desconocida (Wang 2011, 2014) incluso en poblaciones pequeñas con unos pocos individuos muestreados (Wang 2017). Por ejemplo, Breed et al. (2012) utilizaron marcadores moleculares para caracterizar los

CUADRO 2. Herramientas de simulación

Una de las aplicaciones más interesantes, aunque menos utilizadas hasta la fecha en genética y genómica de poblaciones, son los programas que permiten obtener resultados genéticos y/o demográficos *in silico* que resultan de la acción combinada de los procesos y factores que definen cada uno de los escenarios de estudio (Hoban et al. 2012). En general, estos programas informáticos se utilizan para predecir el efecto de algún proceso evolutivo (ej. flujo génico vs. selección) a lo largo de un número de generaciones futuras bajo diferentes escenarios demográficos y ambientales. Esta aproximación se conoce como simulaciones de tiempo hacia delante (*forward in time simulations*) y habitualmente consisten en modelos basados en individuo, donde cada individuo se introduce como un genotipo multilocus con información adicional como edad, sexo, el valor de algún carácter adaptativo objeto de estudio (pigmentación o tamaño del ala como proxy de la capacidad de dispersión). Algunos programas que permiten simular dinámicas evolutivas complejas para cuantificar el efecto de diferentes procesos evolutivos p. flujo génico vs. selección local, como *Nemo* (Guillaume y Rougemont 2006) y *Rangeshifter* (Bocedi et al. 2014). Otros programas además integran las dinámicas evolutivas en un contexto espacialmente explícito, lo que permite evaluar el impacto de los cambios en el paisaje en la diversidad genética, su distribución espacial y el tamaño poblacional. Por ejemplo, Aben et al. (2014) utilizaron *Rangeshifter* para diseñar una red de hábitats forestales que garantizara la viabilidad las poblaciones del bulbul de Cabanis (*Phyllastrephus cabanis*), un passeriforme africano. El modelo basado en individuo y espacialmente explícito cuantificó las tendencias demográficas de las poblaciones en las siguientes décadas a partir de datos de abundancia, demográficos (éxito de reproducción y supervivencia) y de calidad de hábitat. De esta forma, pudieron comparar la trayectoria demográfica de la meta-población con diferentes escenarios donde variaban el número y tamaño de los fragmentos, así como la conectividad entre ellos (Aben et al. 2014). *Rangeshifter* también se ha utilizado para estimar el efecto del cambio climático sobre las tasas de dispersión y sobre algunos caracteres morfológicos (Henry et al. 2014) y para explicar el efecto del tipo de paisaje (forestal vs. deforestado) sobre las dinámicas evolutivas en poblaciones en expansión (Heikkinen et al. 2014). El programa *CDPOP* (Landguth y Cushman 2010) también aplica modelos basados en individuos y espacialmente explícitos y simula los cambios en la composición genética de las poblaciones a partir de datos basados en individuos (genotipos multilocus, probabilidad de dispersión, éxito reproductivo, mortalidad) y resistencia al flujo génico de la matriz del paisaje a través de la cual se dispersan los individuos. *Metasim* (Strand 2002) y su versión disponible como paquete de R, *rmetasim* (Strand y Niehaus 2018) también permiten explorar las consecuencias genéticas y demográficas de escenarios evolutivos complejos de forma espacialmente explícita. Otros programas permiten además evaluar explícitamente las consecuencias genéticas de cambios en el paisaje como *Rangeshifter* (Bocedi et al. 2014) y *IBDsim* (Leblois et al. 2008), que cuantifican los efectos genéticos de cambios en el paisaje (presencia o ausencia de barreras, fragmentación) y de cambios en la densidad de población. Otros programas como *Bottlesim* simulan cuellos de botella en especies de vida larga y con ciclos de vida complejos para cuantificar sus efectos sobre la diversidad genética, pero teniendo en cuenta también los efectos de los cambios en el paisaje. Si el investigador carece de escenarios iniciales, puede generarlos con la mayoría de los programas mencionados, y el paquete de R *skelSim* (Parobek et al. 2017) ofrece una opción gráfica muy intuitiva y sencilla que ayuda diseñar los diferentes escenarios demográficos y genéticos que se quieren simular, al igual que PopPlanner (Ewing et al. 2015) que facilita la definición de escenarios demográficos complejos simulados mediante algoritmos de coancestría (Ewing y Hermisson 2010).

Las simulaciones son, además, una herramienta fundamental para inferir y cuantificar eventos demográficos pasados que expliquen la distribución de la diversidad genética y la estructura demográfica actual (Hoban et al. 2012). En este caso se simulan datos bajo varios escenarios posibles a partir de los cuales se estiman diferentes parámetros demográficos (como el tamaño efectivo poblacional, N_e) y genéticos (como F_{st}) que se comparan con los parámetros observados en la población de estudio para inferir el modelo con mejor soporte estadístico. La mayoría de los modelos usan algoritmos basados en coalescencia y cálculos bayesianos aproximados (Approximate Bayesian Computation, ABC) para inferir la probabilidad de cada escenario testado y de los parámetros demográficos derivados de ellos. A diferencia de las simulaciones de tiempo hacia delante (*forward in time simulations*), los algoritmos de coalescencia reconstruyen la genealogía de los genes o individuos de las poblaciones de estudio desde el presente hacia el pasado (*backward in time simulations*) hasta el momento en el que los linajes coalescen en un ancestro común (Hoban et al. 2012; Wakely 2016). Cuando el número de marcadores genéticos es muy grande, como es habitual en estudios de humanos, las estimas pueden llegar a tener un nivel de precisión muy alto y pueden detectar cuellos de botella ocurridos miles de generaciones atrás (Nielsen et al. 2017). Sin embargo, con unos centenares de SNPs o decenas de SSRs se han obtenido buenas estimas para especies no-modelo, como por ejemplo de la datación del momento en el que dos poblaciones comienzan a divergir (Sharma et al. 2012, 2013), de cuellos de botella (Moodley et al. 2017) o de expansiones demográficas (Wei et al. 2013). El software DIY ABC (Cornuet et al. 2008) ha sido el más utilizado pero paquetes de R como *abc* (Csilléry et al. 2012), *ABCtoolbox* (Wegmann et al. 2010) y *abctools* (Nunes y Prangle 2015) son buenas alternativas.

Una de las principales limitaciones de las herramientas de simulación en genética y genómica del paisaje es que hasta la fecha no permiten simular las dinámicas eco-evolutivas de más de una especie de forma conjunta. Por tanto, las estimas demográficas son limitadas a la hora de inferir el impacto de la pérdida o ganancia de interacciones biológicas o las consecuencias de la competencia o la facilitación entre especies de una comunidad.

Cuadro 2. Herramientas de simulación.

Box 2. Tools for simulating ecological and evolutionary scenarios.

patrones de cruzamiento de los individuos reproductores de *Swietenia macrophylla* en el bosque tropical a lo largo de América Central. Los individuos estudiados eran aislados y rodeados de una matriz agrícola o bien en parches forestales de diferente tamaño. Los investigadores consiguieron demostrar un efecto de depresión por endogamia en la progenie de individuos aislados y particularmente en aquellos procedentes de ambiente secos, en contraste con aquellos que procedían de ambientes más húmedos (Breed et al. 2012).

Inventario de la diversidad y relaciones tróficas

Las técnicas basadas en códigos de barras (*barcoding*) permiten la identificación a nivel de especie (Gross 2012) de diversos tipos de muestras recolectadas en campo a partir de secuencias secuencias cortas de ADN mitocondrial (en animales) o cloroplástico (en plantas) que (Kress et al. 2015); excrementos para la identificación de especies (Walker et al. 2016); restos estomacales para evaluar el contenido de la dieta (Kartzinel et al. 2015); muestras de

suelo para caracterizar la fauna edáfica (Orgiazzi et al. 2015); restos de polen recolectados del cuerpo de los polinizadores que han visitado flores en una comunidad (Ponron et al. 2016); o la identificación de las especies de frugívoros que dispersado semillas en el paisaje (González-Varo et al. 2014; Parejo-Farnés et al. 2018). Por tanto, además de ser una herramienta muy útil para inventariar la diversidad de especies o grupos taxonómicos estas técnicas están contribuyendo a mejorar nuestro conocimiento sobre las funciones de cada una de las especies al ecosistema que habita y de esta forma evaluar las consecuencias ecológicas del ritmo acelerado de pérdida de diversidad (Dirzo et al. 2014).

Cuestiones pendientes que se beneficiarían de la aplicación de la genética y genómica del paisaje

A pesar de los avances conceptuales y metodológicos en genética y genómica del paisaje la falta de integración entre la demografía y la genética en aspectos importantes sigue siendo una limitación

para muchos investigadores (Lowe et al. 2017). Por ejemplo, es habitual estudiar las trayectorias demográficas y genéticas de las poblaciones y de las especies de forma aislada sin tener en cuenta los efectos positivos (e.g. facilitación) y/o negativos (e.g. competencia) de las interacciones bióticas. Por tanto, la genética y genómica del paisaje todavía no resuelve cuestiones importantes a nivel de comunidad, aunque la aplicación de técnicas de barcoding para inventariar la diversidad de una comunidad o resolver relaciones tróficas puede ser un primer paso para escalar a nivel de comunidad (Corlett 2017). El salto del genotipado parcial del genoma con marcadores moleculares a la secuenciación masiva de genotipos completos o casi completos de cada individuo supone un salto cualitativo que abre nuevas vías posibilidades al estudio simultáneo de los procesos neutrales (e.g. flujo genético) y adaptativos (e.g. adaptación local). Esto permitirá caracterizar las dinámicas eco-evolutivas que subyacen en las trayectorias demográficas de las poblaciones naturales y extender este marco conceptual al estudio integrado del impacto de los factores ecológicos sobre la evolución de los caracteres de las especies y viceversa, i.e., el estudio del impacto de las dinámicas evolutivas sobre la ecología de las especies (Schoener 2011). Aproximaciones novedosas como las proyecciones de matrices integrales que incluyen cambios en las frecuencias alélicas, variaciones fenotípicas y heterogeneidad ambiental a lo largo de diferentes estadios demográficos serán de gran utilidad para este nuevo y necesario abordaje integrador (Kokko y López-Sepulcre 2007). Además, es el momento de utilizar las herramientas moleculares más allá de la identificación individual de las especies o de los individuos, por ejemplo, para evaluar de forma cuantitativa el efecto de la diversidad genética en los componentes cuantitativos y cualitativos de la dispersión de polen y semillas de las plantas polinizadas y dispersadas por animales (Valverde et al. 2019) o cuantificar el papel del flujo genético en las dinámicas de co-evolución de especies en paisajes heterogéneos mediante la integración de herramientas genéticas y teoría de redes (Medeiros et al. 2018). La integración de estas herramientas con la biología evolutiva, ecología y biología de la conservación nos permitirá conocer la composición y el funcionamiento de los ecosistemas para poder predecir las respuestas de las poblaciones al efecto conjunto de las actividades antrópicas y cambio climático.

Referencias

- Aben, J., Strubbe, D., Adriaensen, F., Palmer, S.C.F., Travis, J.M.J., Lens, L., Matthysen, E. 2014. Simple individual-based models effectively represent Afrotropical forest bird movement in complex landscapes Fuller, R. *Journal of Applied Ecology* 51: 693-702.
- Adamack, A.T., Gruber, B. 2014. PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution* 5: 384-387.
- Allendorf, F.W. 2017. Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* 26: 420-430.
- Allendorf, F.W., Hohenlohe, P.A., Luikart, G. 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics* 11: 697-709.
- Asner, G.P., Knapp, D.E., Anderson, C.B., Martin, R.E., Vaughn, N. 2016. Large-scale climatic and geophysical controls on the leaf economics spectrum. *Proceedings of the National Academy of Sciences* 113: E4043-E4051.
- Avise, J.C. 2004. *Molecular markers, natural history, and evolution*. Sinauer Associates, Sunderland, MA, Estados Unidos.
- Bacles, C.F.E., Jump, A.S. 2011. Taking a tree's perspective on forest fragmentation genetics. *Trends in Plant Science* 16: 13-18.
- Balkenhol, N., Cushman, S., Storfer, A., Waits, L. 2015. *Landscape genetics: Concepts, Methods, and Applications*. Wiley-Blackwell, Oxford, UK.
- Barbujani, G., Sokal, R.R. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the Academy of Natural Sciences* 87: 1816-1819.
- Beaumont, M.A., Rannala, B. 2004. The bayesian revolution in genetics. *Nature Reviews Genetics* 5: 251-261.
- Beerli, P., Felsenstein, J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations using a coalescent approach. *Proceedings of the National Academy of Science USA* 98 (8): 4563-4568
- Bocedi, G., Palmer, S.C.F., Pe'er, G., Heikkinen, R.K., Matsinos, Y.G., Watts, K., Travis, J.M. 2014. RangeShifter: a platform for modelling spatial eco-evolutionary and species' responses to environmental changes. *Methods in Ecology and Evolution* 5: 388-396.
- Breed, M.F., Gardner, M.G., Ottewill, K.M., Navarro, C.M., Lowe, A.J. 2012. Shifts in reproductive assurance strategies and inbreeding costs associated with habitat fragmentation in Central American mahogany. *Ecology Letters* 15: 444-452.
- Cavalli-Sforza, L.L., Feldman, M.W. 2003. The application of molecular genetic approaches to the study of human evolution. *Nature Genetics* 33: 266-275.
- Charlesworth, B., Charlesworth, D., Coyne, J.A., Langley, C.H. 2016. Hubby and Lewontin on protein variation in natural populations: when molecular genetics came to the rescue of population genetics. *Genetics* 203: 1497-1503.
- Chen, K.-Y., Marschall, E.A., Sovic, M.G., Fries, A.C., Gibbs, H.L., Ludsing, S.A. 2018. assignPOP: An R package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution* 9: 439-446.
- Corlett, R.T. 2017. A Bigger Toolbox: Biotechnology in Biodiversity Conservation. *Trends in Biotechnology* 35: 55-65.
- Cornuet, J.-M., Santos, F., Beaumont, M. a, Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T., Estoup, A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24: 2713-9.
- Csilléry, K., François, O., Blum, M.G.B. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* 3: 475-479.
- Cushman, S.A., Wasserman, T.N., Landguth, E.L., Shirk, A.J. 2013. Re-evaluating causal modeling with mantel tests in landscape genetics. *Diversity* 5, 51-72
- Dewoody, J., Nason, J.D., Hipkins, V.D. 2006. Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes* 6: 951-957.
- Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B., Collen, B. 2014. Defaunation in the Anthropocene. *Science* 345: 401-406.
- Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., Ovenden, J.R. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* 14: 209-214.
- Dray, S., Dufour, A. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software* 22 (4), 1-20.
- Eckert, A.J., Bower, A.D., González-Martínez, S.C., Wegrzyn, J.L., Coop, G., Neale, D.B. 2010. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology* 19: 3789-3805.
- Epperson, B.K. 2003. *Geographical genetics*. Levin, S. A., Horn, H. S. (eds.). Princeton University Press, Princeton.
- Ewing, G., Hermisson, J. 2010. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26: 2064-2065.
- Ewing, G.B., Reiff, P.A., Jensen, J.D. 2015. PopPlanner: visually constructing demographic models for simulation. *Frontiers in Genetics* 6: 150.
- Excoffier, L., Heckel, D.G. 2006. Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics* 7: 745-758.
- Faubet, P., Gaggiotti, O.E. 2008. A New Bayesian Method to Identify the Environmental Factors That Influence Recent Migration. *America* 1504: 1491-1504.
- Fedriani, J.M., Kohn, M.H. 2001. Genotyping faeces links individuals to their diet. *Ecology Letters* 4: 477-483.
- Fisher, R.A. 1930. *The Genetical Theory Of Natural Selection*. At the Clarendon Press.
- Fisher, R.A., Ford, E.B. 1947. The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* 1: 143-174.
- Ford, E.B. 1964. *Ecological genetics*. Chapman y Hall, Londres, Reino Unido.
- Frankham, R., Ballou, J.D., Ralls, K., Dudash, M.R., Fenster, C.B., Sunnucks, P. 2017. *Genetic management of fragmented animal and plant populations*. Oxford University Press. Oxford, Reino Unido.

- Frichot, E., François, O. 2015. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* 6: 925-929.
- Futuyma, D.J. 1998. *Evolutionary Biology*. Sinauer Associates, Sunderland, MA, Estados Unidos.
- García, C., Grivet, D. 2011. Molecular insights into seed dispersal mutualisms driving plant population recruitment. *Acta Oecologica* 37: 632-640.
- García, C., Arroyo, J.M., Godoy, J.A., Jordano, P. 2005. Mating patterns, pollen dispersal, and the ecological maternal neighbourhood in a *Prunus mahaleb* (L.) population. *Molecular Ecology* 14: 1821-1830.
- García, C., Rumeu, B., González-Pérez, M.A., Nogales, M. 2016. Reduced reproductive fitness of an endemic insular juniper population: An ecological mediation hypothesis. *Biological Conservation* 198, 70-77
- García, C., Jordano, P., Godoy, J.A. 2007. Contemporary pollen and seed dispersal patterns in a *Prunus mahaleb* population: patterns in distance and direction. *Molecular Ecology* 16: 1947-1955.
- García, C., Guichoux, E., Hampe, A. 2018. A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (*Juniperus phoenicea* ssp. *turbinata*). *Tree Genetics & Genomes* 14: 87.
- Gauzere, J., Klein, E.K., Oddou-Muratorio, S. 2013. Ecological determinants of mating system within and between three *Fagus sylvatica* populations along an elevational gradient. *Molecular Ecology* 22: 5015-5015.
- Geng, Q.F., Lian, C.L., Goto, S., Tao, J.M., Kimura, M., Islam, M.D.S., Hogetsu, T. 2008. Mating system, pollen and propagule dispersal, and spatial genetic structure in a high-density population of the mangrove tree *Kandelia candel*. *Molecular Ecology* 17: 4724-4739.
- Godoy, J.A., Jordano, P. 2001. Seed dispersal by animals: exact identification of source trees with endocarp DNA microsatellites. *Molecular Ecology* 10: 2275-2283.
- González-Varo, J.P., Arroyo, J.M., Jordano, P. 2014. Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. *Methods in Ecology and Evolution* 5: 806-814.
- Gross, M. 2012. Barcoding biodiversity. *Current Biology* 22: R73-R76.
- Guillaume, F., Rougemont, J. 2006. Nemo: an evolutionary and population genetics programming framework. *Bioinformatics* 22: 2556-2557.
- Guillot, G., Rousset, F. 2013. Dismantling the Mantel tests. *Methods in Ecology and Evolution* 4: 336-344.
- Guillot, G., Mortier, F., Estoup, A. 2005. GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes* 5: 712-715.
- Guillot, G., Santos, F., Estoup, A. 2008. Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* 24: 1406-1407.
- Hardesty, B.D., Hubbell, S.P., Bermingham, E. 2006. Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree. *Ecology Letters* 9: 516-525.
- Hardy, O., 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.
- Heikkinen, R.K., Bocedi, G., Kuussaari, M., Heliölä, J., Leikola, N., Pöyry, J., Travis, J.M.J. 2014. Impacts of land cover data selection and trait parameterisation on dynamic modelling of species' range expansion. *PLOS ONE* 9: e108436.
- Henry, R.C., Bocedi, G., Dytham, C., Travis, J.M.J. 2014. Inter-annual variability influences the eco-evolutionary dynamics of range-shifting. *PeerJ* 1: e228.
- Hether, T.D., Hoffman, E.A. 2012. Machine learning identifies specific habitats associated with genetic connectivity in *Hyla squirella*. *Journal of Evolutionary Biology* 25, 1039-1052.
- Hevroy, T.H., Moody, M.L., Krauss, S.L. 2017. Population genetic analysis reveals barriers and corridors for gene flow within and among riparian populations of a rare plant. *Annals of Botany Plants* 10: plx065.
- Hoban, S. 2012. Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics* 13: 110-122.
- Hoban, S., Bertorelle, G., Gaggiotti, O.E. 2012. Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics* 13: 110-122.
- Hoban, S., Gaggiotti, O., Consortium, C., Bertorelle, G., Borsari, V.L. 2013. Sample Planning Optimization Tool for conservation and population Genetics (SPOTG): a software for choosing the appropriate number of markers and samples. *Methods in Ecology and Evolution* 4(3): 299-303.
- Hodel, R.G.J., Segovia-Salcedo, M.C., Landis, J.B., Crowl, A.A., Sun, M., Liu, X., Gitzendanner, M.A. et al. 2016. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences* 4: 1600025.
- Isagi, Y., Suyama, Y. 2011. *Single-Pollen Genotyping*. Springer. Tokio, Japan.
- Jensen, J.D., Payseur, B.A., Stephan, W., Aquadro, C.F., Lynch, M., Charlesworth, D., Charlesworth, B. 2019. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution* 73: 111-114.
- Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405.
- Jombart, T., Devillard, S., Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94.
- Jones, F.A., Hubbell, S.P. 2006. Demographic spatial genetic structure of the Neotropical tree, *Jacaranda copaia*. *Molecular Ecology* 15: 3205-3217.
- Jordano, P. 2017. What is long-distance dispersal? And a taxonomy of dispersal events. *Journal of Ecology* 105: 75-84.
- Jordano, P., García, C., Godoy, J.A., García-Castaño, J.L. 2007. Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences of the USA* 104: 3278-3282.
- Kalinowski, S.T. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5: 187-189.
- Kalinowski, S.T., Taper, M.L., Marshall, T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099-1106.
- Kartzinel, T.R., Chen, P.A., Coverdale, T.C., Erickson, D.L., Kress, W.J., Kuzmina, M.L., Rubenstein, D.I., et al. 2015. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences* 112: 8019-8024.
- Kern, A.D., Hahn, M.W. 2018. The neutral theory in light of natural selection. *Molecular Biology and Evolution* 35: 1366-1371.
- Kimura, M., Crow, J.F. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49: 725-738.
- Klein, E.K., Carpentier, F.H., Oddou-Muratorio, S. 2011. Estimating the variance of male fecundity from genotypes of progeny arrays: evaluation of the Bayesian forward approach. *Methods in Ecology and Evolution* 2: 349-361.
- Kokko, H., López-Sepulcre, A. 2007. The ecogenetic link between demography and evolution: can we bridge the gap between theory and data? *Ecology Letters* 10: 773-782.
- Kress, W.J., García-Robledo, C., Uriarte, M., Erickson, D.L. 2015. DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution* 30: 25-35.
- Kuhner, M.K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22: 768-770.
- Landguth, E.L., Cushman, S.A. 2010. cdpop: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources* 10: 156-161.
- Lawson, D.J., Falush, D. 2012. Population identification using genetic data. *Annual Review of Genomics and Human Genetics* 13: 337-361.
- Lawson, D.J., Dorp, L. van, Falush, D. 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications* 9: 3258.
- Leblois, R., Estoup, A., Rousset, F. 2008. IBDsim: a computer program to simulate genotypic data under isolation by distance. *Molecular Ecology Resources* 9: 107-109.
- Lefcheck, J.S. 2016. piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution* 7: 573-579.
- Legendre, P., Fortin, M., Borcard, D. 2015. Should the Mantel test be used in spatial analysis?. *Methods in Ecology and Evolution* 6: 1239-1247.
- Lewontin, R.C. 1974. *The genetic basis of evolutionary change*. Columbia University Press. New York, USA.

- Lewontin, R.C., Moore, J.A., Provine, W.B., Wallace, B. 2003. Dobzhansky's Genetics of natural populations.
- Liaw, A., Wiener, M. 2002. Classification and Regression by random Forest. *R News* 2(3), 18-22.
- Lowe, A.J., Cavers, S., Boshier, D., Breed, M.F., Hollingsworth, P.M. 2015. The resilience of forest fragmentation genetics—no longer a paradox—we were just looking in the wrong place. *Heredity* 115: 97-99.
- Lowe, W.H., Kovach, R.P., Allendorf, F.W. 2017. Population genetics and demography unite ecology and evolution. *Trends in Ecology and Evolution* 32: 141-152.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S., Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4 (12): 981-994. Doi.: 10.1038/nrg1226.
- Malécot, G., Blaringhem, L. 1948. *Les mathématiques de l'hérédité*. Masson et Cie, Paris.
- Manel, S., Schwartz, M.K., Luikart, G., Taberlet, P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18: 189-197.
- Medeiros, L.P., Garcia, G., Thompson, J.N., Guimarães, P.R. 2018. The geographic mosaic of coevolution in mutualistic networks. *Proceedings of the National Academy of Sciences* 115: 12017-12022.
- Mondol, S., Bruford, Michael W., Ramakrishnan, U. 2013. Demographic loss, genetic structure and the conservation implications for Indian tigers. *Proceedings of the Royal Society B: Biological Sciences* 280: 20130496.
- Moodley, Y., Russo, I.-R.M., Dalton, D.L., Kotzé, A., Muya, S., Haubensak, P., Bálint, B., et al. 2017. Extinctions, genetic erosion and conservation options for the black rhinoceros (*Diceros bicornis*). *Scientific Reports* 7: 41417.
- Morente-López, J., García, C., Lara-Romero, C., García-Fernández, A., Draper, D., Iriondo, J.M. 2018. Geography and Environment Shape Landscape Genetics of Mediterranean Alpine Species *Silene ciliata* Poirlet. (Caryophyllaceae). *Frontiers in Plant Science* 9: 1698.
- Morin, P.A., McCarthy, M. 2007. Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes* 7: 937-946.
- Muñoz-Pajares, A.J., Garcia, C., Abdelaziz, M., Bosch, J., Perfectti, F., Gómez, J.M. 2017. Driver of genetic differentiation in a generalist insect-pollinated herb across spatial scales. *Molecular Ecology* 26, 1576-1585.
- Nagyilaki, T. 1994. Geographical variation in a quantitative character. *Genetics* 136: 361-381.
- Neale, D.B., Kremer, A. 2011. Forest tree genomics: growing resources and applications. *Nature Reviews* 12: 111-122.
- Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., Willerslev, E. 2017. Tracing the peopling of the world through genomics. *Nature* 541: 302-310.
- Nunes, M.A., Prangle, D. 2015. abctools: An R Package for tuning Approximate Bayesian Computation analyses. *The R Journal* 7: 189-205.
- Oksanen, F.J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. 2019. vegan: Community Ecology Package. *R package version* 2.5-4.
- Orgiazzi, A., Dunbar, M.B., Panagos, P., de Groot, G.A., Lemanceau, P. 2015. Soil biodiversity and DNA barcodes: opportunities and challenges. *Soil Biology and Biochemistry* 80: 244-250.
- Parejo-Farnés, C., Albaladejo, R.G., Camacho, C., Aparicio, A. 2018. From species to individuals: combining barcoding and microsatellite analyses from non-invasive samples in plant ecology studies. *Plant Ecology* 219: 1151-1158.
- Parobek, C.M., Archer, F.I., DePrenger-Levin, M.E., Hoban, S.M., Liggins, L., Strand, A.E. 2017. skeleSim: an extensible, general framework for population genetic simulation in R. *Molecular ecology resources* 17: 101-109.
- Peakall, R., Smouse, P.E. 2012. GenAIC 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28: 2537-2539.
- Pérez-Méndez, N., Jordano, P., García, C., Valido, A. 2016. The signatures of Anthropocene defaunation: cascading effects of the seed dispersal collapse. *Scientific Reports* 6: 24820-24820.
- Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., Pellizzari, C., et al. 2016. Using metabarcoding to reveal and quantify plant-pollinator interactions. *Scientific Reports* 6: 27282.
- Pournosrat, R., Kaya, S., Shaaf, S., Kilian, B., Ozkan, H. 2018. Geographical and environmental determinants of the genetic structure of wild barley in southeastern Anatolia. *PLOS ONE* 13(2): e0192386.
- Pritchard, J.K., Stephens, M., Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Quéméré, E., Amelot, X., Pierson, J., Crouau-Roy, B., Chikhi, L. 2012. Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. *Proceedings of the National Academy of Sciences* 109: 13028-13033.
- Rodríguez-González, P.M., García, C., Albuquerque, A., Monteiro-Henriques, T., Faria, C., Guimarães, J.B., et al. (en prensa). A spatial stream-network approach assists in managing the remnant genetic diversity of riparian forests. *Scientific Reports* 0:00-00. Doi.: 10.1038/s41598-019-43132-7.
- Rousset, F., Ferdy, J.B. 2014. Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography* 37(8): 781-790
- Ryman, N., Palm, S. 2006. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes* 6: 600-602.
- Ryman, N., Palm, S., Andre, C., Carvalho, G.R., Dahlgren, T.G., Jorde, P.E., Laikre, L., et al. 2006. Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology* 15: 2031-2045.
- Schrider, DR., Kern, A.D. 2018. Supervised machine learning for population genetics: a new paradigm. *Trends in Genetics* 34, 301-312
- Schlotterer, C. 2004. The evolution of molecular markers. *Nature Reviews Genetics* 5: 63-69.
- Schoener, T.W. 2011. The Newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science* 331: 426-429.
- Sharma, R., Arora, N., Goossens, B., Nater, A., Morf, N., Salmons, J., Bruford, M.W., et al. 2012. Effective population size dynamics and the demographic collapse of bornean orang-utans. *PLoS One* 7: e49429-e49429.
- Sharma, S., Dutta, T., Maldonado, J.E., Wood, T.C., Panwar, H.S., Seidens-ticker, J. 2013. Spatial genetic analysis reveals high connectivity of tiger (*Panthera tigris*) populations in the Satpura–Maikal landscape of Central India. *Ecology and Evolution* 3: 48-60.
- Sokal, R.R., Oden, N.L. 1978. Spatial autocorrelation in biology: 2. Some biological implications and four applications of evolutionary and ecological interest. *Biological Journal of the Linnean Society* 10: 229-249.
- Sokal, R.R., Smouse, P.E., Neel, J.V. 1986. The genetic-structure of a tribal population, the yanomama indians: patterns inferred by autocorrelation analysis. *Genetics* 114: 259-287.
- Strand, A.E. 2002. METASIM 1.0: an individual-based environment for simulating population genetics of complex population dynamics. *Molecular Ecology Notes* 2: 373-376.
- Strand, A., Niehaus, J. 2018. rmetasim: An individual-based population genetic simulation environment. Disponible en [https://rdrr.io/cran/rmetasim/]
- Sylvester, E.V.A., Bentzen, P., Bradbury, I.R., Clément, M., Pearce, J., Horne, J., Beiko, R.G. 2018. Applications of random forest feature selection for fine-scale genetic population assignment. *Evolutionary Applications* 11: 153-165.
- Valverde, J., Perfectti, F., Gómez, J.M. 2019. Pollination effectiveness in a generalist plant: adding the genetic component. *New Phytologist*, En prensa. doi:10.1111/nph.15743.
- Ver Hoef, J.M., Peterson, E.E., Clifford, D., Shah, R. 2014. SSN: An R package for spatial statistical modeling on stream networks. *Journal of Statistical Software* 56, 1-45. 2014.
- Wakely, J. 2016. *Coalescent Theory: An Introduction*. Macmillan Learning. Londres. Reino Unido.
- Walker, F.M., Williamson, C.H.D., Sanchez, D.E., Sobek, C.J., Chambers, C.L. 2016. Species From feces: order-wide identification of chiroptera from guano and other non-invasive genetic samples. *PLOS ONE* 11: e0162342.

- Wang, J. 2011. coancestry: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources* 11: 141-145.
- Wang, I.J. 2013. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* 67(12): 3403-3411.
- Wang, J. 2014. Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of Evolutionary Biology* 27: 518-530.
- Wang, J. 2017. Estimating pairwise relatedness in a small sample of individuals. *Heredity* 119: 302-313.
- Waples, R.S., Gaggiotti, O.E. 2006. What is a population? An empirical of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* 15: 1419-1439.
- Wegmann, D., Leuenberger, C., Neuenschwander, S., Excoffier, L. 2010. ABCtoolbox: a versatile toolkit for Approximate Bayesian Computations. *BMC Bioinformatics* 11: 116.
- Wei, S.-J., Shi, B.-C., Gong, Y.-J., Jin, G.-H., Chen, X.-X., Meng, X.-F. 2013. Genetic structure and demographic history reveal migration of the diamondback moth *Plutella xylostella* (Lepidoptera: Plutellidae) from the southern to northern regions of China. *PLOS ONE* 8: e59654
- Whitham, T.G., Bailey J.K., Schweitzer J.A., Shuster S.M., Bangert R.K., LeRoy C.J., et al., 2006. A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics* 7(7), 510-523. Doi.:10.1038/nrg1877
- Wilson, G.A., Rannala, B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163: 1177-1191.
- Wright, S. 1931. Evolution in Mendelian Populations. *Genetics* 16: 97-159.
- Wright, S. 1943. Isolation by distance. *Genetics* 28: 114-138.

Anexo

Identificación y estimas de errores de genotipado

Métodos de identificación y cuantificación de errores de genotipado

Una de las tareas más importantes, aunque tediosas, del trabajo de genotipado es la cuantificación de las tasas de error y, en la medida de lo posible la identificación de su origen. Los errores de genotipado se producen cuando el genotipo multilocus obtenido difiere del genotipo real (Pompanon et al. 2005) y hay varias fuentes de error a lo largo del proceso de genotipado que pueden incrementar la probabilidad de errores en el genotipado, especialmente cuando se trabaja con microsatélites. Una baja cantidad y calidad del ADN extraído aumenta enormemente la probabilidad de errores en el genotipado y, por tanto, es muy importante obtener ADN de alta calidad y con una concentración adecuada para minimizar la probabilidad de errores durante la amplificación y posterior asignación de alelos (scoring). La presencia de inhibidores de la polimerasa en el extracto de ADN aumenta su actividad defectuosa y, por tanto, los tejidos a partir de los cuales se extrae ADN deben estar limpios de inhibidores. Compuestos secundarios frecuentes en las hojas de muchas especies de plantas tropicales o mediterráneas actúan de inhibidores. Los errores de pipeteado y de scoring son frecuentes y deben minimizarse para asegurar que el conjunto de genotipos multilocus final representa una muestra fidedigna de los genotipos de la población.

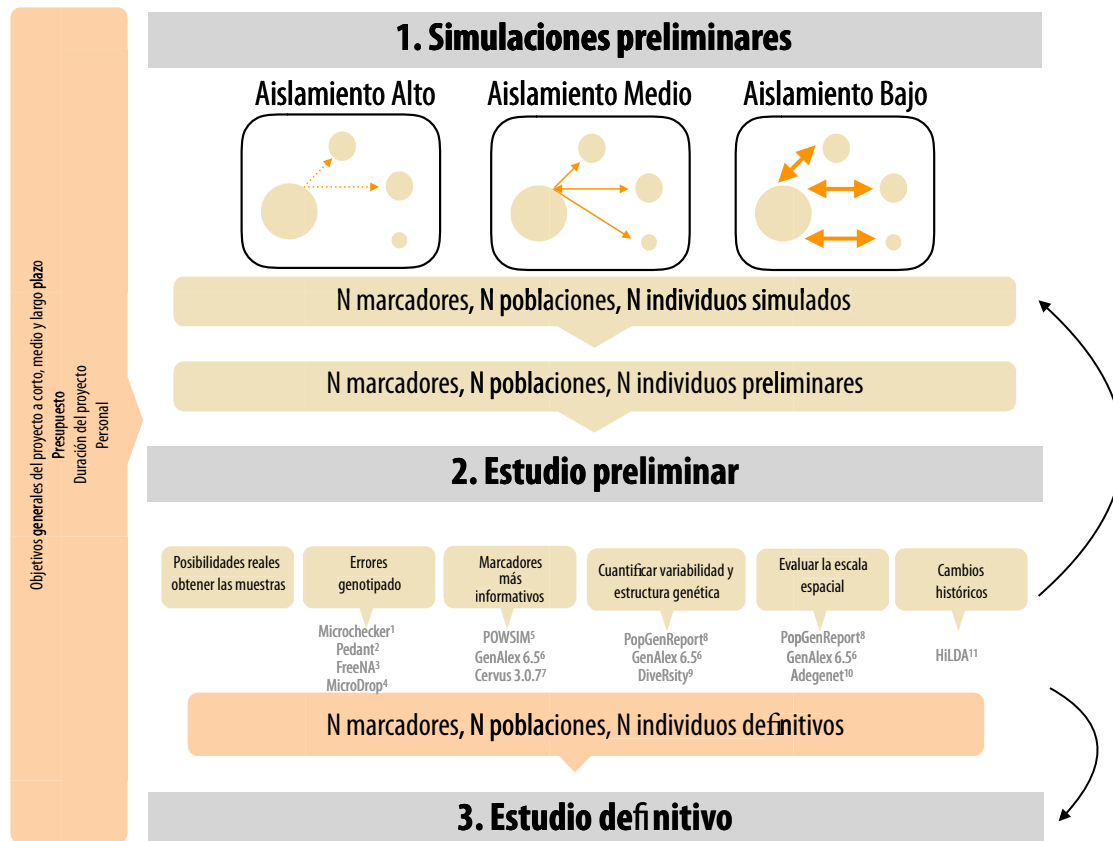
Hay diferentes tipos de errores de genotipado y se pueden clasificar a gran escala como aquellos que producen falta de datos (missing data) y aquellos que introducen datos incorrectos. Entre los primeros están los alelos nulos (null alleles, NA) y el abandono alélico (allelic dropout, ADO). Los alelos nulos ocurren cuando uno de los alelos de un locus heterocigoto no consigue amplificarse y, por tanto, el locus aparece como homocigoto representado por el único alelo que ha conseguido amplificar. Este tipo de errores suele estar asociado a una mutación en la secuencia complementaria de uno de los cebadores que impide una se acople a la cadena de ADN correctamente. El abandono alélico es otra fuente frecuente de errores de genotipado y ocurre cuando de forma estocástica uno de los alelos de un locus no consigue amplificar. Entre los errores que introducen datos incorrectos están los alelos falsos (false alleles, FA) generalmente debidos a errores de identificación de alelos (scoring) cuando se asigna un alelo a un pico que en realidad no lo es (artefacto). La presencia de alelos falsos es frecuente en aquellos marcadores microsatélites que presentan varios picos anteriores o posteriores al pico identificado como alelo (stutter). El stutter se puede reducir optimizando las condiciones de la PCR y generalmente reducir el número de ciclos, disminuir la concentración de MgCl₂, aplicar un ciclo tipo "Touch-down" o aplicar directamente un "hot start Taq", añadir seroalbúmina bovina (BSA) y/o reducir el tiempo final de extensión suele reducir el la frecuencia e intensidad del stutter. Estas modificaciones en la composición y en los ciclos de la PCR también ayudan a reducir el ruido de las amplificaciones que suele ser alto cuando la amplificación no es buena debido generalmente a la baja concentración o baja calidad (DNA degradado o extracción de baja calidad con muchos inhibidores) del DNA. Por último, hay errores de interpretación en la asignación de picos que lleva a identificar alelos erróneos (mistaken alleles, MA). Esto ocurre con frecuencia en microsatélites con repeticiones en tándem de motivos dinucleótidos cuando el patrón de picos puede dar lugar a alelos intermedios que no existen. Por ello, para asegurar que la interpretación es correcta, es siempre conveniente revisar repetidamente el patrón de picos por al menos dos personas de forma independiente y prestar atención a aquellos alelos en los que no haya consenso en la asignación (Pompanon et al. 2005).

Para reducir la frecuencia de los errores de genotipado es recomendable aplicar de forma sistemática el siguiente protocolo (Pompanon et al. 2005; Dewoody et al. 2006): (i) en la fase de muestreo y almacenamiento de muestras, se recomienda preservar la muestra de forma que se limite la degradación del DNA. El uso de nitrógeno líquido de forma inmediata preserva correctamente las muestras. Alternativamente, el uso de silica gel y secado de las muestras que evite la presencia de hongos es esencial; (ii) en la fase de extracción, modificar el método de extracción para reducir presencia de inhibidores y aumentar el DNA de alta calidad; (iii) la PCR se debe optimizar para conseguir un patrón de picos lo más limpio posible, libre de ruido y de stutter; (iv) se recomienda genotipar varias veces (al menos 2) un 10%-20% de las muestras para poder estimar posteriormente las tasas de error basados en medidas repetidas. Es importante tener en cuenta este aumento de muestras genotipadas en el presupuesto del proyecto; (v) revisar repetidamente a identificación de alelos por al menos dos personas de forma independiente; (vi) aplicar software adecuado para identificar y cuantificar los casos de ADO, NA y FA y reportar sus respectivas frecuencias (Pompanon et al. 2005); (vii) una vez cuantificadas las tasas de error tratar de reducirlas al máximo (por ejemplo, modificando los protocolos de extracción y PCR si fuera necesario). También puede ser de interés eliminar alguna muestra problemática o algún marcador problemático que consistentemente introduzca ADO, FA o NA; (viii) una vez cuantificadas las tasas de error, incorporarlas a los sucesivos análisis de datos. Por ejemplo, INEST 2.0 (Chybicki et al. 2011) calcula estimas de la heterocigosidad observada y esperada (H_o , H_e) corregidas por la frecuencia de alelos nulos en los datos. Además, tiene en cuenta el nivel de endogamia a la hora de estimar la frecuencia de los alelos nulos por lo que produce estimas de frecuencia de alelos nulos más certeras que otros métodos (Campagne et al. 2012).

Todos los datos de genotipos multilocus tendrán un error de genotipado que es imposible eliminar del todo. El investigador debe procurar minimizar este porcentaje de error y eliminar del estudio aquellas muestras o marcadores sistemáticamente produzcan una alta tasa de error. Una vez reportado la tasa de error correspondiente, ésta debería incorporarse a posteriores análisis utilizando métodos de análisis que incorporen una tasa de error en sus estimaciones (Figura A1; Tabla A1).

Consecuencias de los errores de genotipado

Los errores de genotipado pueden conducir a errores en los resultados obtenidos y interpretaciones equivocadas sobre los patrones y procesos presentes en nuestro sistema de estudio. Los errores que producen falta de datos (FA, MA) típicamente llevan a una sobre estimación de la homocigosidad de las poblaciones porque la frecuencia de heterocigotos será menor de la esperada (Pompanon et al. 2005). Si la frecuencia de FA o NA es alta podríamos detectar falsos cuellos de botella, estimar niveles de endogamia por encima de los reales o subestimar la magnitud del flujo génico. Además, si este error es más frecuente en algunas poblaciones podríamos inferir falsas estructuras genéticas en el paisaje (Morin et al. 2009) o incluso concluir posibles procesos de selección natural. Los análisis de paternidad/maternidad son muy sensibles a este tipo de errores (incluso cuando se acepta algún marcador no coincidente o *mismatch* entre los genotipos de los parentales y la descendencia). Por ejemplo, (Hoffman y Amos 2005) estimaron que tasas bajas de error ~1% resultan en un 20% de errores de asignación en los estudios de paternidad y suelen llevar a conclusiones erróneas sobre la varianza reproductiva entre individuos de una población. Los errores que introducen alelos erróneos (FA, MA) nos pueden llevar a la conclusión de alta diferenciación (i.e., estructura) debido a que generalmente sobreestiman la presencia y la frecuencia de alelos raros. En general, los análisis de estructura espacial y asignación a poblaciones son robustos a una frecuencia moderada o baja de alelos nulos, siempre y cuando éstas no afecten de forma diferencial a ciertos marcadores o poblaciones (Cornuet et al. 1999). Sin embargo, los análisis de paternidad y maternidad son más sensibles a la presencia de alelos nulos e incluso tasas moderadas de alelos nulos (~20%) reducen a la mitad el poder de asignación a los parentales correctos (Dakin y Avise 2004; Hoffman y Amos 2005) lo cual aumenta en un 18% la frecuencia en la que se excluyen los parentales correctos.



¹Van Oosterhout et al. (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535-538 ²Johnson & Haydon (2007) *Bioinforma. Biol. Insights* **1**, 71-75 ³Chapuis & Estoup (2007) *Mol. Biol. Evol.* **24**, 621-631. ⁴Wang, C., et al. (2012) *Genetics* **192**, 651-669 ⁵Ryman & Palm (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Mol. Ecol. Notes* **6**, 600-602 ⁶Peakall & Smouse (2012) GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics* **28**, 2537-2539 ⁷Kalinowski (2007) *Mol. Ecol.* **16**, 1099-1106 ⁸Adamack & Gruber (2014) *Methods Ecol. Evol.* **5**, 384-387 (2014) ⁹Keenan et al (2013) *Methods Ecol. Evol.* **4**, 782-788 ¹⁰Jombart (2008) *Bioinformatics* **24**, 1403-1405 ¹¹Fuchs et al. (2015) Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. *Glob. Change Biol.* **21**, 299-313.

Figura A1. Protocolo para cuantificar de forma robusta los errores de genotipado.

Figure A1. Protocol to quantify genotyping errors in a robust way and to identify the source of those errors

Tabla A1. Listado de software que tienen en cuenta las tasas de error de genotipado para estimar diversos parámetros de genética de poblaciones.

Table A1. List of software that accomodates genotyping errors to estimate different parameters of interest in population genetics.

Software	Referencia	Tipo de error	Tipo de análisis	URL
INEST 2.0	Chybicki et al. (2011)	Allelos nulos (NA)	Estimas de heterocigosidad observada y esperada (H_o , H_e) y F_{is} corregidas por la frecuencia de alelos nulos en los datos.	https://www.ukw.edu.pl/pracownicy/stro na/igor_chybicki/software_ukw/
FreeNA	Chapuis y Estoup (2007)	Allelos Nulos (NA)	Estimas de frecuencia de alelos nulos y estimas de F_{st} corregidas por la frecuencia de alelos nulos (unbiased F_{st})	https://www1.montpellier.inra.fr/CBGP/software/FreeNA/
Cervus 3.0.7	Kalinowski et al. (2007)	Frecuencia global de error de genotipado sin especificar qué tipo de error	Paternidad, maternidad y parentesco. Posibilidad de estimar el porcentaje de asignación al parental/parentales correcto en función de tasas de error crecientes para decidir qué tasa de error es tolerable en el contexto del estudio.	http://www.fieldgenetics.com/pages/aboutCervus_Overview.jsp
Micro-Drop	Wang et al. (2012)	Estimas de frecuencias alélicas corregidas en presencia de ADO	Las estimas alélicas corregidas se pueden utilizar posteriormente en multitud de análisis de genética de poblaciones	http://rosenberglab.stanford.edu/software.html

Referencias del anexo

- Campagne, P., Smouse, P.E., Varouchas, G., Silvain, J.-F., Leru, B. 2012. Comparing the van Oosterhout and Chybicki-Burczyk methods of estimating null allele frequencies for inbred populations. *Molecular Ecology Resources* 12: 975-982.
- Chapuis, M.P., Estoup, A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24:621-631
- Chybicki, I.J., Oleksa, A., Burczyk, J. 2011. Increased inbreeding and strong kinship structure in *Taxus baccata* estimated from both AFLP and SSR data. *Heredity* 107: 589-600.
- Cornuet, J.-M., Piry, S., Luikart, G., Estoup, A., Solignac, M. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153: 1989-2000.
- Dakin, E.E., Avise, J.C. 2004. Microsatellite null alleles in parentage analysis. *Heredity* 93: 504-509.
- Dewoody, J., Nason, J.D., Hipkins, V.D. 2006. Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes* 6: 951-957.
- Hoffman, J.I., Amos, W. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* 14: 599-612.
- Kalinowski, S.T., Taper, M.L., Marshall, T.C. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099-1106.
- Morin, P.A., Leduc, R.G., Archer, F.I., Martien, K.K., Huebinger, R., Bickham, J.W., Taylor, B.L. 2009. Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources* 9: 498-504.
- Pompanon, F., Bonin, A., Bellemain, E., Taberlet, P. 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews* 6: 847-859.
- Wang, C., Schroeder, K.B., Rosenberg, N.A. 2012. A maximum likelihood method to correct for allelic dropout in microsatellite data with no replicate genotypes. *Genetics* 192, 651-669.