

PUESTA A PUNTO DE MARCADORES MOLECULARES PARA INVESTIGAR LA BIOGEOGRAFÍA DEL MADROÑO (*ARBUTUS UNEDO* L.)

Paula Catarina Sánchez Pedreira

RESUMEN

Arbutus unedo es un árbol de la familia Ericaceae ampliamente distribuido por el Mediterráneo pero que también presenta algunas poblaciones disjuntas en la región Atlántica (Irlanda). En este trabajo se buscaron marcadores polimórficos y fácilmente amplificables del cpADN (ADN del cloroplasto) útiles para estudiar la filogeografía de esta especie. Además, se realizó una evaluación preliminar de la variabilidad haplotípica de esta planta. Para ello, obtuvimos secuencias de árboles muestreados en 10 poblaciones a lo largo de su rango de distribución (desde Irlanda hasta Turquía). Se ensayaron cinco fragmentos no codificantes del cpADN, de los cuales cuatro resultaron ser filogeográficamente útiles: Intrón *L*, *trnL-rpl32*, *trnH-psbA*, y *trnT-trnL*. Para obtener secuencias legibles de estos cuatro fragmentos, fue necesario utilizar una polimerasa de alta fidelidad debido a la presencia repetida de largas regiones poli-A. Los resultados obtenidos indican que las poblaciones Atlánticas tienen menor diversidad genética que las Mediterráneas. El examen de las relaciones filogenéticas entre haplotipos sugiere la existencia de tres grupos genéticos a lo largo del rango de distribución de *Arbutus unedo*: Atlántico-Norte de África, Mediterráneo Occidental y Mediterráneo Oriental. Las muestras recogidas en Irlanda están filogenéticamente relacionadas con las del norte de la Península Ibérica pero no con las del Atlántico francés.

INTRODUCCIÓN

Filogeografía es un término acuñado en 1987 por Avise *et al.* para describir una disciplina relativamente nueva y actualmente en expansión que trata de la disposición espacial de los linajes genéticos dentro y entre especies muy relacionadas (Avise *et al.* 2009). La filogeografía se puede dividir en intraespecífica e interespecífica. La filogeografía intraespecífica observa cómo están distribuidos geográficamente los distintos haplotipos presentes dentro de una misma especie mientras que la interespecífica estudia la disposición de linajes genéticos en distintas especies, pero muy relacionadas entre sí. Los estudios de filogeografía intraespecífica están siendo cada vez más aceptados para focalizarlos hacia actuaciones para la conservación de especies, pues estos estudios permiten obtener una medida de diversidad genética, facilitando su gestión para mantener el potencial evolutivo de las especies (Newton 1999).

Un principio básico en filogeografía es el de la concordancia genealógica (Avise 2000). La presencia de dicha concordancia se considera una evidencia de que los factores responsables de una estructura filogeográfica han debido ser muy fuertes (ej. fragmentación de las poblaciones durante una glaciación). Se pueden distinguir, al menos, cuatro aspectos de la concordancia genealógica. El primero es la concordancia intra-locus: detectar múltiples sustituciones nucleotídicas a lo largo de una secuencia de ADN que de manera consistente distingue unos haplotipos de otros. Esto evidencia que ha ocurrido una fuerte división histórica puesto que la acumulación de nuevas mutaciones es un proceso bastante lento. El segundo es la concordancia multi-locus: detectar una profunda división en múltiples genes distinguiendo los mismos conjuntos de poblaciones. Esto sugiere la existencia de fuerzas evolutivas capaces de provocar efectos genómicos generalizados (ej. impedimento perdurable para la dispersión). Un tercer aspecto es el multi-específico: observar las mismas divisiones genealógicas espaciales en múltiples especies. Esta concordancia multi-específica indica que las fuerzas evolutivas responsables deben haber tenido amplias repercusiones en el ámbito de las comunidades bióticas o ecosistemas. Finalmente, otro aspecto de la concordancia hace referencia a la coincidencia entre distintas líneas de evidencia empírica (ej. morfología, distribución de subespecies, geología).

Históricamente, los estudios filogeográficos comenzaron cuando fue posible obtener datos del ADN mitocondrial (mtADN). La información del mtADN daba una perspectiva genealógica a la estructura de las poblaciones de una especie. El mtADN es especialmente apropiado para los estudios filogeográficos de animales por varios motivos. Por un lado, las secuencias del mtADN de los animales suelen evolucionar muy rápidamente (mucho más rápido que otros genomas) lo que hace que una especie suela presentar muchos haplotipos. Por otra parte, el mtADN se transmite de manera no recombinante, lo que facilita ordenar filogenéticamente esos haplotipos (i.e. descubrir qué haplotipos derivan de otros haplotipos). En la mayoría de los casos, se trata de una herencia maternal, de manera que las secuencias del mtADN describen la ascendencia materna de un individuo. Desde un principio, se encontró que distintas especies mostraban patrones filogeográficos diferentes (ej. especies con una clara subdivisión en dos o más grupos de subpoblaciones, especies sin ninguna subdivisión a lo largo de su rango geográfico) que guardaban relación tanto con la movilidad de esas especies como con la fragmentación de su ambiente (pasada y presente). Una observación frecuente en muchos estudios es que las diferencias genealógicas entre poblaciones de distintas regiones suelen ser mucho más profundas que las diferencias entre poblaciones de una misma región. Esto llevó al concepto de Unidades Evolutivamente Significativas (*Evolutionary Significant Units*, ESUs) para identificar grupos matrilineales que están claramente diferenciados de otros grupos matrilineales de la misma especie (Moritz 1994).

A diferencia de los animales, el ADN mitocondrial (mtADN) no es el más apropiado para este tipo de estudios en plantas ya que presenta una tasa de evolución muy lenta (aproximadamente 100 veces más despacio que en animales) (Avice et al. 2009). Por ello, los estudios filogeográficos en plantas se deben llevar a cabo con marcadores genéticos de otros genomas como el nuclear (nADN) o el del cloroplasto (cpADN). Este último evoluciona tres o cuatro veces más rápido que el mtADN, aunque sigue siendo mucho más lento que el mtADN de animales (Avice et al. 2009). El nADN se dispersa principalmente por el polen, por lo que la información derivada de los marcadores de este genoma puede ser útil para estudiar el estado actual de la especie (historia reciente). Por otra parte, a diferencia del cpADN y mtADN, las secuencias del nADN sí están sometidas a

recombinación lo que complica enormemente la reconstrucción de su filogenia. Esto explica que, con diferencia, una mayoría de estudios filogeográficos con plantas se hayan centrado en el cpADN. Al igual que el mtADN, el cpADN se dispersa únicamente por las semillas pues, en general, es heredado maternalmente (excepto en las Gimnospermas). Otra ventaja que tiene el uso del cpADN es que la mayoría de los individuos presentan una única secuencia haplotípica y varios individuos dentro de una misma especie a veces muestran diferencias detectables en las secuencias de cpADN (Avise et al. 2009). En el presente trabajo, como el objetivo es poner las bases para desentrañar las posibles rutas migratorias responsables del patrón de distribución actual de un árbol típico de la flora Mediterránea (*Arbutus unedo*), nos centramos en obtener marcadores filogeográficamente informativos de su cpADN para obtener la historia antigua de la especie. Para ello, prestamos atención especial a marcadores radicados en regiones no codificantes del cpDNA. Generalmente, las regiones no codificantes tienden a evolucionar más rápido que las secuencias codificantes, por lo que previsiblemente generarán un mayor número de haplotipos.

La especie objeto de este estudio es el madroño (*Arbutus unedo* L. 1753), perteneciente a la familia Ericaceae y con dotación genética diploide. Se trata de un arbusto o árbol pequeño que puede llegar a medir hasta 9 metros de altura (tamaño que puede tardar en alcanzar hasta 25 años). Florece entre septiembre y enero, presentando flores hermafroditas y polinización entomógama. La dispersión de la semilla es mediante zoocoria (concretamente endozoocoria), principalmente a principios del invierno. Esto se debe a que las semillas se encuentran en el interior de un fruto carnoso (baya de 10-20 mm de diámetro, que tarda un año en madurar) de color rojo anaranjado cuando está maduro, por lo que atrae a mamíferos y aves que son los encargados de dispersar las semillas. Cada fruto contiene numerosas semillas de tamaño pequeño y color pardo. La pulpa del fruto contiene inhibidores de la germinación, fenómeno denominado latencia química (Nikolaeva, 1997). La endozoocoria asegura tanto la dispersión de las semillas a largas distancias, como la liberación de las semillas de la pulpa del fruto, permitiendo así la germinación de éstas.

El madroño tiene un buen desarrollo en lugares de clima suave y húmedo durante el invierno, y seco y cálido durante el verano. Puede crecer en suelos

alcalinos o relativamente ácidos. *Arbutus unedo* puede resistir temperaturas de hasta -12°C en invierno (Mereti et al., 2001), de manera que esta especie muestra una fuerte resistencia a condiciones ambientales desfavorable (Özcan et al., 2007). El área de distribución natural de *Arbutus unedo* L. es típicamente circunmediterránea (España, Francia, Italia, Albania, Grecia, antigua Yugoslavia, Córcega, Cerdeña, Sicilia, Baleares, Creta, Norte de África, Palestina), estando limitado a zonas costeras y penetrando hacia el interior sólo en lugares con clima benigno. De hecho, las poblaciones más extensas de la Península Ibérica se encuentran en zonas Atlánticas con influencia oceánica, suelos oligotróficos y lluvia abundante, mientras que es muy escaso o inexistente en zonas frías muy continentales del interior (Torres et al. 2003, Villar 1993). Fuera de la cuenca Mediterránea, también hay poblaciones en Irlanda, mostrando así un patrón de distribución disjunto conocido como distribución lusitánica (i.e. especies presentes en Irlanda y la Península Ibérica pero que no habitan en Gran Bretaña). El origen de esta distribución disjunta sigue siendo motivo de discusión. Más recientemente, el madroño se puede encontrar también en Macaronesia (Islas Canarias) donde ha sido introducido por el hombre.

Los objetivos del presente trabajo son: 1. Identificar y poner a punto marcadores polimórficos del cpADN para estudiar la filogeografía de *Arbutus unedo*; 2. Obtener una evaluación preliminar de la variación haplotípica de *Arbutus unedo* a lo largo de su rango de distribución; 3. Investigar si los datos obtenidos proporcionan algún indicio sobre el origen de la distribución actual de *Arbutus unedo*, prestando especial atención al caso de las poblaciones disjuntas de Irlanda.

MATERIAL Y MÉTODOS

Muestreo y extracción de ADN

El estudio se hizo con muestras recogidas en diez localidades que cubren todo el rango de distribución de la especie (Fig. 1). En cada localidad, se recogieron hojas de 24-30 árboles dejando una distancia entre árbol y árbol de 10 metros. Para el presente estudio preliminar, se seleccionaron al azar 4-5 muestras (árboles) de cada población. Las hojas recogidas se secaron en silica gel y se almacenaron en oscuridad hasta el momento de la extracción del ADN. El ADN se obtuvo

usando un kit de extracción específico para plantas y hongos (Realpure Genomic ADN, REAL) siguiendo las instrucciones del fabricante. Las extracciones se hicieron en tandas de extracción de 22 muestras combinando plantas de diferentes poblaciones más 2 controles negativos. La cantidad e integridad del ADN extraído y de los controles negativos se examinó por medio de geles de agarosa (1.5%).

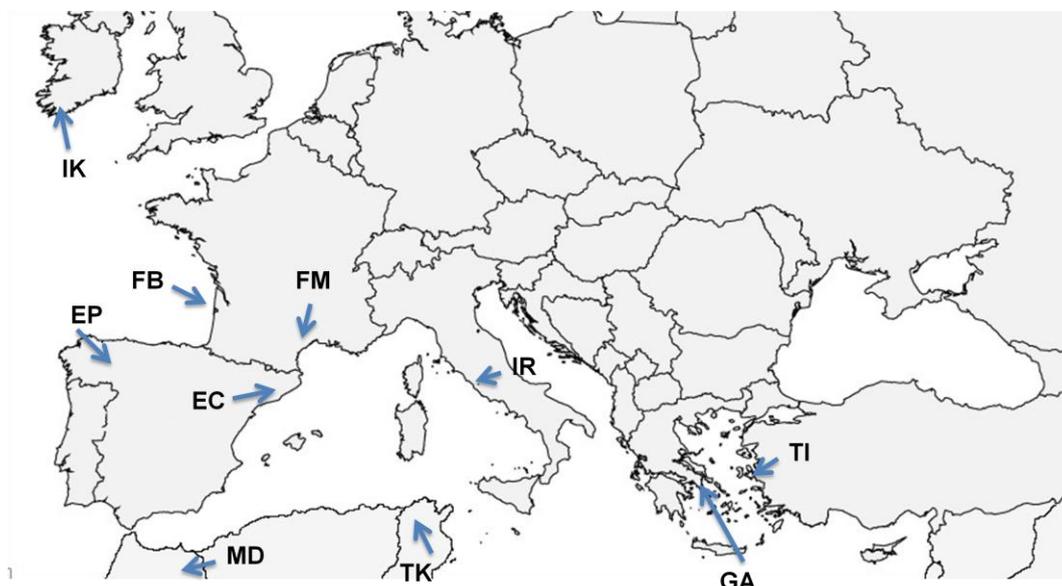


Fig. 1. Localidades de muestreo de *Arbutus unedo*. IK (Irlanda_Killarney), FB (Francia_Burdeos), EP (España_Ponferrada), MD (Marruecos_Debdou), TK (Túnez_Kroumerie), FM (Francia_Montpellier), EC (España_Cataluña), IR (Italia_Roma), TI (Turquía_Izmir), GA (Grecia_Atenas)

Secuenciación

Tras unas pruebas iniciales, se determinó que la concentración de ADN ideal para las reacciones de secuenciación era con el ADN extraído diluido 1:10. Todos los productos de PCR fueron visualizados en geles de agarosa al 1.5% y posteriormente purificados con 1 μ L de Exonucleasa I (20 U/ μ L) y 2 μ L de FastAP (10 U/ μ L; Fermants, Waltham, Madison, USA) antes de su secuenciación.

Tras una revisión de estudios filogeográficos con otros árboles, se seleccionaron cinco regiones no codificantes del cpADN para evaluar su facilidad de amplificación y polimorfismo en *Arbutus unedo*: *trnS*(GCU)-*trnG*(UUC), *trnH*(GUG)-*psbA*, *trnLUA2-rpL32F2*, *trnT*(UGU)-*trnL*(UAA), Intrón *L*. Inicialmente, las reacciones de amplificación se hicieron siguiendo las

condiciones que habían sido usadas previamente por otros autores para estos mismos fragmentos del cpADN. Así, el Intrón *L* y de la región *trnT-trnL* se amplificaron siguiendo a Taberlet et al. (1991). En concreto, las reacciones de amplificación del espaciador intergénico *trnT-trnL* se realizaron en 25 μ L con los cebadores a y b de Taberlet et al. (1991). Cada reacción contenía 1x Buffer, 2 mM/L $MgCl_2$, 0.2 mM/L de cada dNTP, 0.5 μ mol/L de cada cebador, 1 μ L de ADN, y 0.05 U/ μ L de *Taq* polimerasa. Para la amplificación del Intrón *L* se usaron los cebadores c y d de Taberlet et al. (1991). La mezcla de PCR para este fragmento fue 1x Buffer, 1.5 mM/L $MgCl_2$, 0.2 mM/L de cada dNTP, 0.5 μ mol/L de cada cebador, 1 μ L de ADN diluido (1:10) y 0.018 U/ μ L de *Taq* polimerasa. Por otra parte, la amplificación de los otros tres fragmentos se hizo siguiendo a Shaw et al. (2005) (*trnS-trnG* y *trnH-psbA*) y Shaw et al. (2007) (*trnL-rpl32*). En los tres casos, la amplificación se hizo con 1x Buffer, 2 mM/L $MgCl_2$, 0.2 mM/L de cada dNTP, 0.5 μ mol/L de cada cebador, 1 μ L de ADN, y 0.06 U/ μ L de *Taq* polimerasa. Para los cinco fragmentos se siguió un perfil de PCR estándar consistente en una desnaturalización a 94°C durante 2'; 35 ciclos de 94°C durante 1', 50°C durante 1', y 72°C durante 90"; y un paso final de elongación de 3' a 72°C.

La presencia generalizada de problemas en la secuenciación (secuencias ilegibles), obligó a buscar protocolos alternativos capaces de generar secuencias legibles. Éstas se obtuvieron utilizando una polimerasa de alta fidelidad (Phusion High-Fidelity ADN Polymerase, Thermo Scientific). Todas las reacciones se hicieron en un volumen de 20 μ L con una mezcla de PCR que contenía: 10 μ L de 2x Phusion Master Mix (incluye 1U de Phusion High-Fidelity ADN Polymerase), 0.5 μ mol/L de cada cebador, y 1 μ L de ADN. El perfil de termociclado consistió en una desnaturalización a 98°C durante 30"; 30 ciclos de 98°C durante 10", 20" a la temperatura de *annealing* de cada par de cebadores, y 72°C durante 30"; y un paso final de elongación de 7' a 72°C. Las temperaturas de *annealing* variaron dependiendo del fragmento a amplificar: 60.8°C para el fragmento *trnT-trnL*, 59.7°C para el *trnL-rpl32*, 61.8°C para el Intrón *L*, 71.8°C para el *trnS-trnG*, y 66.8°C para el fragmento *trnH-psbA*.

Todos los fragmentos purificados se secuenciaron bidireccionalmente en un servicio externo (Macrogen Europe) utilizando BigDye Terminator en un

secuenciador automático 3730XL (Applied Biosystems). La calidad de las secuencias obtenidas se examinó con la ayuda del programa Geneious R6 v.6.1.4 creado por Biomatters Ltd. (www.geneious.com). El mismo programa se utilizó para alinear las secuencias de los distintos individuos.

Análisis de los datos

Dada la naturaleza no recombinante del cpADN, los análisis se hicieron para cada fragmento por separado así como para las cuatro regiones que presentaron polimorfismo concatenadas como una única secuencia. Las estimas de diversidad genética (nº haplotipos, diversidad haplotípica y nucleotídica, y número de sustituciones e indels) para cada población y para el conjunto de los datos se obtuvieron con la ayuda del programa Arlequín v.3.5.1.2 (Excoffier & Lischer 2010). El alineamiento de las secuencias concatenadas se convirtió a formato Roehl con la ayuda de DnaSP v.5 (Librado & Rozas 2009) para cargarlo en el software Network v.4.6.1.1 (disponible en fluxus-engineering.com). En Network, se construyó una red haplotípica usando el algoritmo *median-joining* (MJ) con la opción de máxima parsimonia (Polzin et al. 2003). Previamente, los huecos de varios indels contiguos se codificaron como presencia/ausencia y se les dio un peso cuatro veces superior a las sustituciones nucleotídicas o a las indels de un solo nucleótido.

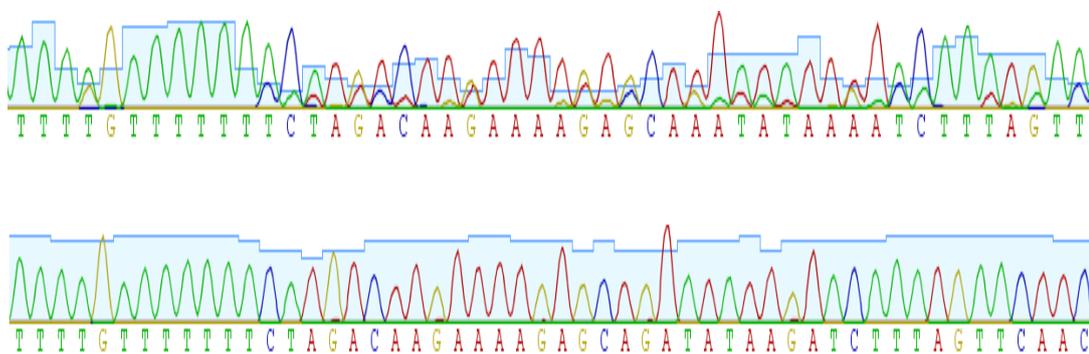


Fig. 2. Influencia del tipo de polimerasa sobre la calidad de las secuencias de regiones no codificantes de cpADN en *Arbutus unedo*. Arriba, secuencia de mala calidad obtenida con una *Taq* polimerasa (AmpliTaq DNA Polymerase, Applied Biosystems); abajo, la misma secuencia obtenida con una polimerasa de alta fidelidad (Phusion High-Fidelity DNA Polymerase, Thermo Scientific).

RESULTADOS

Influencia del tipo de polimerasa en la calidad de la secuenciación

La presencia de largos fragmentos poli-A causaron problemas de secuenciación (secuencias ilegibles) al utilizar la *Taq* polimerasa. El examen de los cromatogramas obtenidos reveló que los problemas de legibilidad aparecían justo a partir de la primera región poli-A que encontraba la polimerasa (Fig. 2 arriba). Esto sugería problemas de fidelidad (*slippage*) durante la copia con la *Taq* polimerasa. Una búsqueda de protocolos alternativos para solucionar este problema nos llevó a concluir que la mejor alternativa podría ser utilizar una polimerasa de alta fidelidad (Fazekas et al. 2010, Thomas et al. 2011). Al aplicar una polimerasa de este tipo a nuestros fragmentos comprobamos que las secuencias eran perfectamente legibles en cuatro de las cinco regiones amplificadas (Fig. 2 abajo). No se pudo solucionar el problema en el caso del fragmento *trnS-trnG*.

Diversidad genética

Cuando las cuatro secuencias filogeográficamente útiles (Intrón *L* y los espaciadores *trnT-trnL*, *trnH-psbA* y *trnL-rpl32*) se analizaron por separado, se detectaron 99 sitios polimórficos que incluyeron 95 indels y cuatro mutaciones puntuales. 19 indels y tres mutaciones puntuales fueron detectadas en la región *trnT-trnL*, una mutación puntual y tres indels en el Intrón *L*, 37 indels en el *trnH-psbA* y 36 indels en el *trnL-rpl32* (Tabla 1).

Las estimas de diversidad genética obtenidas con los diferentes fragmentos no codificantes del cpADN mostraron distintos resultados (Tabla 1). Por un lado, el fragmento *trnL-rpl32* mostró una variabilidad nula dentro de cada una de las poblaciones mientras que, en el otro extremo, el Intrón *L* presentó cierta diversidad genética en todas ellas salvo IK. Por otra parte, analizando los datos de la Tabla 1 se observa que el fragmento *trnT-trnL* generó unas estimas de diversidad muy diferentes en las poblaciones de la región Atlántica frente a las de la región Mediterránea. Únicamente las poblaciones de la región Mediterránea presentaron diversidad intrapoblacional mientras que las de la región Atlántica mostraron una variabilidad nula. Los otros tres marcadores fueron menos

informativos respecto a la diversidad genética de cada una de las poblaciones. Un ejemplo extremo es *trnL-rpl32* que presentó una variabilidad nula en todas y cada una de las poblaciones. En todo caso, todos los marcadores fueron informativos respecto a la estructura de población a lo largo del rango de distribución de la especie, como pone de manifiesto que las estimas de diversidad totales siempre tienen valores distintos de cero. Por ejemplo, el marcador *trnL-rpl32*, a pesar de mostrar nula variabilidad intrapoblacional, separó las poblaciones IK y EP del resto (algo fácilmente observable con el propio alineamiento).

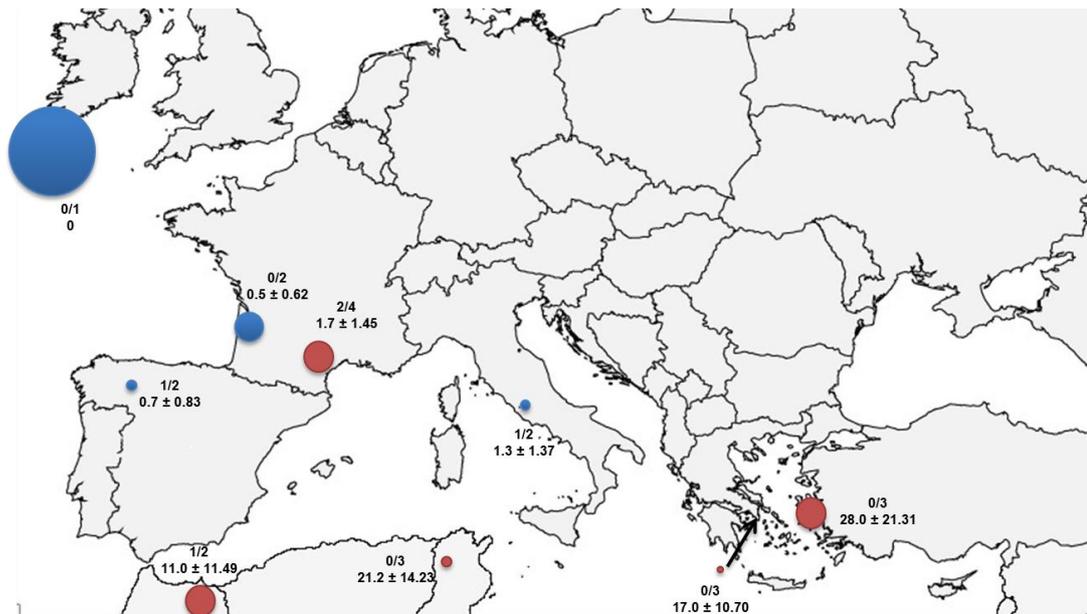


Fig. 3. Diversidad haplotípica en *Arbutus unedo*. El tamaño de los círculos es proporcional a la desviación con respecto a la media para todas las poblaciones. Los círculos azules representan diversidades por debajo de la media y los rojos por encima. Los números al lado de cada círculo indican el número de haplotipos privados respecto al número total de haplotipos así como la estima de la diversidad nucleotídica (en nº de diferencias) con su desviación estándar.

La estima de diversidad haplotípica (H_d) total obtenida concatenando las secuencias de los cuatro fragmentos fue 0.91 ± 0.02 mientras que la diversidad nucleotídica (π , en nº de diferencias) fue 38.3 ± 18.99 (Tabla 1). Todas las poblaciones pertenecientes a la región Atlántica presentaron una diversidad genética menor que las de la región Mediterránea, posiblemente debido a la influencia de los valores obtenidos con el fragmento *trnT-trnL* (Fig. 3). Las estimas de diversidad haplotípica más bajas se registraron en IK y FB (0 y $0.5 \pm$

0.265 respectivamente), ambas pertenecientes al región Atlántica. La otra población Atlántica, EP, mostró una diversidad haplotípica igual a la de IR (0.67 ± 0.314), pero en IR la diversidad nucleotídica fue mayor.

Cuatro de las diez poblaciones estudiadas (EP, MD, FM e IR) mostraron uno o varios haplotipos privados (haplotipos únicos dentro de esas poblaciones) sin patrón geográfico aparente. Las estimas de diversidad nucleotídica revelaron que las poblaciones del Norte de África y de la Península Balcánica-Asia tienen variabilidades 10-20 veces más altas que el resto de sitios estudiados. Estos valores más altos indicaron que sólo en estas poblaciones conviven haplotipos separados por largos huecos de indels.

Relaciones filogenéticas y distribución geográfica de los haplotipos

Los cuatro marcadores filogeográficamente informativos se concatenaron en una única secuencia de 2448-2487 pb (dependiendo del número de indels presentes en cada individuo). Las secuencias del Intrón *L* presentaron una longitud de 561 o 563 pb, las del *trnT-trnL* de 843, 852 u 862 pb, las del *trnH-psbA* de 309 o 346 pb y las del *trnL-rpl32* de 690 o 726 pb (Tabla 1). Todas juntas, las cuatro regiones no codificantes definieron trece haplotipo, siendo los más extendidos el Hap_3, observado en cuatro de las diez poblaciones estudiadas (FM, TK, GA y TI) y el Hap_4, encontrado en 3 (FM, TK e IR). El resto de haplotipos estaban presentes en una o dos poblaciones (Fig. 4).

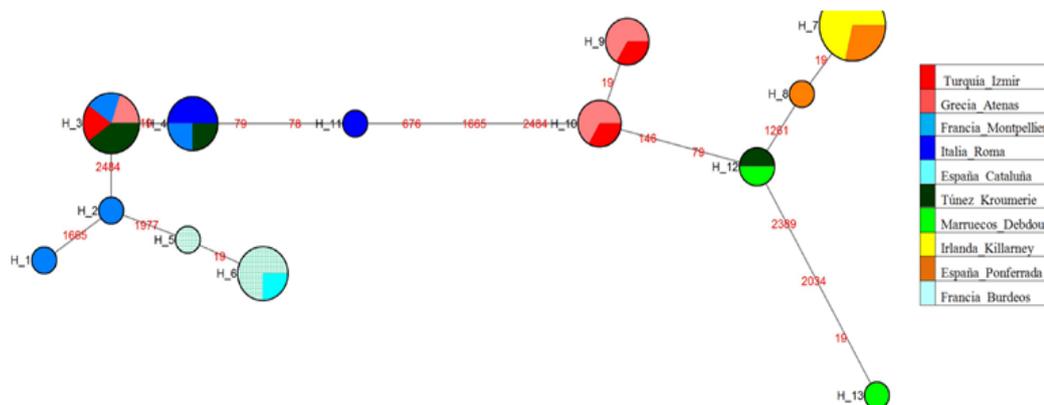


Fig. 4. Red haplotípica construida con el algoritmo MJ seguido de la opción de máxima parsimonia. Cada círculo identifica un haplotipo combinado de regiones no codificantes del cpADN. El tamaño del círculo es proporcional al número de individuos y los sectores coloreados indican las poblaciones de origen de esos individuos.

La red haplotípica construida con el algoritmo MJ presentó una disposición esencialmente lineal. A lo largo de esta disposición lineal es posible distinguir tres grupos de haplotipos filogenéticamente próximos (Fig. 4). Esos tres grupos mostraron una disposición espacial que en gran medida coincide con las tres grandes regiones geográficas distinguibles a lo largo del rango de distribución de *Arbutus unedo* (Fig. 5): región Atlántica-norte de África (IK, EP, MD y TK), región Mediterránea Occidental (FB, FM, EC e IR) y región Mediterránea Oriental (GA y TI). Los haplotipos de la región Mediterránea Oriental (Hap_9 y Hap_10) ocuparon una posición más central dentro de la red que sugiere que serían haplotipos más ancestrales. Del mismo modo, los haplotipos de Francia-Cataluña (Hap_1, Hap_2, Hap_5 y Hap_6), norte de África (Hap_12 y Hap_13), y los del noroeste de la Península Ibérica e Irlanda (Hap_7 y Hap_8), aparecieron en los extremos de la red, lo que sugeriría un origen más reciente. Además, la red haplotípica mostró una gran separación entre la población FB (Atlántico francés) y la de las otras dos de la región Atlántica (IK y EP).

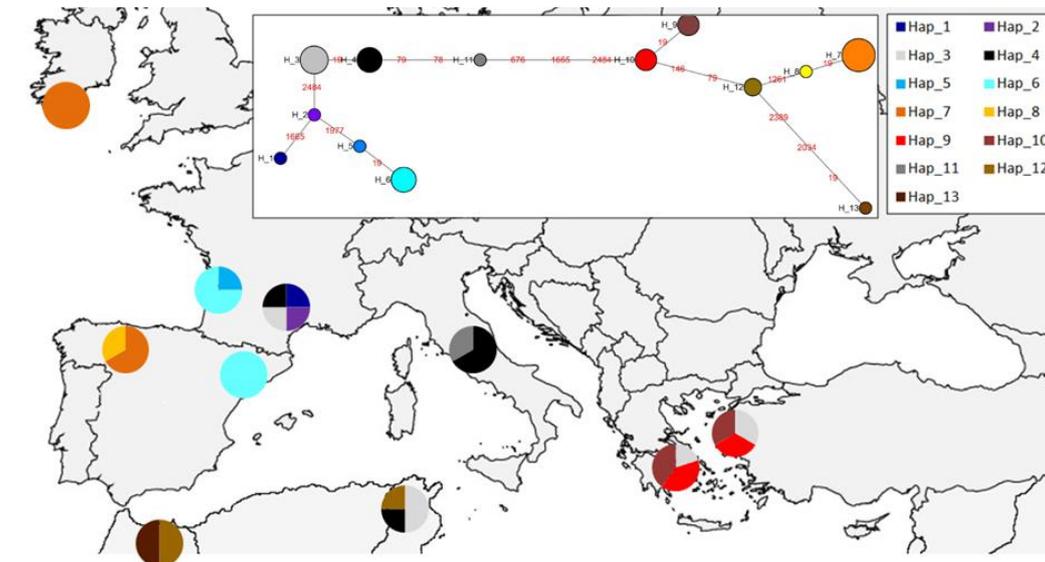


Fig. 5. Distribución de los haplotipos del cpADN de *Arbutus unedo*. Cada color identifica un haplotipo combinado de los fragmentos Intron *L*, *trnH-psbA*, *trnT-trnL*, *trnL-rpl32*. La figura insertada reproduce la red haplotípica para mostrar las relaciones filogenéticas entre los haplotipos.

DISCUSIÓN

Puesta a punto marcadores polimórficos del cpADN para estudiar la filogeografía de *Arbutus unedo*

Tras probar cinco regiones no codificantes del cpDNA utilizadas con frecuencia en estudios filogeográficos de otras plantas, hemos comprobado que al menos cuatro de ellas pueden ser informativas para investigar *Arbutus unedo*. En el caso del único fragmento con el que no hemos tenido éxito (*trnS-trnG*), los problemas son exclusivamente de secuenciación. A pesar de que conseguimos productos de PCR también para este fragmento, las secuencias obtenidas fueron ilegibles. De hecho, uno de los mayores impedimentos para este estudio fueron los problemas de legibilidad de las secuencias. Las primeras reacciones de PCR se llevaron a cabo utilizando una *Taq* polimerasa, el enzima universalmente más utilizado. Sin embargo, aunque obtuvimos producto de PCR para todos los fragmentos, sus secuencias eran ilegibles en todos los casos. La solución para estos problemas de secuenciación fue la utilización de una polimerasa de nueva generación para generar los productos de PCR. Estas polimerasas de nueva generación tienen niveles de actividad 2 o más veces mayores que las *Taq* polimerasa convencionales así como unas tasas de error más bajas. Del mismo modo, Fazekas et al. (2010) demostraron que los problemas de secuenciación debidos a largos fragmentos mononucleótidos sólo se solucionan con este tipo de enzimas, mientras que otro tipo de soluciones (disminuir la temperatura de extensión, incluir coadyuvantes en la PCR, o cambiar el número de ciclos) son poco eficaces.

Variación haplotípica y filogeografía en *Arbutus unedo*

Los resultados obtenidos estudiando los cuatro marcadores del cpADN de *Arbutus unedo* muestran ya indicios interesantes sobre la filogeografía de este árbol mediterráneo. Uno de los aspectos más interesantes del patrón de distribución del madroño es la presencia de poblaciones en Irlanda pero no en Inglaterra y otras localidades intermedias del Atlántico. Este tipo de distribución disjunta se conoce como distribución Lusitánica (especies presentes en la Península Ibérica e Irlanda pero no en localidades intermedias) y se observa también en otras especies animales (*Geomalacus maculosus*, *Semilimax*

pyrenaicus, *Elona quimperiana*) y vegetales (*Calluna* spp., *Daboecia cantabrica*) (Beatty & Provan 2013). El origen de esta distribución no está totalmente resuelto.

Nuestros resultados muestran que la población de FB no presenta ningún haplotipo coincidente con EP ni con IK, por lo que se podría rechazar la idea de que *Arbutus unedo* siguiera la ruta migratoria que la mayoría de los árboles que se encuentran hoy en día establecidos en Irlanda han podido seguir después del LGM (último máximo glacial) (Mitchell 2006). Estos árboles habrían seguido una ruta a lo largo de la costa Atlántica pasando luego desde la costa Atlántica francesa directamente hacia Irlanda sin pasar por Gran Bretaña. Frente a esta hipótesis, analizando los datos obtenidos en este trabajo podemos desechar que esto ocurriera con el *Arbutus unedo* puesto que no existe ninguna relación genética entre la población de la costa Atlántica de Francia con la población de Irlanda. De hecho, los haplotipos del Atlántico francés se encuentran en el otro extremo de la red haplotípica respecto a los haplotipos irlandeses. Según el estudio de análisis de polen llevado a cabo por Mitchell (1993) y las dataciones de restos de madera realizadas por van Rijn (2004), se ha concluido que *Arbutus unedo* lleva asentado en Irlanda desde al menos hace 4000 años. La cuestión que se plantea ahora es ¿cómo pudo llegar el *Arbutus unedo* a Irlanda, si no estaba allí antes del LGM y no siguió la ruta de migración seguida por la mayoría de las especies de árboles que se encuentran ahora en Irlanda? Nuestros resultados sugieren que las poblaciones irlandesas derivan de migrantes procedentes del Noroeste de la Península Ibérica. Si *Arbutus unedo* realmente hubiese llegado a Irlanda hace 4000 años, sus semillas pudieron ser transportadas por el ser humano a finales del Mesolítico pues podrían utilizar sus frutos como fuente de alimento. La hipótesis alternativa sería que *Arbutus unedo* estuviese presente en Irlanda antes del LGM. En este escenario alternativo, el sur de Irlanda habría servido de refugio, aunque pasando por un cuello de botella, al verse sus poblaciones afectadas por las bajas temperaturas y disminuir en densidad. Esto explicaría la nula diversidad genética observada en las muestras de Irlanda. Sin embargo, esta hipótesis es menos probable ya que estudios recientes han corroborado que toda la isla de Irlanda se encontraba cubierta por una gran placa de hielo durante el LGM (Coxon 2005). Además, se sabe que las semillas de *Arbutus unedo* no son capaces de germinar por debajo de 4°C (Narbona et al. 2003). Aún así, también se ha comprobado que

el madroño es capaz de soportar temperaturas de hasta -12°C durante el invierno (Mereti et al. 2001) y se sabe que es una especie que está bien adaptada a condiciones ambientales duras (Özcan et al. 2007).



Fig. 6. Posibles rutas de dispersión de *Arbutus unedo*.

Además, la red haplotípica y la disposición geográfica de los haplotipos sugiere que *Arbutus unedo* pudo haber seguido al menos dos rutas migratorias diferentes a lo largo de su historia (Fig. 6). Posiblemente la ruta que más clara puede estar fue la que va desde la Península Balcánica-Asia hacia Península Ibérica-Irlanda pasando por el Norte de África. Por otra parte, otra ruta se concentraría en la región Mediterránea. Aunque sus detalles son más confusos se puede especular que pudo haber una radiación desde Italia hacia el Este (Península Balcánica), Sur (Norte de África) y Oeste (Francia), llegando hasta la costa nororiental de la Península Ibérica. En general, las poblaciones de la región Mediterránea (MD, TI, FM, EC, IR, TK y GA) presentan mayor diversidad genética que las de la región Atlántica (IK, FB y EP). Esto puede indicar que *Arbutus unedo* lleva más tiempo asentado en estas regiones, lo que sería coherente con el hecho de que el madroño es una especie que se desarrolla mejor en los lugares secos y cálidos durante el verano.

Finalmente, nuestra red haplotípica indica que en *Arbutus unedo* se podrían distinguir hasta tres ESUs. Uno estaría formado por las poblaciones Atlánticas

Noroeste de la Península Ibérica e Irlanda, otro por las poblaciones francesas y catalanas circumpirenaicas y, por último, una ESU geográficamente mucho más amplia que abarcaría el resto del rango de distribución de *Arbutus unedo*. En el caso de que estas ESUs sufriesen un fuerte cambio en las fuerzas evolutivas, tal vez podría llegar a formar una subespecie. Estas ESUs deberían considerarse unidades independientes desde el punto de vista de la gestión y conservación de la especie.

Tabla 1. Diversidad genética de *Arbutus unedo*. Estimaciones de diversidad para cuatro fragmentos del genoma del cloroplasto y para su secuencia combinada. S/Indel, número de sustituciones / número de sitios afectados por inserciones/delecciones; H, número de haplotipos; H_d , diversidad haplotípica; π , diversidad nucleotídica (como n° de diferencias); pb, longitud de la secuencia; N, número de secuencias individuales analizadas dentro de cada una de las poblaciones. Los asteriscos (*) indican presencia de uno o varios haplotipos privados en esa población. Códigos de las poblaciones como en Fig. 1.

Marcador	Atlántico			Mediterráneo Occidental					Mediterráneo Oriental		Total	
	IK	FB	EP	MD	FM	EC	IR	TK	GA	TI		
<i>trnT-trnL</i>	pb	852	862	852	852/843	852	852/862	852	852	852	852	843/862
	S/Indel	0/0	0/0	0/0	1/9	2/0	1/10	1/0	2/0	2/0	2/0	3/19
	N	5	4	4	3	4	3	4	5	5	3	40
	H	1	1	1	2	2	2	1	2	2	2	5
	H_d	0	0	0	0,67 ± 0,314	0,83 ± 0,222	0,67 ± 0,314	0,50 ± 0,265	0,40 ± 0,237	0,40 ± 0,237	0,67 ± 0,314	0,69 ± 0,047
	π	0	0	0	6,7 ± 5,40	1,2 ± 1,11	7,3 ± 5,90	0,5 ± 0,62	0,8 ± 0,80	0,8 ± 0,8	1,3 ± 1,4	4,2 ± 2,36
Intrón L	pb	563	563	563	563	563	563	561/563	563	561/563	561/563	561/563
	S/Indel	0/0	1/2	1/0	1/0	1/0	1/0	0/2	1/2	1/2	1/2	1/3
	N	5	5	5	4	4	4	4	5	5	5	46
	H	1	3	2	2	2	2	2	3	3	3	6
	H_d	0	0,70 ± 0,218	0,60 ± 0,175	0,67 ± 0,204	0,50 ± 0,265	0,50 ± 0,265	0,50 ± 0,265	0,80 ± 1,640	0,80 ± 0,164	0,80 ± 0,164	0,82 ± 0,022
	π	0	1,2 ± 1,06	0,6 ± 0,65	0,7 ± 0,75	0,5 ± 0,62	0,5 ± 0,62	1,0 ± 0,99	1,4 ± 1,19	1,4 ± 1,19	1,8 ± 1,44	1,8 ± 1,16
<i>trnH-psbA</i>	pb	346	309/346	346	346	309	309	309	309/346	309/346	309/346	309/346
	S/Indel	0/0	0/37	0/0	0/0	0/0	0/0	0/0	0/37	0/37	0/37	0/37
	N	5	5	4	4	4	2	3	5	5	4	41
	H	1	2	1	1	1	1	1	2	2	2	2
	H_d	0	0,40 ± 0,237	0	0	0	0	0	0,50 ± 0,265	0,40 ± 0,237	0,60 ± 0,175	0,51 ± 0,022
	π	0	14,8 ± 9,37	0	0	0	0	0	18,5 ± 12,49	14,8 ± 9,37	22,2 ± 13,85	18,8 ± 9,43
<i>trnL-rpl32</i>	pb	690	726	690	726	726	726	726	726	726	726	690/726
	S/Indel	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/36
	N	5	5	5	5	5	4	4	5	5	5	48
	H	1	1	1	1	1	1	1	1	1	1	2
	H_d	0	0	0	0	0	0	0	0	0	0	0,34 ± 0,069
	π	0	0	0	0	0	0	0	0	0	0	12,1 ± 6,19
Combinado	pb	2451	2460	2451	2478/2487	2450	2460	2448/2450	2450/2487	2450/2485	2450/2485	2450/2487
	S/Indel	0/0	1/0	1/0	2/9	3/0	-	0/2	3/39	3/39	3/39	4/95
	N	5	4	3	2	4	1	3	4	5	3	34
	H	1	2	2*	2*	4*	-	2*	3	3	3	13
	H_d	0	0,50 ± 0,265	0,67 ± 0,314	1,00 ± 0,500	1,00 ± 0,177	-	0,67 ± 0,314	0,83 ± 0,222	0,80 ± 0,164	1,00 ± 0,272	0,91 ± 0,024
	π	0	0,5 ± 0,62	0,7 ± 0,83	11,0 ± 11,49	1,7 ± 1,45	-	1,3 ± 1,37	21,2 ± 14,23	17,0 ± 10,70	28,0 ± 21,31	38,3 ± 18,99

REFERENCIAS

- Andrew R.L., Bernatchez L., Bonin A., Buerkle C.A., Carstens B.C., Emerson B.C., Garant D., Giraud T., Kane N.C., Rogers S.M., Slate J., Smith H., Sork V.L., Stone G.N., Vines T.H., Waits L., Widmer A. & Rieseberg L.H. (2013) A road map for molecular ecology. *Molecular Ecology* **22**, 2605-2626
- Avise, J.C. (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA
- Avise, J.C. (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography* **36**, 3-15
- Beatty G. E., and J. Provan. 2013. Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the Lusitanian plant species *Daboecia cantabrica* (Ericaceae). *Journal of Biogeography* **40**, 335–344
- Excoffier L. & Lischer H.E.L. (2010) Arlequin suite v.3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567
- Falchi A., Paolini J., Desjobert J-M, Melis A., Costa J. & Varesi L. (2009) Phylogeography of *Cistus creticus* L. on Corsica and Sardinia inferred by the TRNL-F and RPL32-TRNL sequences of cpDNA. *Molecular Phylogenetics and Evolution* **52**, 538-543
- Fazekas A.J., Steeves R. & Newmaster S.G. (2010) Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques* **48**, 277-285
- García X.R. (2008) *Guía das Plantas de Galicia*. Edicións Xerais de Galicia, S.A.
- Librado P. & Rozas J. (2009) DnaSP v.5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452
- Li G-D, Yue L-L, Sun H. & Qian Z-G (2012) Phylogeography of *Cyananthus delavayi* (Campanulaceae) in Hengduan Mountains inferred from variation in nuclear and chloroplast DNA sequences. *Journal of Systematics and Evolution* **50**, 305-315

- López L. & Barreiro R. (2013) Patterns of chloroplast DNA polymorphism in the endangered polyploid *Centaurea borjae* (Asteraceae): implications for preserving genetic diversity. *Journal of Systematics and Evolution* **9999**, 1-10
- McCauley D.E. (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Tree* **10**
- Mereti M., Grigoriadou K. & Nanos G.D (2002) Micropropagation of the strawberry tree, *Arbutus unedo* L. *Scientia Horticulturae* **93**, 143-148
- Mitchell F.J.G. (2006) Where did Ireland's trees come from? *Biology and environment: proceedings of the Royal Irish Academy* **106**, 251-259
- Mitchell F.J.G. (1993) The biogeographical implications of the distribution and history of the strawberry tree, *Arbutus unedo* in Ireland. En: M.J. Costello & K. S. Kelly (eds) *Biogeography of Ireland: Past, Present and Future*, vol. **2**, 35-44. Irish Biogeographical Society, Dublin
- Molina M., Pardo-de-Santayana M., Aceituno L., Morales R. & Tardío J. (2011) Fruit production of strawberry tree (*Arbutus unedo* L.) in two Spanish forests. *Forestry* **84**
- Moritz C. (1994) Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology* **3**, 401-411
- Narbona E., Arista M. & Ortiz P.L. (2003) Germinación de las semillas del madroño (*Arbutus unedo* L., Ericaceae) *Acta Botanica Malacitana* **28**, 73-78
- Newton, A.C., Allnut, T.R., Gillies, A.C.M., Lowe, A.J. & Ennos, R.A. (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Elsevier Science* **14**, 140-145
- Nikolaeva, M.G. (1977) Factors controlling the seed dormancy pattern. En Khan, A.A. (ed) *The Physiology and Biochemistry of Seed Dormancy and Germination*, 51-74. North-Holland, Amsterdam.
- Özcan M.M. & Haciseferoğulları H. (2007) The Strawberry (*Arbutus unedo* L.) fruits: Chemical composition, physical properties and mineral contents. *Journal of Food Engineering* **78**, 1022-1028

- Palmer, J. (1985) Chloroplast DNA and Molecular Phylogeny. *BioEssays* **2**, 263-267
- Polzin T. & Daneschmand S.V. (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* **31**, 12-20
- Ricoi H. N. & Silvar C. (2001) *Guía das Árbores de Galicia*. Baía Edicións, Galigraf Galicia, S.L.
- Shaw J., Lickey E.B., Beck J.T., Farmer S.B., Liu W., Miller J., Siripun K.C., Winder C.T., Schilling E.E., Small R.L. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**, 142-166
- Sipos B., Massingham T., Stütz A.M. & Goldman N. (2012) An Improved Protocol for Sequencing of Repetitive Genomic Regions and Structural Variations Using Mutagenesis and Next Generation Sequencing. *PLOS ONE* **7**
- Taberlet P., Gielly L., Pautou G. & Bouvet J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**, 1105-1109
- Thomas D.C., Hughers M., Phutthai T., Rajbhandary S., Rubite R., Ardi W.H. & Richardson J.E. (2011) A non-coding plastid DNA phylogeny of Asian *Begonia* (Begoniaceae): Evidence for morphological homoplasy and sectional polyphyly. *Molecular Phylogenetics and Evolution* **60**, 428-444
- Torres, J.A., Valle F., Pinto C., García-Fuentes A., Salazar C. & Cano E. (2002) *Arbutus unedo* L. communities in southern Iberian Peninsula mountains. *Plant Ecology* **160**, 207-223
- van Rijn P. (2004) The analysis of charcoal from Ross Island. En: W. O'Brien (ed). *Ross Island Mining, Metal and Society in Early Ireland*, 386-401 Bronze Age Studies n°6 Glaway. Department of Archaeology, National University of Ireland, Galway

Villar L. (1993) *Arbutus* L. En: Castroviejo S., Aedo C., Gómez C., Láinz M., Montserrat P., Morales R., Muñoz F., Nieto G., Rico E., Talavera S., Villar L. (ed). *Flora Ibérica*, vol. **IV**, 514-516 CSIC, Madrid