Consejo Superior de Investigaciones Científicas
Real Jardín Botánico de Madrid

Universidad Autónoma de Madrid
Facultad de Ciencias

Filogeografía, evolución de nicho y especiación en un linaje mediterráneo–macaronésico: *Lavatera maritima–Lavatera acerifolia* (Malvaceae)

Irene Villa Machío

Memoria de Tesis Doctoral
Madrid, 2017

Consejo Superior de Investigaciones Científicas
Real Jardín Botánico de Madrid

Universidad Autónoma de Madrid
Facultad de Ciencias

Filogeografía, evolución de nicho y especiación en un linaje mediterráneo–macaronésico: *Lavatera maritima–Lavatera acerifolia* (Malvaceae)

Memoria para optar al grado de Doctor en Biología que presenta

Irene Villa Machío

Directores

Dr. Francisco Javier Fuertes Aguilar
Dr. Gonzalo Nieto Feliner

Tutor académico

Dr. Roberto Gamarra Gamarra

Madrid, 2017

Imagen de la portada: *Lavatera maritima*. J. V. Sandoval-Sierra

Diseño de Portada: L. Martín-Torrijos

A mi familia,
malagueña y madrileña

AGRADECIMIENTOS

Porque una tesis no la hace una persona sola, ¿cómo simplificar en pocas palabras toda la ayuda que he recibido?

Me gustaría agradecer a mis directores de tesis. A Javier por darme la oportunidad de iniciarme en esta ardua carrera investigadora, y a Gonzalo por toda su dedicación, interés y esfuerzo demostrado para que esta tesis salga adelante.

A mi tutor académico, Roberto Gamarra, por la atención recibida y su eficacia en resolver las dudas burocráticas.

A los diferentes grupos de investigación que me han acogido en sus laboratorios. A Pilar Catalán y todo su grupo: Luis Ángel, Rubén, Ernesto y a Diana, gracias por esas tardes de aprendizaje. Gracias también a Viruel y a Esther por sus sabios consejos. Gracias por hacerme disfrutar de una estancia "tan provechosa" tanto profesional como personalmente. Y si pienso en Huesca, no puedo dejar de agradecer todo el cariño que recibí de mis padres adoptivos, gracias Simón y Tere! Quiero agradecer a Joan Vallès y Teresa Garnatje por compartir conmigo sus conocimientos sobre "tamaño genómico". También a Marcela Rosato, por su dedicación a las "lavateras" en el estudio citogenético, y a Juli Caujapé por facilitarnos la estancia en esa maravillosa isla.

A las personas que me han acompañado en los viajes de trabajo de campo: Gonzalo, Jano, Juankis y Fátima, así como a todos los que me habéis ayudado con la aportación de muestras: Javier Fuertes, Pedro Escobar, Juan Carlos Moreno, Saul Manzano, Alex Baumel, F. Médail, Stephan Scholz, Ricarda Riina, Inés Álvarez....

Al personal del RJB del Herbario, de la Biblioteca, de Administración, de Seguridad, de Limpieza, gracias por facilitar el trabajo día a día, ya sea proporcionando pliegos, facilitando el trámite de pedidos, ordenes de servicio....., por abrirme la puerta del botánico cada día (y por supuesto dejarme salir a altas horas de la noche), o por un simple saludo acompañado de una sonrisa.

Al personal técnico más veterano: Guillermo Sanjuanbenito, Fátima Durán, Alberto Herrero, Emilio Cano, Yolanda Ruiz, y a la nueva generación: Marta, Lucía, Olga y Jose, por esas interminables horas de trabajo codo a codo, por esos días de risa (y por los de bajón también). Gracias! A Yolanda Ruiz por facilitarme el uso del servicio de Microscopía electrónica.

A todos mis compañeros y amigos de las diferentes etapas vividas durante la tesis: Mario Mairal, Andrea S. Meseguer, Marieta, Alicia, Melisa, José Luis, Isa, Mario, Pedro, Ana, Yurena, Arantxa, Jano, Juan Carlos, Vladi, Laura, Kina, Javi, Javi Fajardo, Iván, Vicky, Bea Rumeu, Angélica, Isabel Sanmartín, Riki, Tamara... por generar tan buen ambiente de trabajo, además de aportar buenas ideas y consejos siempre constructivos. A los estudiantes que nos han visitado durante estancias más o menos largas: Vicky, Jose, Marisa, Carla, Donis, Pamela, Farideh, Yeter, porque han contribuido positivamente en mi proceso de formación, y por los buenos ratos vividos. Y a los pequeños que han ido creciendo durante este tiempo: Simón, Rubén y Noa, Jairo, Emma, Raúl y Alejandra.

De manera muy especial, a los amigos que habéis estado muy cerca, siempre apoyando con vuestras ideas, conocimientos, logística, ingeniando planes y dándome mensajes de apoyo. A la familia de Delicias, al hogar de Esperanza! GRACIAS!

A los amigos que ya estaban cuando empecé este camino doctoral, y que pese a la distancia y el aislamiento que conlleva la tesis, siguen ahí esperando noticias mías. Agradecer a mi tocaya Irene, la precursora de esta carrera.

A mi familia, gracias por apoyarme en todo momento, gracias por estar siempre dándome las fuerzas necesarias para enfrentar esta etapa. A mi princess (para cuando empiece a leer), por hacerme disfrutar como una enana de las cosas pequeñas.

Gracias a todos los que os habéis involucrado, porque una tesis no se hace sola....

ÍNDICE

INTRODUCCIÓN

Durante siglos, biólogos y naturalistas han observado que el número de especies, tanto en ambientes marinos como terrestres, no se encuentra distribuido de forma aleatoria, ni a escala global ni a escala local, es decir, que existe una tendencia de los grupos de seres vivos a concentrarse en ciertas áreas geográficas o regiones. La ocupación de un área geográfica particular por un taxón es una consecuencia tanto de factores históricos como ecológicos (Lomolino *et al.*, 2006). La distribución geográfica de los organismos es objeto de estudio de la biogeografía, que al ser parte de la biología evolutiva trata de averiguar cómo han evolucionado los organismos en el espacio a través del tiempo hasta llegar a su distribución actual. Dentro de la biogeografía, tradicionalmente ha habido dos grandes corrientes, la histórica que intenta explicar cómo los eventos geológicos y climáticos pasados han dado forma a la distribución de los organismos (Crisci, Katinas & Posadas, 2003), y otra ecológica que se ha centrado más en los factores medioambientales que explican la distribución espacial en el momento actual.

El concepto de biodiversidad fue acuñado por Wilson (1988) como como la "variedad de la vida a lo largo de todos los niveles de organización, desde la diversidad génica en poblaciones, pasando por la diversidad de especies, hasta la diversidad de ecosistemas". Esta biodiversidad es el resultado de cientos de millones de años de evolución influidos por procesos naturales y, en tiempos recientes, antropogénicos. La diversidad génica en poblaciones está relacionada con los cambios que se producen en el genoma de cada individuo, debidos a mutaciones o a recombinación genética, entre otros factores influyentes. Los biólogos moleculares evolutivos han hecho uso de la acumulación de estos cambios a lo largo del tiempo para reconstruir la historia evolutiva de las especies, infiriéndola a partir de los patrones de diversidad genética que presentan las diferentes poblaciones dentro una especie.

Biogeografía de la región Mediterránea

La región mediterránea es especialmente atractiva para estudiar el efecto que los procesos geológicos y climáticos tienen sobre el origen y el patrón de distribución de las especies vegetales debido a los conocimientos que existen sobre su compleja historia geológica y paleoclimática (Dercourt *et al.*, 2000; Krijgsman, 2002), su diversidad biótica bien documentada (Blondel, 2010) y su gran variabilidad de tipos de hábitat (Blondel

& Aronson, 1999). La región mediterránea, incluyendo la región macaronésica (Quézel, 1985), se considera uno de los 25 *hotspots* de biodiversidad del mundo y una encrucijada biogeográfica excepcional para las regiones eurosiberiana, irano-turanias y saharo-síndica (Manafzadeh, Staedler & Conti, 2016; Médail & Quezel, 1997; Myers *et al.*, 2000; Quézel, 1985). El paisaje en mosaico, la presencia de islas, el que haya funcionado como refugio de diversidad que se perdió en otras latitudes en el Pleistoceno y la gran diversidad de condiciones climáticas hace que la cuenca mediterránea acumule un elevado número de especies vegetales endémicas, albergando unas 22.500 plantas vasculares, de las cuales 13.000 son endémicas (Thompson *et al.*, 2005). Además, la flora mediterránea cuenta con elementos florísticos de diferente origen geográfico, y alberga especies relictas de origen subtropical que existían previamente al establecimiento del clima mediterráneo actual (Blondel, 2010; Thompson *et al.*, 2005).

La alta diversidad de especies ha sido tradicionalmente explicada, en parte, por los diferentes procesos paleo-climáticos y geológicos que han tenido lugar, en particular tres episodios. El primero es la crisis de la salinidad del Messiniense, a finales del Mioceno (5.96–5.33 millones de años AP; Bocquet, Widler & Kiefer, 1978; Krijgsman *et al.*, 1999) en la que la evaporación en el Mediterráneo superó la precipitación recogida por los ríos que drenaban en él y se redujo el intercambio de agua con el Atlántico hasta llegar a desconectar completamente, lo que causó una rápida caída del nivel del mar Mediterráneo y la deposición de grandes cantidades de sal en el fondo marino. El segundo es el establecimiento del clima Mediterráneo (3.2 millones de años AP; Suc, 1984), donde se consolida la estacionalidad con veranos secos y calurosos e inviernos húmedos y lluviosos. El tercero son los episodios de glaciación ocurridos durante el Cuaternario (Hewitt, 2000). Los cambios climáticos del Cuaternario, que comienzan a inicios del Pleistoceno (2.6 millones de años AP), han modelado los patrones geográficos de diversidad de muchas especies de forma diferente (Bennett & Provan, 2008; Stewart *et al.*, 2010). En Europa, durante los períodos de glaciación, el descenso del nivel del mar permitió la extensión de la superficie actual de islas que facilitó la conexión entre áreas de distribución previamente más alejadas. En el otro extremo, en las zonas montañosas, sobre todo de las masas continentales, la distribución geográfica de muchas especies

se limitó a áreas aisladas que actuaron como refugio, principalmente en las penínsulas mediterráneas del sur (Comes & Kadereit, 1998; Taberlet *et al.*, 1998). Estos refugios representan zonas climáticamente más estables que las circundantes, contienen mayor cantidad de especies y diversidad genética que aquellas colonizadas tras los períodos de glaciación debido a un efecto acumulativo, y se consideran reservorios de biodiversidad (Hewitt, 1996; Kadereit *et al.*, 2005; Médail & Diadema, 2009; Nieto Feliner, 2011). Durante los períodos interglaciares se produjo una re-expansión, y las áreas habitables del norte de Europa fueron recolonizadas a partir de de los refugios, pero principalmente solo por las poblaciones más próximas, lo que explica la menor diversidad genética de las poblaciones septentrionales hoy en día, de acuerdo con el modelo de recolonización llamado "leading edge" (Nichols & Hewitt, 1994). Estos ciclos de contracción-expansión se repitieron, al menos, cuatro veces (Weising & Freitag, 2007). Los ciclos glaciares tuvieron un notable efecto tanto en la distribución y evolución de las especies de plantas como en los tamaños de sus poblaciones. Así pues, en los períodos más fríos, donde las especies menos tolerantes a las bajas temperaturas sufrieron la contracción de su rango geográfico, la formación de refugios, la especiación alopátrica y los eventos de vicarianza pudieron verse favorecidos. Otros efectos, como el flujo genético entre poblaciones previamente aisladas en refugios separados, se produjeron durante los períodos interglaciares, no solo en zonas de refugio sino también en zonas de tránsito de expansión y retracción (Petit *et al.*, 2003; Widmer & Lexer, 2001). La cuenca mediterránea alberga grupos de plantas con diferentes patrones filogeográficos más o menos comunes, por ejemplo los que representan diferentes niveles de diversidad genética a lo largo de un eje este-oeste o un gradiente decreciente sur-norte debido a las oscilaciones climáticas y a los procesos de recolonización ocurridos durante el Pleistoceno (Nieto Feliner, 2014). Pero la influencia de factores específicos del linaje en cuestión, históricos o biológicos, han evitado en muchos casos el que puedan detectarse patrones filogeográficos comunes. Otro factor que ha modificado la flora mediterránea es la persistente influencia del ser humano, que se ha convertido en la principal causa de amenaza y extinciones de especies. Como consecuencia, poblaciones individuales de determinadas especies han sufrido repetidos aislamientos, y poblaciones de diferente origen se han ido mezclando

a lo largo de la historia de la cuenca (Besnard *et al.*, 2013; Mateu-Andrés *et al.*, 2015). Así mismo, la actividad humana también ha provocado colonizaciones muy rápidas de especies útiles (Fineschi *et al.*, 2000; Vendramin *et al.*, 2008).

Las penínsulas Ibérica, Itálica y Balcánica son los principales refugios que han permitido mantener la alta biodiversidad del Mediterráneo. Éstos se diferencian en sus orígenes, edades, conexiones y componentes. La Península Itálica representa una región relativamente joven, mientras que la importante orografía de la Península Ibérica favoreció la existencia de múltiples refugios (modelo "refugios dentro de refugio", Gómez & Lunt, 2007). La Península Balcánica es considerada como el mayor *hotspot* de Europa, con gran cantidad de especies endémicas, gracias a la encrucijada geográfica que representa y la mayor estabilidad climática desde finales del Terciario con una incidencia menor de las glaciaciones que la Península Ibérica (Hewitt, 2011). Sin embargo, en los últimos años comienza a cobrar fuerza la idea de que otras áreas han podido jugar también un papel importante como refugios, tal es el caso del Norte de África (Husemann *et al.*, 2014; Veríssimo *et al.*, 2016). Esta región alberga numerosos *hotspots* de biodiversidad de plantas y endemismos (Médail & Quezel, 1997), sin embargo no es una región tan estudiada como Europa. Debido a su historia climática y a su notable topografía montañosa, los estudios filogeográficos de especies presentes en el norte de África detectan una divergencia evolutiva y una riqueza alélica mayor en las poblaciones de esta región en comparación con las poblaciones europeas (Hampe *et al.*, 2003). Por otro lado, el Norte de África es un área en la que el avance del calentamiento global, en particular el significativo aumento de la aridez y el avance del desierto del Sahara en los últimos 10.000 años (Houghton *et al.*, 2001), está dejando marcados efectos negativos (Jiménez-López *et al.*, 2016).

Biogeografía de Islas

Debido al dinamismo evolutivo, por un lado, y a la precisa circunscripción de las áreas de distribución impuesta por los océanos, por el otro, las islas se han considerado laboratorios naturales que ayudan a comprender la evolución. Dentro de esta idea, se ha prestado mucha atención a la biogeografía, tratando de interpretar la diversidad de especies en islas en función de procesos biogeográficos. El modelo del equilibrio dinámico

de islas de (MacArthur & Wilson, 1963, 1967) postula que el número de especies de un taxón dado encontrado en una isla se encuentra en equilibrio dinámico entre la tasa de extinción y la tasa de migración de dicha isla. Este número en equilibrio se encuentra determinado por dos factores: la distancia al continente y el tamaño de la isla, elementos que llevan a la ganancia y a la pérdida de especies dando lugar a una renovación continua a lo largo del tiempo. Las islas de menor dimensión están expuestas a una mayor tasa de extinción, y consecuentemente, a una menor diversidad de especies, mientras que el nivel de inmigración se espera que sea mayor en las islas más cercanas al continente (Fig. 1).

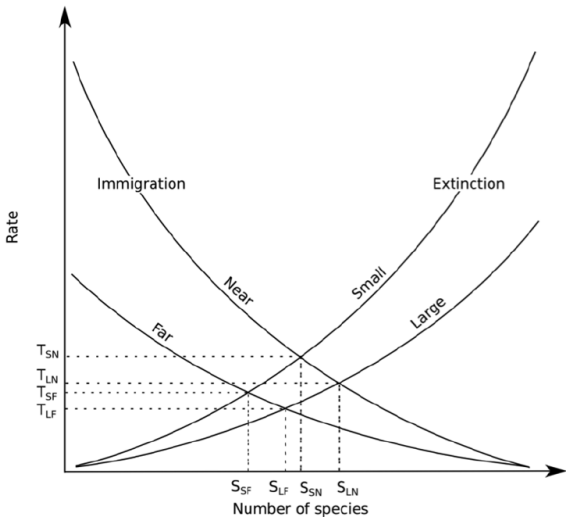


Figura 1. Modelo del equilibrio dinámico en la biogeografía de islas. Figura extraída de Wilson & McArthur (1963)

La teoría del equilibrio dinámico se convirtió en el paradigma de la biogeografía de islas durante años y ha tenido una gran influencia en la biología de la conservación. Sin embargo, Lomolino (2000), más de 30 años después, reflexionó acerca de esta teoría y de su adecuación a los avances conseguidos en el conocimiento de la complejidad de la naturaleza. A día de hoy, se consideran una mayor variedad de factores explicativos y escalas temporales para explicar la diversidad actual, al tiempo que se buscan herramientas analíticas más versátiles (Sanmartín, Van Der Mark & Ronquist, 2008). Según Lomolino (2000), la teoría de equilibrio dinámica es demasiado sencilla, fue desarrollada para explicar la variación de especies en un archipiélago durante un tiempo ecológico, sin tener en cuenta suficientemente la influencia de la evolución.. Además, Wilson y MacArthur sólo tuvieron en cuenta factores físicos de las islas (distancia y tamaño)

como factores influyentes en la colonización y extinción de las especies, asumiendo que todas las especies presentan la misma capacidad para establecerse en nuevos ambientes. Lomolino (2000) realizó algunas modificaciones y presentó alternativas a la teoría del equilibrio para conseguir mejorar la comprensión de los factores que estructuran las comunidades insulares. Entre otros, propone que esta teoría debe ser simple y estar basada en tres principales procesos biogeográficos: inmigración, extinción y evolución (Fig.2).

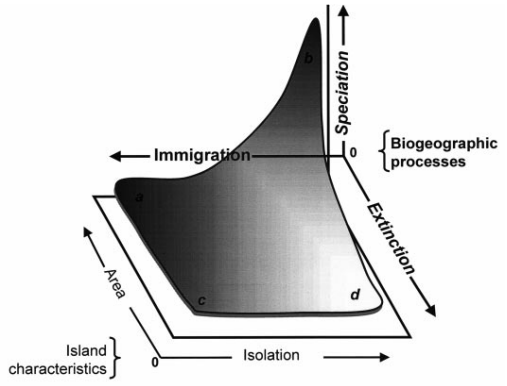


Figura 2. Modelo tripartito de biogeografía de islas de Lomolino (2000) ilustrando la variación en los 3 procesos biogeográficos fundamentales (inmigración, extinción y evolución) como una función de las características de las islas en cuestión.

Región Macaronésica - Islas Canarias

La región mediterránea, incluyendo la región macaronésica (Mittermeier *et al.*, 1998), representa un escenario idóneo para un estudio comparativo de biología evolutiva entre islas y continente, ya que además de ser un área singular con elevada diversidad de hábitats y de especies, comprende islas continentales (Córcega, Islas Baleares, Malta, etc.), que son masas de tierra de corteza continental separadas de éste por una franja de mar; e islas oceánicas (región macaronésica: Madeira, Islas Canarias o Azores entre otras), que nunca han tenido contacto con el continente. Las islas representan, debido a su separación física del continente, sistemas ideales para estudiar la influencia que tienen las barreras en el patrón de colonización y especiación de los seres vivos.

El concepto de Macaronesia fue introducido por el botánico P. Barker-Webb en el siglo XIX para designar la región biogeográfica natural constituida por cinco archipiélagos atlánticos: Azores, Madeira, Canarias, Cabo Verde y Salvajes, clasificación que se sigue manteniendo hoy día (Fernández-Palacios *et al.*, 2011). El grado de aislamiento de cada archipiélago, medido como distancia al continente, varía entre los 96 Km que dista

Fuerteventura (Islas Canarias) de la costa del Sahara y los 1450 Km que dista Azores de la costa portuguesa. El origen volcánico de todos los archipiélagos conlleva la existencia de estructuras geológicas y paisajísticas similares entre ellos. El archipiélago actual más antiguo es Salvajes, que alcanza los 27 millones años (Geldmacher *et al.*, 2001).

Las Islas Canarias, además de ser las más próximas al continente, son las islas con mayor superficie y altitud de la región (7480 km2 y 3718 m del Teide, respectivamente; Fernández-Palacios & Andersson, 2000). Este archipiélago está formado por siete islas (El Hierro, La Gomera, La Palma, Tenerife, Gran Canaria, Lanzarote y Fuerteventura) y varios islotes como La Graciosa, Montaña Clara, Alegranza y el islote de Lobos. Debido a su gran extensión, las islas Canarias constituyen el archipiélago con mayor riqueza de especies (Fernández-Palacios & Andersson, 2000), con 1300 especies de plantas vasculares, de las cuales el 44.3% son endémicas (Reyes-Betancort *et al.*, 2008; Whittaker & Fernández-Palacios, 2007).

En la historia geológica de las islas volcánicas se pueden reconocer varias fases: nacimiento, crecimiento hasta alcanzar su máxima área y elevación, seguido de un proceso de erosión caracterizado por distintos eventos catastróficos (vulcanismo, megadeslizamientos, etc), y un posterior hundimiento por debajo del nivel del mar. La edad que alcanzan las diferentes islas que conforman el archipiélago canario nos permite observar esas fases de desarrollo en diferentes islas (Fig. 3). En estado primario, de nacimiento, se encuentra Las Hijas, una colina submarina que aún no ha emergido. En el estado de crecimiento o desarrollo se encuentran La Palma, El Hierro y Tenerife, islas en construcción como prueban su considerable actividad volcánica, que

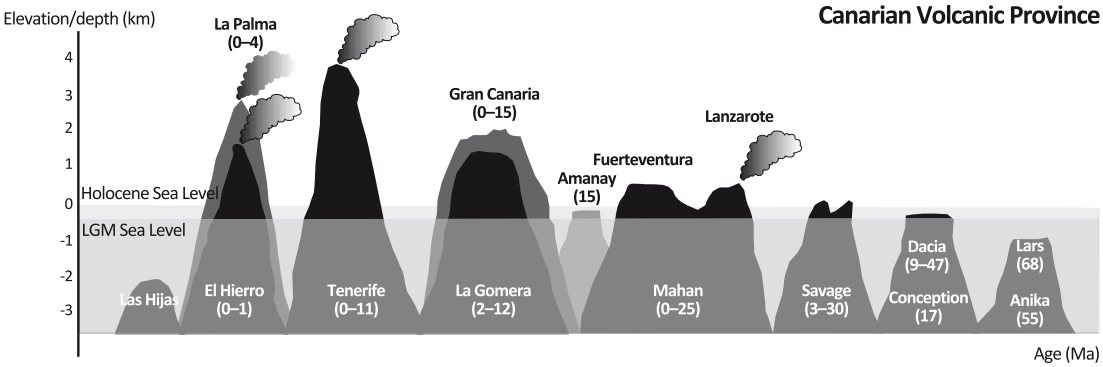


Figura 3. Diferentes estados del ciclo de las islas volcánicas representado en las Islas Canarias. Figura adaptada de Fernández-Palacios *et al*. (2011).

ha provocado las mayores altitudes del archipiélago (3000-4000m). En un período de erosión y desmantelamiento se encuentra Gran Canaria y La Gomera, islas en las que los procesos de destrucción son mayores que los procesos de construcción y que, en un pasado, alcanzaban mayor área y altitud que en la actualidad. En un estado "plano basal" se encuentra Mahan, el edificio volcánico que alberga tanto a Fuerteventura como a Lanzarote. Estas islas pudieron alcanzar una altitud de hasta 3300m, sin embargo hoy día no superan los 800m (en el caso de Fuerteventura). La fase de desaparición terminal se puede observar en una isla del archipiélago Salvajes (Salvaje pequeña), que pertenece al mismo hotspot volcánico que las Islas Canarias. Y por último, la fase "guyot" (montes submarinos con la cima plana) que aparece en Amanay, Concepción y Dacia, colinas sumergidas que pertenecen al archipiélago canario (Fernández-Palacios *et al.*, 2011). Las colinas sumergidas más antiguas de la región macaronésica --Ormonde en la provincia volcánica de Madeira y Lars en Canarias-- datan de hace unos 60 millones de años, en el Paleoceno (Geldmacher *et al.*, 2001), por lo que estas colinas pudieron llegar a emerger con las variaciones del nivel eustático del nivel del mar durante la época glacial del Pleistoceno. Estos cambios de nivel del mar, con la consecuente afloración de islas, facilitaron una vía de colonización llamada en *stepping stone* entre las islas actuales y la Península Ibérica y el Norte de África. Diversos estudios apoyan que estas islas actualmente sumergidas tuvieron un papel importante en la formación de la actual biota macaronésica (Carine *et al.*, 2004; Fernández-Palacios *et al.*, 2011).

Sin embargo, hay casos en que la dinámica geológica es algo más compleja, tal es el caso de Tenerife. Se considera que esta isla emergió inicialmente como dos o tres islas separadas hace unos 11 millones de años. Estas paleo-islas se fusionaron en los últimos 3,5 millones de años, por el afloramiento de material volcánico en la zona intermedia actual de la isla, dando lugar a lo que conocemos hoy como la isla de Tenerife. Por otro lado, Guillou *et al.* (2004) en su estudio con datos magnetostratigráficos y de K-Ar sugieren que la capa central de Tenerife data del Mioceno, a partir de la cual se desarrollaron Teno (Mioceno tardío) y Anaga (Plioceno), datos que serían consistentes con la actividad volcánica este-oeste que presenta la cadena de las Islas Canarias.

Dos especies hermanas del género *Lavatera*

Esta tesis está enfocada a investigar la historia evolutiva de dos especies de plantas del género *Lavatera*: *Lavatera maritima* y *Lavatera acerifolia*. Filogenéticamente, son especies hermanas dentro de la alianza *Malva*, en el grupo *Lavatera-Malva* (familia Malvaceae, Escobar et al., 2009, Fig. 4). Las Malváceas constituyen una de las familias de Angiospermas más importantes en zonas neo-tropicales, tanto por el número de especies (120 géneros y1500 especies, Bayer & Kubitzki, 2003) como por su abundancia en los ecosistemas. El grupo *Lavatera-Malva* constituye un grupo filogenético muy diversificado formado por plantas herbáceas perennes de distribución mediterránea y del suroeste de Asia, con el principal centro de diversidad en el oeste de la cuenca mediterránea y en el Oriente Medio (Bates, 1968; Escobar *et al.*, 2009; Tate *et al.*, 2005). En este grupo de especies la poliploidía es un elemento importante ya que contiene desde genomas hexaploides, como es el caso de *L. maritima* y *L. acerifolia*, hasta 16-ploides (Devesa Alcaraz & Luque, 1986; Escobar *et al.*, 2009).

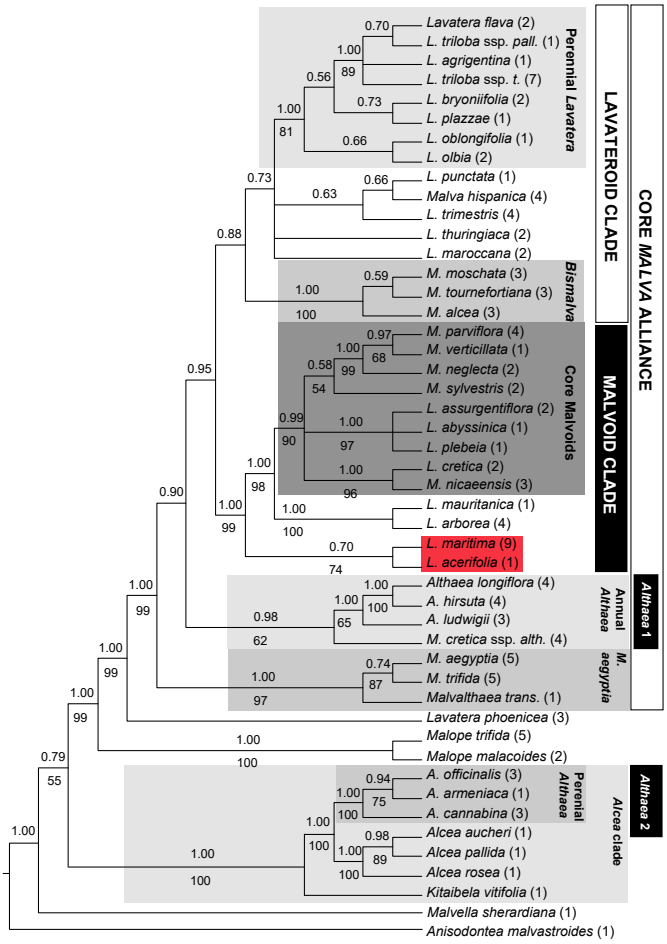


Figura 4. Árbol consenso (*majority rule consensus tree 50%*) basado en secuencias ITS obtenido en el análisis bayesiano (adaptado de Escobar *et al.*, 2009)

A continuación se presentan las descripciones de las especies:

- *Lavatera maritima* Gouan. (Malvavisco marino)

Arbusto de 0,3-1,5 m de altura, muy ramificado, cubierto en todas sus partes por tomento denso y bajo. Hojas de unos 6 cm de longitud, suborbiculares, poco profundamente lobadas (3-5), con lóbulos de redondeados a subagudos con el margen de crenado a subentero; pecíolo 0,6-3 cm. Flores generalmente solitarias, a veces axilares. Corola con pétalos de 1,5-3 cm de longitud, rosa pálidos, con la uña purpúrea. Esquizocarpo con 10-13 mericarpos. Polen esferoidal, grande, polipantoporado con exina equinada. 2n= 44 (hexaploide). Matorrales en medios alterados sobre crestones calizos más o menos verticales, y más raramente sobre margas y esquistos, en lugares muy secos del litoral, pudiendo aparecer en zonas que distan hasta 300 Km del litoral. 0-900 m.s.n.m. Se encuentra distribuida por el oeste de la región mediterránea (Península Ibérica, Francia, Italia, Córcega, Cerdeña, Túnez, Argelia y Marruecos), y se considera una especie rara en el oeste de Portugal (Castroviejo et al., 1993) (Fig. 5).

- *Lavatera acerifolia* Cav. (Malva de risco)

Arbusto alto de hasta 2-5 m. Hojas glabras palmatilobadas de 5-7 cm de longitud; lóbulos muy marcados, irregularmente dentados; peciolos de 4-9 cm de longitud. Flores en pequeños racimos terminales o axilares, a veces, solitarias, con pedúnculos largos. Corola con los pétalos de 4-5 cm de longitud, de color malva con bases oscuras o, en muy pocas ocasiones, blancos. Polen esferoidal, grande, polipantoporado con exina equinada. 2n= 44 (hexaploide). En acantilados o en pies de paredes rocosas de Lanzarote, Fuerteventura, Gran Canaria, Tenerife, La Palma y La Gomera (Bramwell & Bramwell, 1990). Aparece en el piso de vegetación de los bosques termófilos de Canarias (150-500m de altitud), un ecosistema joven caracterizado por inviernos fríos y húmedos frente a veranos cálidos y secos que apareció al comienzo del Cuaternario, hace unos 2,5 millones de años (Fernández-Palacios *et al.*) (Fig. 6).

Lavatera maritima y *L. acerifolia* exhiben adaptaciones claras al hábitat en el que viven y específicamente al clima, el tipo de sustrato y la incidencia de luz solar (Fig. 5, 6). Se trata de plantas leñosas perennes, cuyo ciclo de vida dura más de un año, y xerófitas, adaptadas a vivir en ambientes secos, por lo que poseen raíces muy desarrolladas para

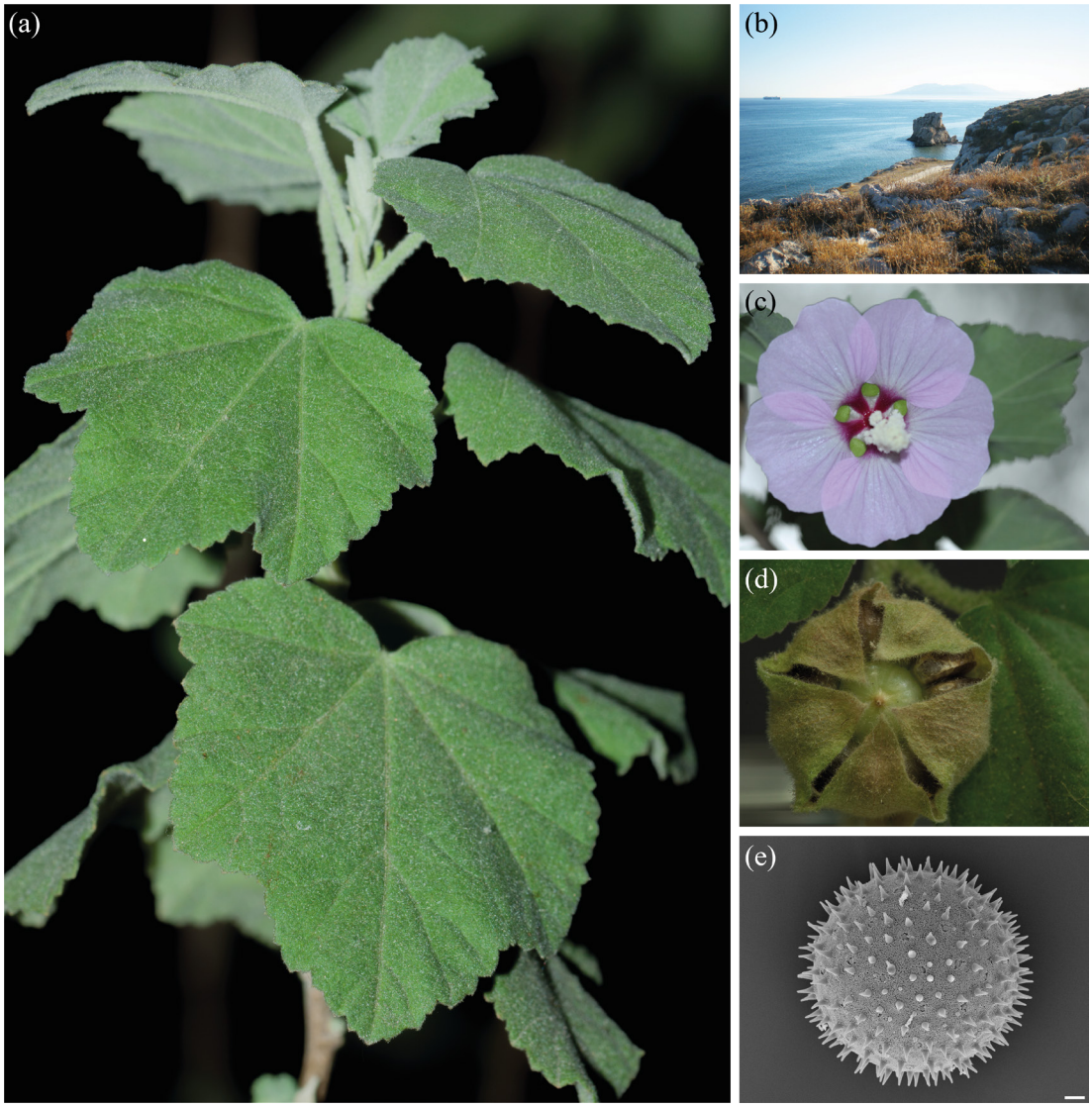


Figura 5. Fotografías de *Lavatera maritima*. (a) Hojas, (b) hábitat, (c) flor, (d) fruto e (e) imagen del polen obtenida con el microscopio electrónico de barrido. Barra de escala = 10 µm.

llegar a mayor profundidad. En cuanto al sustrato, *Lavatera maritima* es una especie calcícola, que viven en suelos ricos en carbonato cálcico (con pH>7, alcalino). Sin embargo, algunas de sus poblaciones también aparecen sobre suelos volcánicos, como *L. acerifolia*. Este tipo de suelo se considera muy rico en nutrientes, y de pH variable según su composición (Dahlgren, Saigusa & Ugolini, 2004). Tanto *Lavatera maritima* como *L. acerifolia* son especies halo-nitrófilas, es decir, habitan suelos con alto contenido en nutrientes como nitratos, potasio o fósforo, y con cierta concentración de sales, de ahí que vivan en suelos alterados o degradados por el hombre y en acantilados cerca de la línea de costa (Castroviejo *et al.*, 1993). En cuanto a la insolación, *L. maritima* está recubierta

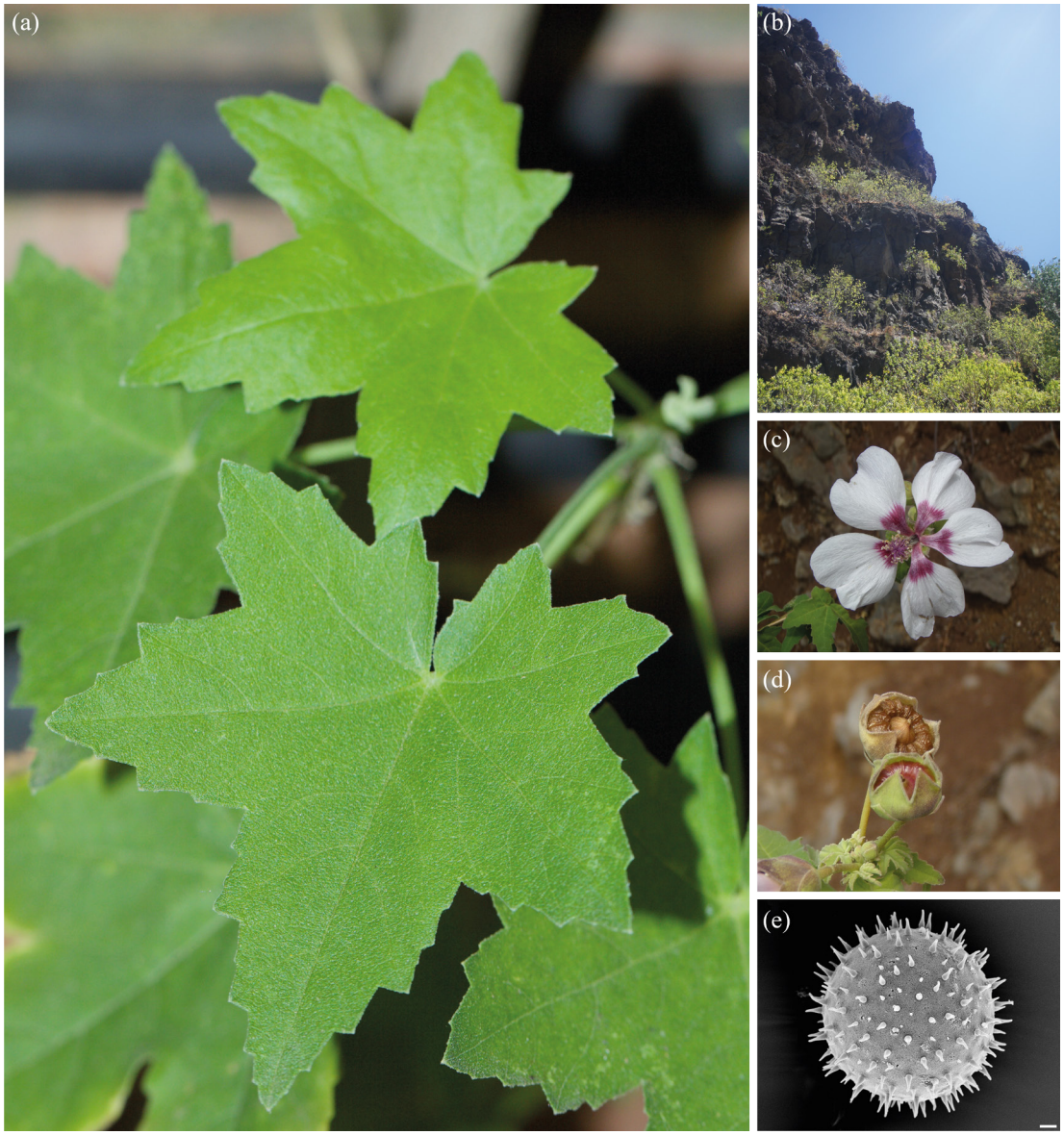


Figura 6. Fotografías de *Lavatera acerifolia*. (a) Hojas, (b) hábitat, (c) flor, (d) fruto e (e) imagen del polen obtenida con el microscopio electrónico de barrido. Barra de escala = 10 µm.

de un indumento de color plateado que es efectivo al reflejar los rayos ultravioleta y proteger de la pérdida de agua por transpiración (Fig. 5). Esta característica es muy común en especies adaptadas al clima mediterráneo.

Proceso de divergencia y especiación

El hecho de que *L. acerifolia* y *L. marítima* son especies hermanas que presentan una distribución geográfica disyunta, la primera endémica de un archipiélago oceánico y la segunda distribuida en el continente e islas continentales, constituye un caso de estudio favorable y de considerable interés para intentar esclarecer los factores que han podido

influir en el proceso de divergencia y especiación de ambas especies.

La especiación es un proceso evolutivo mediante el cual las poblaciones de una especie quedan aisladas y evolucionan hasta dar lugar a uno o varios linajes diferentes. El modo de aislamiento de esas poblaciones puede ser debido a barreras geográficas o genéticas. En plantas, como en otros organismos, se dan varios modelos de especiación: simpátrida, parapátrida y alopátrida. La especiación simpátrica ocurre cuando una especie da lugar a dos nuevas especies debido a un aislamiento por barreras genéticas; las poblaciones divergen hasta conseguir una independencia evolutiva dentro del mismo rango geográfico. La especiación parapátrica ocurre cuando existe aislamiento reproductivo en poblaciones que tienen una distribución adyacente en el espacio, pero sigue existiendo un modesto flujo génico entre ellas. La especiación geográfica o alopátrica ocurre cuando una especie da lugar a dos nuevas especies debido al aislamiento geográfico de algunas de sus poblaciones, dado que es probable que al estar sometidas a ambiente diferentes, evolucionen de forma distinta. Para que se generen especies nuevas tiene que darse aislamiento reproductivo entre las poblaciones de la especie ancestral. Dependiendo del tamaño de las poblaciones que quedan aisladas, dentro de la especiación alopátrida, se puede diferenciar entre especiación peripátrica, cuando el tamaño de las poblaciones que quedan aisladas es muy desigual (principalmente se debe a poblaciones que sufren efecto fundador), y vicarianza, cuando el tamaño de las poblaciones que se aíslan es similar (Coyne & Orr, 2004).

APROXIMACIONES METODOLÓGICAS APLICADAS

La relación filogenética de *L. maritima* y *L. acerifolia*, junto a las características ecológicas de su nicho, nos lleva a explorar la distribución geográfica y genética que presentan estas especies, así como los factores que dieron lugar al proceso de divergencia evolutiva y especiación entre ellas.

Estudio filogeográfico

La filogeografía es una rama de la biogeografía que estudia los principios y procesos que gobiernan la distribución geográfica de las genealogías génicas de una especie, o de especies estrechamente relacionadas (Avise, 2000). La filogeografía integra la

filogenética y la genética de poblaciones para estudiar con detalle la historia evolutiva de una especie en el contexto de la historia geoclimática del planeta. Se basa en el análisis de la variabilidad molecular de los organismos en un contexto geográfico e histórico e integra conceptos y técnicas de genética molecular, genética de poblaciones, demografía, sistemática filogenética, etología y paleontología (Avise, 2000), haciendo uso de marcadores moleculares. Existen numerosos tipos de marcadores moleculares aplicables a estudios filogeográficos: secuencias de ADN citoplasmático (mitocondrial y plastidial) y nuclear, microsatétiles, AFLPs, SNPs, entre otros. La elección de los marcadores moleculares para un estudio depende de los objetivos y del nivel taxonómico de estudio (género, especie...) y de la naturaleza del organismo que se esté estudiando, por ejemplo si es diploide o poliploide.

Marcadores moleculares: secuencias de ADN (secuenciación de Sanger)

Cuando nació la filogeografía, se defendía el uso de secuencias de ADN citoplasmático, no recombinante y de herencia uniparental como marcadores moleculares ideales para encontrar variabilidad genética en los individuos que conforman una población y en definitiva estudiar la distribución espacial de las genealogías génicas. En el caso de las plantas, se ha usado mayoritariamente el ADN plastidial, que se hereda normalmente por vía materna (Corriveau & Coleman, 1988). La mayoría de los individuos posee una única secuencia de ADN plastidial, que constituye su haplotipo, y varios individuos, dentro de una especie, pueden mostrar diferencias detectables en estas secuencias. La naturaleza no recombinante y la haploidía de la molécula hacen que sea adecuada para estimar la historia evolutiva en poblaciones de plantas (Avise, 2009) aun cuando su tasa de mutación sea menor que la de las regiones mitocondriales propuestas como idóneas por Avise. Son numerosos los estudios filogenéticos de plantas basados en secuencias moleculares donde se utilizan exclusivamente genes y regiones no codificantes, tales como espaciadores, del genoma plastidial (Catalán, Kellogg & Olmstead, 1997; Clegg, 1993; Chen *et al.*, 2014; Kim *et al.*, 2016; Soltis & Soltis, 1998). Sin embargo, su herencia uniparental hace que las genealogías de regiones plastidiales tengan muchas limitaciones para poder reflejar la filogenia de las especies. Además de ser filogenias génicas y no de especies solo reflejan la historia genealógica de uno de sus progenitores. Para compensar esta desventaja se

produjo un incremento del uso de secuencias de marcadores nucleares, sobre todo los espaciadores transcrito interno (ITS) del ADN ribosómico nuclear (Álvarez & Wendel, 2003; Rieseberg & Soltis, 1991; Rieseberg & Wendel, 1993; Warwick *et al.*, 2010). Algunos de los motivos que justifican el uso generalizado de marcadores nucleares son, además de su herencia biparental, la disponibilidad de varios sets de cebadores universales para PCR útiles en un gran número de grupos taxonómicos (Gardes & Bruns, 1993; White *et al.*, 1990), su estructura multicopia que facilita la amplificación, y el hecho de que, debido a su nivel de variabilidad, frecuentemente son marcadores moleculares bastantes informativos para estudios evolutivos a nivel de especie (Feliner & Rosselló, 2007). Más del 90% de los estudios realizados en plantas han hecho uso de marcadores plastidiales y ribosomales (Claudel *et al.*, 2017; Li *et al.*, 2008; Tomasello *et al.*, 2015).

Marcadores moleculares: Single nucleotide polymorphism (SNPs)

Desde los años 80, la secuenciación de Sanger (secuenciación por dideoxinucleótidos basado en la replicación de ADN tras amplificar los fragmentos de ADN mediante la reacción en cadena de la polimerasa o PCR) ha sido la técnica estrella para los estudios genéticos, pero actualmente hay una gran demanda de métodos conocidos como "herramientas de secuenciación masiva" o "Next Generation Sequencing" (NGS), capaces de producir millones de secuencias de ADN reduciendo el costo y el tiempo de generación. Con estas nuevas tecnologías se consigue explorar y responder preguntas biológicas usando información representativa del genoma completo, por lo que están teniendo un gran impacto en la investigación. Las plataformas o secuenciadores de nueva generación están ayudando a revolucionar o abrir nuevas áreas de investigación biológica, como la exploración de genomas antiguos (Mardis, 2008), así como a entender mejor la fisiología de especies de cultivo. Algunos ejemplos son los análisis eQTL (expression quantitative trait loci) de genes relacionados con la domesticación y la tolerancia a la sequía en el arroz (Degenkolbe *et al.*, 2009; Li, Zhou & Sang, 2006), o con la resistencia a enfermedades en el maíz (Chung *et al.*, 2010).

El desarrollo de NGS, junto con el desarrollo de análisis basados en métodos de coalescencia a nivel de especies y de poblaciones, están revolucionando la filogeogeografía (Carstens, Lemmon & Lemmon, 2012; Carstens *et al.*, 2013; McCormack *et al.*, 2013;

Papadopoulou & Knowles, 2016). Durante los últimos veinte años, se han desarrollado varias tecnologías de marcadores moleculares que han sido aplicadas al análisis de genomas de plantas, principalmente para detectar diferencias entre individuos dentro de una misma especie. Los SNPs o polimorfismos de un único nucleótido son los principales marcadores usados hoy día en los análisis genéticos de plantas. Estos SNPs son marcadores directos que proporcionan información de la naturaleza exacta de cada una de las variantes alélicas (Chikara *et al.*, 2014). Los investigadores han reconocido durante años que los SNPs, a pesar de su bajo polimorfismo, son mucho más comunes en el genoma que los microsatélites (Brumfield *et al.*, 2003); sin embargo, resultaba muy difícil, o casi imposible, aprovechar la información contenida en estos marcadores cuando se estudia un bajo número de ellos. Con la llegada de la secuenciación masiva, no solo se producen SNPs en un número suficientemente alto para llevar a cabo estudios de filogeografía, sino que se pone de manifiesto la ventaja del uso de SNPs frente a la secuenciación del genoma completo de un organismo. La nueva tecnología NGS aporta grandes ventajas para las investigaciones sobre estructura poblacional, la delimitación de especies o la demografía histórica que normalmente se hacen con marcadores neutrales (Edwards, Shultz & Campbell-Staton, 2015).

La nueva era de la secuenciación: NGS y filogeografía

Existen diferentes técnicas para la generación de datos NGS aplicables a estudios filogenéticos y filogeográficos (Fig. 7). El más sencillo de ellos es la secuenciación de amplicones, que consiste en secuenciar, mediante plataformas de NGS, productos de PCR que han sido previamente generados por Sanger (McCormack *et al.*, 2013). Otro método es el denominado "target enrichment", también llamado "sequence capture", que implica una captura selectiva de regiones del genoma previas a la secuenciación NGS (Mamanova *et al.*, 2010). Esta técnica implica un conocimiento genómico previo para diseñar las sondas encargadas de marcar el ADN. Y, por último, los métodos de secuenciación de ADN "asociados a sitios de restricción" o RADseq, en el cual se secuencian fragmentos de ADN que previamente han sido digeridos por enzimas de restricción que cortan en regiones específicas del genoma. Se trata de técnicas que generan una representación reducida del genoma. RADseq comprende técnicas como RAD (restriction site associated

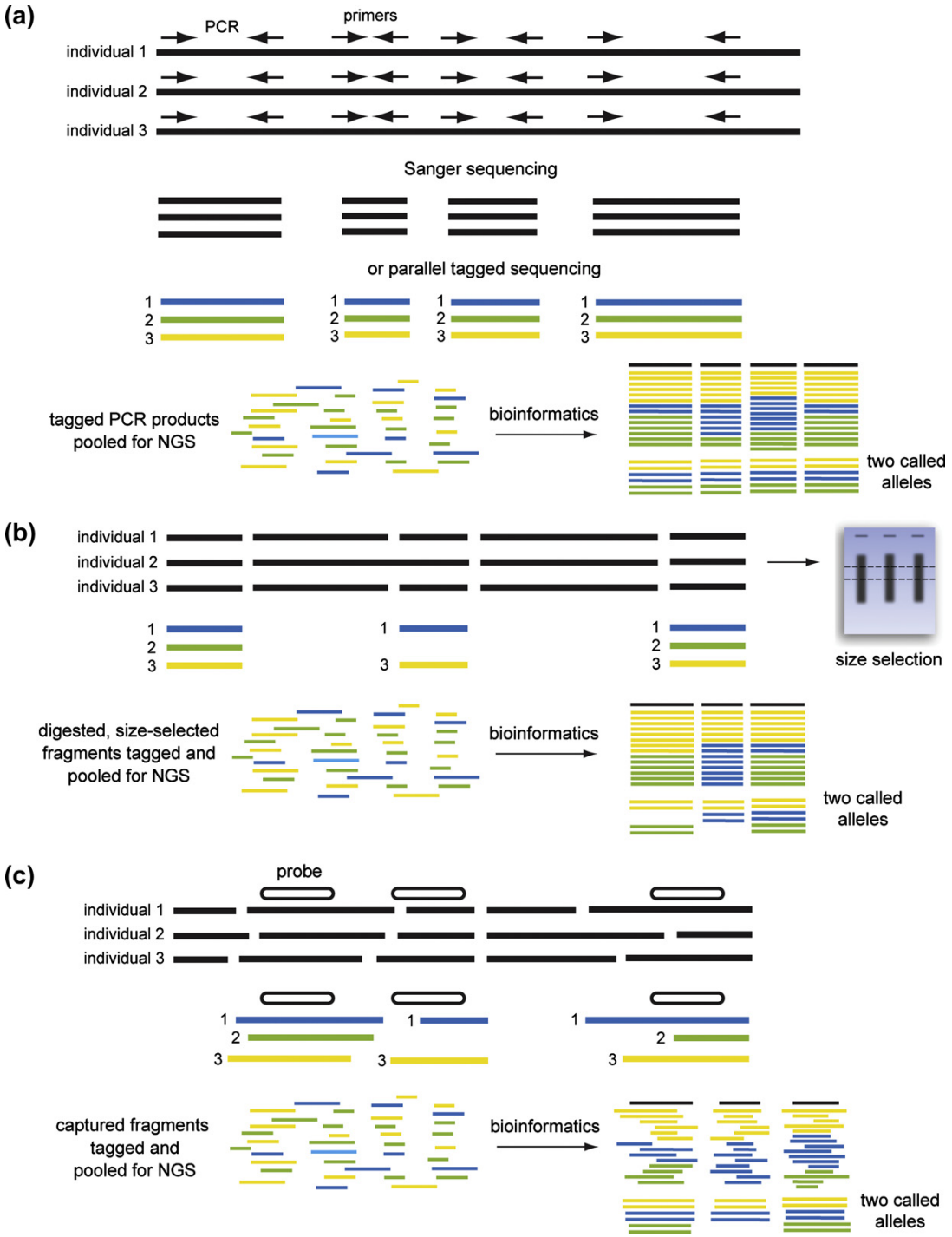


Figura 7. Métodos básicos de preparación de muestras para NGS. (a) Secuenciación de amplicones, (b) métodos basados en enzimas de restricción y (c) "target enrichment". (McComack *et al.*, 2013)

DNA), ddRAD (double digestión RAD) o GBS (Genotyping-by-sequencing) entre otros, que se diferencian principalmente en el número de enzimas empleadas para fragmentar el genoma y por la existencia de un paso adicional para la selección del tamaño de los fragmentos (Andrews *et al.*, 2016).

Genotyping-by-sequencing (GBS)

En esta memoria doctoral, la detección de SNPs se ha llevado a cabo aplicando la técnica GBS, un método que genera fragmentos de ADN de corta longitud usando una sola enzima de digestión (Fig. 8). Se trata de una técnica simple, con alta capacidad para procesar muestras independientes de forma conjunta (multiplexar) y adecuada para estudios poblacionales, también para caracterización de germoplasmas, estudios de sistemas de reproducción o "trait mapping" en diversos organismos (Elshire *et al.*, 2011). Esta técnica, de bajo coste, está basada en la secuenciación de nueva generación de fragmentos genómicos generados mediante enzimas de restricción. La tecnología NGS secuencia una representación del genoma completo de varios especímenes que permite detectar un elevado número de SNPs, y con ello estimar la diversidad genética existente dentro de especies y en comparación con especies próximas. Los fragmentos de ADN de diferentes muestras se marcan con unas secuencias de ADN cortas y únicas (barcodes), lo que permite reunir todas las muestras (pool) en un mismo canal de secuenciación.

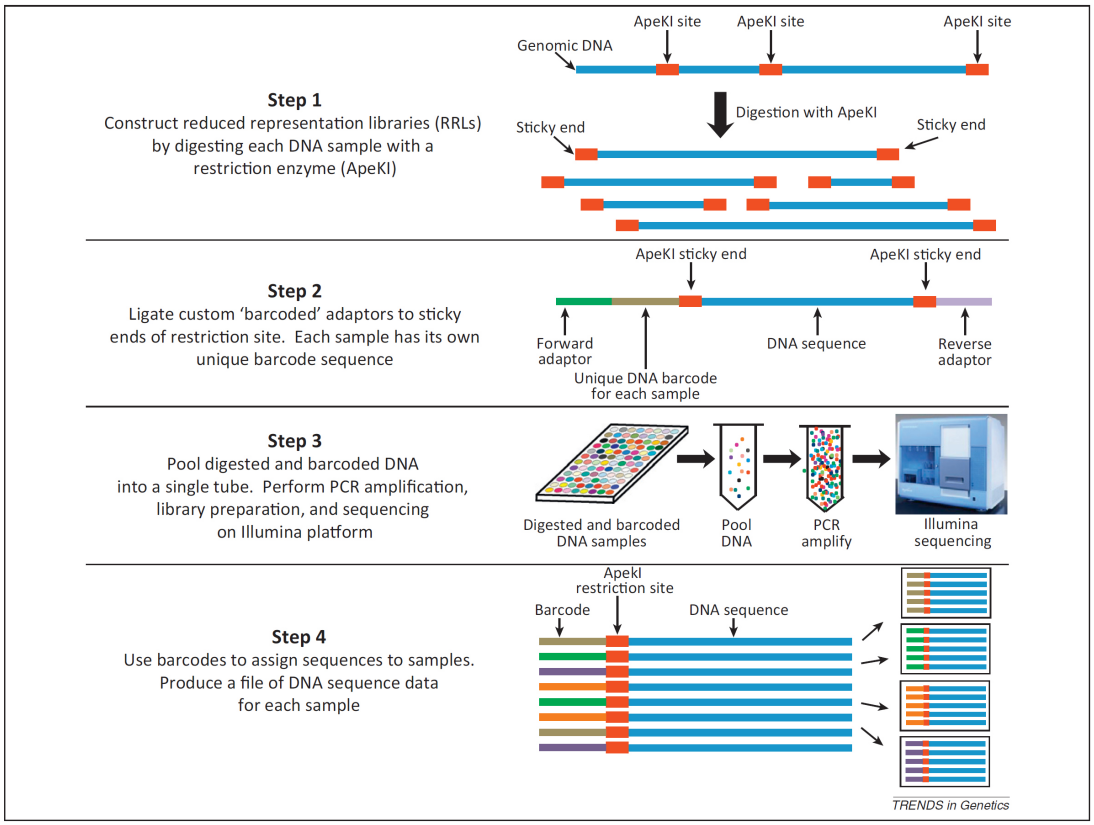


Figura 8. Genotyping-by-sequencing (GBS). (Myles, 2013)

Este método (DNA troceado de forma aleatoria por enzimas de restricción seguido de marcaje por barcodes) funciona muy bien para especies con genomas pequeños. Reducir la complejidad del genoma con enzimas de restricción es fácil, rápido, extremadamente específico, reproducible, y puede alcanzar regiones importantes del genoma que resultan inaccesibles con otros métodos de secuenciación (como "sequence capture"). Eligiendo la enzima de restricción adecuada, se pueden evitar las regiones repetitivas del genoma y las regiones de bajo número de copias pueden ser marcadas con mayor eficiencia, lo que simplifica tremendamente el posterior difícil alineamiento mediante procedimientos informáticos en especies con altos niveles de diversidad genética (Elshire *et al.*, 2011).

Baird *et al.* (2008) fue el primero que demostró las ventajas de secuenciar ADN genómico asociado a sitios de restricción para extraer SNP en una elevada cantidad y poder genotipar. Comparado con otros métodos RADseq, GBS es una técnica más sencilla en la que las muestras están sometidas a un menor procesado para obtener los fragmentos de ADN ligados a los adaptadores o barcodes. De este modo, se reduce la manipulación de las muestras en el laboratorio, evitando provocar posibles resultados erróneos. Entre los pasos reducidos en la técnica GBS que se aplican en otros métodos RADseq están la sonicación aleatoria de los fragmentos de ADN que previamente han sido digeridos por la/s enzima/s de restricción y la selección de tamaño de los fragmentos de ADN (Davey *et al.*, 2011).

Modelos de distribución de especies (SDM)

La componente ecológica del estudio se ha abordado con ayuda de los modelos de distribución de especies (SDM, por *species distribution modeling*) y midiendo el grado de solapamiento de los nichos bioclimáticos. Los SDM asocian los registros de presencia de una especie con los datos ambientales (por lo general climáticos) de esas localidades con el fin de estimar la distribución geográfica potencial y las condiciones ambientales idóneas para esa especie (Guisan & Zimmermann, 2000). Además, este conjunto de condiciones ambientales puede ser proyectado a épocas pasadas y futuras para las que existen datos, identificando áreas geográficas que muestran alta idoneidad para la presencia de dicha especie. Con estas proyecciones es posible inferir la distribución potencial de la especie en diferentes etapas temporales asumiendo que los requerimientos, el nicho en general, no haya cambiado (Nogués-Bravo, 2009).

En los últimos años ha cobrado una gran relevancia la pregunta de si los linajes tienden a conservar o a desplazar su nicho en el transcurso de la evolución, sobre todo a través de los episodios de especiación (Pearman *et al.*, 2008), al margen de que, en términos generales, las plantas podrían tener mayor dificultad para cambiar de nicho que los animales(Donoghue & Smith, 2004; Sanmartín & Ronquist, 2004). Hay un punto de vista común que considera que los nichos ecológicos tienden a estar conservados a lo largo de la evolución, como ocurre en los procesos de radiación y en la especiación alopátrica (Losos, 2008; Wiens & Donoghue, 2004; Wiens & Graham, 2005). Sin embargo, existen pruebas de que el cambio de nicho ocurre y constituye un factor clave en diversos procesos evolutivos relacionados con el cambio climático, invasiones biológicas, cambios macroecológicos y en la especiación simpátrica (Broennimann *et al.*, 2007; Levin & Lammers, 2005).

La disponibilidad de capas paleoclimáticas nos proporciona la oportunidad de contrastar, con evidencia independiente, hipótesis filogeográficas que hayan sido influidas por cambios climáticos del pasado. Anteriormente, las áreas de refugio y la distribución histórica de las especies eran inferidas mediante la presencia de fósiles, datos palinológicos o macro-restos, o por la concentración de diversidad genética dentro de un área localizada. El uso combinado de métodos filogeográficos y modelos de distribución de especies ayuda a entender los patrones de biodiversidad en la actualidad y a inferir la distribución potencial de especies actual y del pasado para reconstruir los patrones de colonización, diferenciación y biogeografía de las especies (Vega *et al.*, 2010).

Además, estos modelos de distribución de especies permiten precisar las condiciones ecológicas que precisa el hábitat una especie, aunque la relación entre la distribución de las especies y las variables predictivas depende de la adecuación de los predictores usados en el modelo (Araujo & Guisan, 2006). El nicho ecológico que ocupa una especie ha sido descrito como un hipervolumen de *n*-dimensiones, donde cada dimensión corresponde a los factores bióticos y abióticos con los que el organismo se relaciona. Hutchinson (1957) definió dos tipos de nicho: 1) nicho fundamental, el espacio *n*- dimensional donde la especie es capaz de persistir en ausencia de competición con otras especies (factores abióticos), 2) nicho realizado, es la parte del nicho fundamental

donde la especie es capaz de vivir en presencia de competición con otra especie (factores bióticos, Araujo & Guisan, 2006). Hay autores que consideran que los modelos proporcionan una representación espacial del nicho fundamental (Soberon & Peterson, 2005), mientras que otros consideran que los modelos proporcionan una aproximación del nicho realizado (Austin, Nicholls & Margules, 1990; Pearson & Dawson, 2003). En la mayoría de casos, los modelos se construyen teniendo en cuenta sólo los factores abióticos (es decir, el área geográfica que presenta buenas condiciones ambientales para la especie) sin la interacción con los factores bióticos, por lo tanto, sería más conservador pensar que obtenemos una representación del nicho fundamental de la especie. Sin embargo, no puede descartarse que la no colonización de parte del nicho potencial no se deba a poca capacidad competitiva en esas zonas.

Solapamiento de nicho bioclimático

La transición ecológica o su ausencia determinan los distintos modelos de especiación en plantas. De forma general, se asume que el desplazamiento del nicho de una especie se produce en dos pasos: primero, el establecimiento de poblaciones mal adaptadas en lugares con nicho distinto pero donde las oportunidades ecológicas lo permiten, y segundo, el refinamiento genético de tales poblaciones que les permite integrarse en nuevas comunidades y hábitats. Estos pasos son más fáciles de lograr en floras no saturadas, como las islas, donde la competición es menos intensa (Levin, 2004).

Entender cómo el aislamiento geográfico y/o ecológico influye en el proceso de especiación es objetivo de distintas líneas de investigación. El estudio de especies hermanas, una de ellas con un área de distribución muy definida por su naturaleza insular, puede permitir calibrar la importancia relativa que tienen estos dos factores en la especiación, ya que estas especies están más relacionadas entre sí que con cualquier otra, lo que les permite conservar una mayor huella de su atributos geográficos y ecológicos al tiempo que se simplifica el componente evolutivo. Así mismo, el grado de solapamiento de nicho bioclimático actual entre especies hermanas puede ayudar a comprender el papel del aislamiento geográfico en la especiación, ya que permite excluir algunos escenarios. Por ejemplo, un gran solapamiento de nicho actual sería incompatible con una especiación simpátrica (Anacker & Strauss, 2014)

OBJETIVO PRINCIPAL

El objetivo principal de esta memoria doctoral consiste en analizar los procesos evolutivos que dieron lugar a la divergencia entre *Lavatera maritima* y *Lavatera acerifolia* y su posterior expansión, realizando estudios de modelización de distribución de especies con proyección a escenarios pasados, y estudios filogeográficos. Una hipótesis en cuanto al primer aspecto es que determinar los nichos de ambas especies debería ayudar a comprender los procesos evolutivos implicados en la especiación. En cuanto a la filogeografía, se pretende conocer la historia evolutiva de *L. maritima,* así como el origen geográfico de *L. acerifolia* y el modelo de colonización del archipiélago Canario. De este modo se podría responder a una pregunta relevante en biología y biogeografía de islas: ¿Las especies endémicas de islas oceánicas divergen y especian localmente en las islas? o ¿proceden de antecesores pre-adaptados que colonizaron el archipiélago ya diferenciadas de sus especies hermanas?

Los **objetivos específicos** que pretendemos alcanzar para examinar las hipótesis de trabajo son los siguientes:

- Evaluar la estructura genética de *L. maritima*.
- Calcular la diversidad genética distribuida en las poblaciones de *L. maritima*.
- Estimar las relaciones filogenéticas entre las secuencias de ADN plastidial a nivel individual de *L. maritima.*
- Examinar la distribución potencial actual de *L. maritima*, y su proyección a escenarios pasados, concretamente al Máximo Interglaciar (22000 años) y al Último Máximo Glaciar (130000 años).
- Estimar la estructura genética del endemismo canario *L. acerifolia*.
- Valorar la diversidad genética a lo largo de la distribución geográfica en las Islas Canarias de *L. acerifolia*.
- Averiguar las posibles rutas de colonización de *L. acerifolia* en el archipiélago canario.
- Analizar la distribución potencial de *L. acerifolia* en la actualidad y en tiempos pasados (mitad del Holoceno, 6000 años, y en Último Máximo Glaciar).
- Esclarecer la aproximación metodológica bioinformática más adecuada para conseguir numerosos y fiables SNPs de genomas poliploides a partir de datos GBS.

- Estimar el tiempo de divergencia entre *L. maritima* y *L. acerifolia*.
- Valorar el solapamiento de nicho ambiental entre las dos especies hermanas, evaluando el nicho climático de *L. maritima* y *L. acerifolia*.
- Inferir las variables ambientales más influyentes que determinan el nicho ecológico de ambas especies (*L. maritima* y *L. acerifolia*).
- Dilucidar el posible modo de especiación y proceso de divergencia que tuvo lugar entre las dos especies (*L. maritima* y *L. acerifolia*).

En el **capítulo 1** se detalla el patrón de distribución geográfica y la estructura poblacional de *L. maritima*, teniendo en cuenta factores ecológicos, geográficos y genéticos, haciendo uso de secuencias de ADN plastidial y de modelos de distribución de especie en la actualidad y su proyección al pasado.

En el **capítulo 2** se estima la estructura genética de una especie endémica canaria, *L. acerifolia*, y se describe el patrón de colonización a lo largo de las Islas Canarias, a partir de polimorfismos en un único nucleótido (SNPs) extraídos desde datos GBS, además del estudio de modelización de distribución de especies.

En el **capítulo 3** se estudia el proceso de divergencia entre *L. maritima* y *L. acerifolia*, teniendo en cuenta el solapamiento de nicho ambiental y las variables climáticas más influyentes en el hábitat que ocupan cada una, así como una comparación del nicho bioclimático de los grupos genéticos que conforman las poblaciones de ambas especies obtenidos con datos GBS.

REFERENCIAS

Álvarez I, Wendel JF. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417-434.

Anacker BL, Strauss SY. 2014. The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proceedings of the Royal Society of London B: Biological Sciences, 281(1778), 20132980.*

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17: 81-92.

Araujo MB, Guisan A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33: 1677-1688.

Austin M, Nicholls A, Margules CR. 1990. Measurement of the realized qualitative niche: environmental niches of five Eucalyptus species. *Ecological monographs* 60: 161-177.

Avise JC. 2000. *Phylogeography: the history and formation of species*. Cambridge, MA: Harvard University Press.

Avise JC. 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography* 36: 3-15.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one* 3: e3376.

Bates DM. 1968. Generic Relationships in the Malvaceae, tribe Malveae. *Gentes Herbariorum*. 10: 117–135.

Bayer C, Kubitzki K. 2003. Malvaceae. *The Families and Genera of Vascular Plants*. Berlin: Springer. 5: 225-311.

Bennett KD, Provan J. 2008. What do we mean by 'refugia'? *Quaternary Science Reviews* 27: 2449-2455.

Besnard G, Khadari B, Navascués M, Fernández-Mazuecos M, El Bakkali A, Arrigo N, Baali-Cherif D, de Caraffa VB-B, Santoni S, Vargas P. 2013. The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society of London B:* (Vol. 280, No. 1756, p. 20122833).

Blondel J. 2010. *The Mediterranean region: biological diversity in space and time*. Oxford University Press.

Blondel J, Aronson J. 1999. *Biology and wildlife of the Mediterranean region*. Oxford University Press, USA.

Bocquet G, Widler B, Kiefer H. 1978. The Messinian model-A new outlook for the floristics and systematics of the Mediterranean area. *Candollea* 33: 269-287.

Bramwell D, Bramwell ZI. 1990. Flores silvestres de las Islas Canarias. *Madrid: Editorial Rueda xvi, 376p.-illus., col. illus., maps.. ISBN 847207062X Sp Icones, Keys. Translated from the original English Wild flowers of the Canary Islands (publ. 1974) Geog* 1.

Broennimann O, Treier UA, Müller-Schärer H, Thuiller W, Peterson A, Guisan A. 2007. Evidence of climatic niche shift during biological invasion. *Ecology Letters* 10: 701-709.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* 18: 249-256.

Carine MA, Russell SJ, Santos-Guerra A, Francisco-Ortega J. 2004. Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in *Convolvulus* (Convolvulaceae). *American Journal of Botany* 91: 1070-1085.

Carstens B, Lemmon AR, Lemmon EM. 2012. The promises and pitfalls of next-generation sequencing data in phylogeography. *Systematic Biology* 61: 713-715.

Carstens BC, Brennan RS, Chua V, Duffie CV, Harvey MG, Koch RA, McMahan CD, Nelson BJ, Newman CE, Satler JD. 2013. Model selection as a tool for phylogeographic inference: an example from the willow Salix melanopsis. *Molecular Ecology* 22: 4014-4028.

Castroviejo S, Aedo C, Cirujano S, Lainz M, Monserrat P, Morales R, Munoz Garmendia F, Navarro C, Paiva J, Soriano C. 1993. Plumbaginaceae (partim) Capparaceae. *Flora Iberica. Plantas vasculares de la Península Ibérica e Islas Baleares* 3: 730.

Catalán P, Kellogg EA, Olmstead RG. 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndhF* gene sequences. *Molecular Phylogenetics and Evolution* 8: 150-166.

Claudel C, Buerki S, Chatrou LW, Antonelli A, Alvarez N, Hetterscheid W. 2017. Large-scale phylogenetic analysis of *Amorphophallus* (Araceae) derived from nuclear and plastid sequences reveals new subgeneric delineation. *Botanical Journal of the Linnean Society* 184: 32-45.

Clegg MT. 1993. Chloroplast gene sequences and the study of plant evolution. *Proceedings of the National Academy of Sciences* 90: 363-367.

Comes HP, Kadereit JW. 1998. The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in plant science* 3: 432-438.

Corriveau JL, Coleman AW. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*: 1443-1458.

Coyne JA, Orr HA. 2004. Speciation. Sunderland, MA: Sinauer Associates, Inc.

Crisci JV, Katinas L, Posadas P. 2003. *Historical Biogeography*. Harvard University Press.

Chen Y, Li B, Olmstead RG, Cantino PD, Liu E, Xiang C. 2014. Phylogenetic placement of the enigmatic genus *Holocheila* (Lamiaceae) inferred from plastid DNA sequences. *Taxon*, 63(2), 355-366.

Chikara SK, Pandey M, Pandey S, Vaidya K, Chaudhary S. 2014. Next generation sequencing: a revolutionary tool for plant variety improvement. *American Journal of Social Issues and Humanities*.

Chung C-L, Jamann T, Longfellow J, Nelson R. 2010. Characterization and fine-mapping of a resistance locus for northern leaf blight in maize bin 8.06. *Theoretical and Applied Genetics* 121: 205-227.

Dahlgren RA, Saigusa M, Ugolini FC. 2004. The Nature, Properties and Management of Volcanic Soils *Advances in Agronomy*: Academic Press. Volume 82: 113-182.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12: 499-510.

Degenkolbe T, Do PT, Zuther E, Repsilber D, Walther D, Hincha DK, Köhl KI. 2009. Expression profiling of rice cultivars differing in their tolerance to long-term drought stress. *Plant Molecular Biology* 69: 133-153.

Dercourt J, Gaetani M, Vrielynck B, Barrier E, Biju-Duval B, Brunet M, Cadet J, Crasquin S, Sandulescu M. 2000. Peri-Tethys Palaeogeographical Atlas 2000. *Paris: Commission de la Carte Géologique du Monde/Commission for the Geologic Map of the World*.

Devesa Alcaraz JA, Luque T. 1986. Contribución al estudio citotaxonómico del género Lavatera (Malvaceae) en España.

Donoghue MJ, Smith SA. 2004. Patterns in the assembly of temperate forests around the Northern Hemisphere. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 359: 1633-1644.

Edwards SV, Shultz AJ, Campbell-Staton S. 2015. Next-generation sequencing and the expanding domain of phylogeography. *Folia Zoologica* 64 (3): 187–206.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* 6: e19379.

Escobar P, Schönswetter P, Fuertes Aguilar J, Nieto Feliner G, Schneeweiss GM. 2009. Five molecular markers reveal extensive morphological homoplasy and reticulate evolution in the *Malva* alliance (Malvaceae). *Molecular Phylogenetics and Evolution* 50: 226-239.

Feliner GN, Rosselló JA. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44: 911-919.

Fernández-Palacios JM, Andersson C. 2000. Geographical determinants of the biological richness in the Macaronesian region. *Acta Phytogeographica Suecica* 85: 41-50.

Fernández-Palacios JM, de Nascimento L, Otto R, Delgado JD, García-del-Rey E, Arévalo JR, Whittaker RJ. 2011. A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography* 38: 226-246.

Fernández-Palacios JM, Otto R, Delgado JD, Arévalo JR, Naranjo A, Artiles FG, Morici C, Barone R. Los bosques.

Fineschi S, Taurchini D, Villani F, Vendramin G. 2000. Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* Mill.(Fagaceae) throughout southern European countries. *Molecular Ecology* 9: 1495-1503.

Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular ecology* 2: 113-118.

Geldmacher J, Hoernle K, van den Bogaard P, Zankl G, Garbe-Schönberg D. 2001. Earlier history of the ≥70-Ma-old Canary hotspot based on the temporal and geochemical evolution of the Selvagen Archipelago and neighboring seamounts in the eastern North Atlantic. *Journal of Volcanology and Geothermal Research* 111: 55-87.

Gómez A, Lunt DH. 2007. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula *Phylogeography of southern European refugia*: Springer: 155-188.

Guillou H, Carracedo JC, Paris R, Torrado FJP. 2004. Implications for the early shield-stage evolution of Tenerife from K/Ar ages and magnetic stratigraphy. *Earth and Planetary Science Letters* 222: 599-614.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.

Hampe A, Arroyo J, Jordano P, Petit RJ. 2003. Rangewide phylogeography of a bird-dispersed Eurasian shrub: contrasting Mediterranean and temperate glacial refugia. *Molecular Ecology* 12: 3415-3426.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907-913.

Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological journal of the Linnean Society* 58: 247-276.

Hewitt GM. 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica* 139: 617-638.

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson C. 2001. *Climate change 2001: the scientific basis*. The Press Syndicate of the University of Cambridge.

Husemann M, Schmitt T, Zachos FE, Ulrich W, Habel JC, Riddle B. 2014. Palaearctic biogeography revisited: evidence for the existence of a North African refugium for Western Palaearctic biota. *Journal of Biogeography* 41: 81-94.

Hutchinson GE. 1957. Concluding remarks. *Cold spring harbor symposium on quantitative biology* 22: 415-427.

Jiménez-López FJ, Ortiz MA, Berjano R, Talavera S, Terrab A. 2016. High population genetic substructure in *Hypochaeris leontodontoides* (Asteraceae), an endemic rupicolous species of the Atlas Mountains in NW Africa. *Alpine Botany* 126: 73-85.

Kadereit JW, Arafeh R, Somogyi G, Westberg E. 2005. Terrestrial growth and marine dispersal? Comparative phylogeography of five coastal plant species at a European scale. *Taxon* 54: 861-876.

Kim SC, Kim JS, Chase MW, Fay MF, Kim JH. 2016. Molecular phylogenetic relationships of Melanthiaceae (Liliales) based on plastid DNA sequences. *Botanical Journal of the Linnean Society* 181: 567-584.

Krijgsman W. 2002. The Mediterranean: *Mare Nostrum* of earth sciences. *Earth and Planetary Science Letters* 205: 1-12.

Krijgsman W, Hilgen F, Raffi I, Sierro F, Wilson D. 1999. Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400: 652-655.

Levin DA. 2004. The ecological transition in speciation. *New Phytologist* 161: 91-96.

Levin DA, Lammers TG. 2005. Isolate selection and ecological speciation. *Systematic Botany* 30: 233-241.

Li C, Zhou A, Sang T. 2006. Genetic analysis of rice domestication syndrome with the wild annual species, Oryza nivara. *New Phytologist* 170: 185-194.

Li M, Wunder J, Bissoli G, Scarponi E, Gazzani S, Barbaro E, Saedler H, Varotto C. 2008. Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. *Cladistics* 24: 727-745.

Lomolino MV. 2000. A call for a new paradigm of island biogeography. *Global Ecology and Biogeography* 9: 1-6.

Lomolino MV, Riddle BR, Brown JH, Brown JH. 2006. *Biogeography*. Sinauer Associates Sunderland, MA.

Losos JB. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters* 11: 995-1003.

MacArthur RH, Wilson EO. 1963. An equilibrium theory of insular zoogeography. *Evolution*: 373-387.

MacArthur RH, Wilson EO. 1967. The theory of island biogeography. *Princeton, NJ*.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7: 111-118.

Manafzadeh S, Staedler YM, Conti E. 2016. Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. *Biological Reviews*.

Mardis ER. 2008. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* 9: 387-402.

Mateu-Andrés I, Ciurana M-J, Aguilella A, Boisset F, Guara M, Laguna E, Currás R, Ferrer P, Vela E, Puche MF. 2015. Plastid DNA Homogeneity in *Celtis australis* L. (Cannabaceae) and *Nerium oleander* L. (Apocynaceae) throughout the Mediterranean Basin. *International Journal of Plant Sciences* 176: 421-432.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66: 526-538.

Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* 36: 1333-1345.

Médail F, Quezel P. 1997. Hot-spots analysis for conservation of plant biodiversity in the Mediterranean Basin. *Annals of the Missouri Botanical Garden*: 112-127.

Mittermeier RA, Myers N, Thomsen JB, Da Fonseca GAB, Olivieri S. 1998. Biodiversity Hotspots and Major Tropical Wilderness Areas: Approaches to Setting Conservation Priorities. *Conservation Biology* 12: 516-520.

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858.

Myles S. 2013. Improving fruit and wine: what does genomics have to offer? *Trends in Genetics* 29: 190-196.

Nichols RA, Hewitt GM. 1994. The genetic consequences of long distance dispersal during colonization. *Heredity* 72: 312-317.

Nieto Feliner G. 2011. Southern European glacial refugia: a tale of tales. *Taxon* 60: 365-372.

Nieto Feliner G. 2014. Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics* 16: 265-278.

Nogués-Bravo D. 2009. Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography* 18: 521-531.

Papadopoulou A, Knowles LL. 2016. Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences* 113: 8018-8024.

Pearman PB, Guisan A, Broennimann O, Randin CF. 2008. Niche dynamics in space and time. *Trends in Ecology & Evolution* 23: 149-158.

Pearson RG, Dawson TP. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12: 361-371.

Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG. 2003. Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. *Science* 300: 1563-1565.

Quézel P. 1985. Definition of the Mediterranean region and the origin of its flora. *Geobotany*.

Reyes-Betancort JA, Guerra AS, Guma IR, Humphries CJ, Carine MA. 2008. Diversity, rarity and the evolution and conservation of the Canary Islands endemic flora. *Anales del Jardín Botánico de Madrid,* 25-45.

Rieseberg LH, Soltis D. 1991. *Phylogenetic consequences of cytoplasmic gene flow in plants.* Evolutionary Trends in Plants.

Rieseberg LH, Wendel JF. 1993. Introgression and its consequences in plants. *Hybrid Zones and the Evolutionary Process*: 70-109.

Sanmartín I, Ronquist F. 2004. Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology* 53: 216-243.

Sanmartín I, Van Der Mark P, Ronquist F. 2008. Inferring dispersal: A Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography* 35: 428-449.

Soberon J, Peterson AT. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiv. Inf. Biodiversity Informatics* 2.0 (2005): n. pag. http://dx.doi.org/10.17161/bi.v2i0.4

Soltis DE, Soltis PS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis *Molecular systematics of plants II*: Springer: 1-42.

Stewart JR, Lister AM, Barnes I, Dalén L. 2010. Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society of London B: Biological Sciences* 277: 661-671.

Suc JP. 1984. Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature* 307: 429-432.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453-464.

Tate JA, Aguilar JF, Wagstaff SJ, La Duke JC, Slotta TAB, Simpson BB. 2005. Phylogenetic relationships within the tribe Malveae (Malvaceae, subfamily Malvoideae) as inferred from ITS sequence data. *American Journal of Botany* 92: 584-602.

Thompson JD, Lavergne S, Affre L, Gaudeul M, Debussche M. 2005. Ecological differentiation of Mediterranean endemic plants. *Taxon* 54: 967–976.

Tomasello S, Álvarez I, Vargas P, Oberprieler C. 2015. Is the extremely rare Iberian endemic plant species *Castrilanthemum debeauxii* (Compositae, Anthemideae) a 'living fossil'? Evidence from a multi-locus species tree reconstruction. *Molecular Phylogenetics and Evolution* 82, Part A: 118-130.

Vega R, Fløjgaard C, Lira-Noriega A, Nakazawa Y, Svenning JC, Searle JB. 2010. Northern glacial refugia for the pygmy shrew Sorex minutus in Europe revealed by phylogeographic analyses and species distribution modelling. *Ecography* 33: 260-271.

Vendramin GG, Fady B, González-Martínez SC, Hu FS, Scotti I, Sebastiani F, Soto Á, Petit RJ. 2008. Genetically depauperate but widespread: the case of an emblematic Mediterranean pine. *Evolution* 62: 680-688.

Veríssimo J, Znari M, Stuckas H, Fritz U, Pereira P, Teixeira J, Arculeo M, Marrone F, Sacco F, Naimi M, Kehlmaier C, Velo-Antón G. 2016. Pleistocene diversification in Morocco and recent demographic expansion in the Mediterranean pond turtle *Mauremys leprosa*. *Biological Journal of the Linnean Society* 119: 943-959.

Warwick SI, Mummenhoff K, Sauder CA, Koch MA, Al-Shehbaz IA. 2010. Closing the gaps: phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. *Plant Systematics and Evolution* 285: 209-232.

Weising K, Freitag H. 2007. Phylogeography of halophytes from European coastal and inland habitats. *Zoologischer Anzeiger - A Journal of Comparative Zoology* 246: 279-292.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications* 18: 315-322.

Whittaker RJ, Fernández-Palacios JM. 2007. *Island biogeography: ecology, evolution, and conservation*. Oxford University Press.

Widmer A, Lexer C. 2001. Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology & Evolution* 16: 267-269.

Wiens JJ, Donoghue MJ. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* 19: 639-644.

Wiens JJ, Graham CH. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Syst.* 36: 519-539.

Wilson EO. 1988. The current state of biological diversity. *Biodiversity* 521: 3-18.

CAPÍTULO 1

Out of North Africa twice. Phylogeography and species distribution model of the western Mediterranean *Lavatera maritima* (Malvaceae)

Villa-Machío, I., Fernández de Castro, A. G., Fuertes-Aguilar, J., Nieto Feliner, G.
Submitted to *Botanical Journal of Linnean Society*

ABSTRACT

Despite the importance of the North African region in the complex climatic and geological history of the Western Mediterranean basin, the level of sampling of that region in biogeographic and phylogeographic studies is remarkably lower than in the European part. Aiming to contribute to fill this gap, the evolutionary history of *Lavatera maritima* is reconstructed using sequence data from three cpDNA regions (*trnD-trnT*, *trnG*, *matK*), species distribution modeling and divergence time analysis. Of the nine haplotypes identified, six occur in North Africa and four are exclusive to that region. The origin of the species is estimated to be Plio-Pleistocene (c. 2.77 Myr ago) and the projection of the climatic model on the Last Interglacial indicates very low suitability for this species in Europe. North Africa is inferred to be the main genetic reservoir for this species and the source for the colonization of the European populations. Our data suggest that such colonization occurred along two waves, one mainly across Iberia northwards to southern France and the other involving long distance dispersal between continents and islands. The existence of successful possibly bird-mediated LDD events is consistent with the specific niche (limestone organic matter-rich cliffs) and the lack of phylogeographic structure found.

INTRODUCTION

The Mediterranean region is one of the 25 world biodiversity hotspots (Médail & Quezel, 1997; Myers *et al.*, 2000) hosting ca. 22.500 vascular plant species of which ca. 13.000 are endemics (Thompson *et al.*, 2005). Due to such concentration of diversity, both at the species and genetic levels, and to an active climatic and geological history that is rather well known (Dercourt *et al.*, 2000; Krijgsman, 2002), the Mediterranean region has been the object of much biogeographic interest (e.g. Médail & Diadema, 2009; Salvi, Bisconti & Canestrelli, 2016; Sanmartín, 2003; Sfenthourakis & Svenning, 2011; Valente & Vargas, 2013). A large part of the studies aiming to understand the causes for the current biotas have focused at major abiotic forces determining active speciation and range shifts through time (Caujapé-Castells & Jansen, 2003; Molins, Mayol & Rosselló, 2009; Salvo *et al.*, 2010). Under such view, climatic and geographic events such as the Messinian salinity crisis (5.96–5.33 Myr BP; Bocquet, Widler & Kiefer, 1978; Krijgsman *et al.*, 1999), the establishment of a mediterranean climate (3.2 Myr BP; Suc, 1984) and the glacial episodes along the Pleistocene have been dominant features in the biogeographic and phylogeographic literature (Fiz-Palacios & Valcárcel, 2013). However, a too strong focus on causative abiotic events and specially an unrealistic expectation of common phylogeographic patterns from similarly distributed species have been recently criticized (Papadopoulou & Knowles, 2015). Lineage-specific factors are critical too both for explaining current distribution patterns and unveiling species evolutionary histories summarized in phylogeographies (Benito-Garzón, Ruiz-Benito & Zavala, 2013). Therefore, one of the challenges of phylogeography is integrating lineage-specific factors such as environmental niche, breeding system and dispersal capacity to estimate their effects in modulating those of major common abiotic factors such as climate and geographic barriers.

Phylogeographic studies in the Mediterranean region have thrown light on some of those topics in the last years and underlined the importance of widely focused research approaches (reviewed in Nieto Feliner, 2014). However, such a demanded integration requires more empirical evidence, e.g., to explain why important geographic elements in the western Mediterranean such as the Gibraltar strait and the Sicilian channel have

played contrasting roles in exchanges between North African and European biotas, as barriers or bridges (Fernández-Mazuecos & Vargas, 2010; Hewitt, 2011; Santiso *et al.*, 2016). To accomplish some of these goals, North Africa was identified as one of the two more significant geographic sampling gaps (Nieto Feliner, 2014). Despite some researches focused on that region (Lumaret *et al.*, 2002; Naciri, Cavat & Jeanmonod, 2010; Ortiz *et al.*, 2007; Petit *et al.*, 2002), scarcity of studies with a representative sampling in North Africa still hampers addressing general questions such as the biogeographic exchanges between the western Mediterranean and the Irano-Turanian biotas through time or the role that African populations may have played in maintaining the genetic diversity of species along the Pleistocene. To fill such geographic gap, not only adequate samplings should be undertaken in North Africa in phylogeographic studies but also more studies are needed in which North Africa hosts current or past genetic reservoirs.

Despite being Malvaceae one of the most dominant plant families in neotropical areas (120 genera and 1500 species, Bayer & Kubitzki, 2003), the object of this study, *Lavatera maritima* Gouan, belongs to a highly diversified lineage of the mostly temperate subtribe Malvineae, the Malva Alliance (Bates, 1968; Tate *et al.*, 2005). *Lavatera maritima* (tree mallow) is a Mediterranean hexaploid (2n=44; Escobar *et al.*, 2009) shrub up to 2 m tall, distributed along the western Mediterranean coasts (Spain, France, Italy, Corsica, Sardinia, Tunisia, Algeria, Morocco and one single small population in Portugal) but also occurring in some continental enclaves up to 900 m.a.s.l. (Fernandes, 1993).

The interest of reconstructing the phylogeography of this species is threefold. The first focus of interest is *L. maritima* specific ecological requirements being a halo-nitrophilous shrub (Ghermaoui, Hassaine & Moulaï, 2016) growing in disturbed soils on limestones cliffs, rarely on loams and shale lands. Such a narrow niche leads to a very discontinuous range that demands dispersal and colonizing explanations for interpreting current distributional patterns. Second, coastal species facilitate inferring distributional history and identifying the geographical patterns of genetic variation because the range dimensionality is reduced. A linear distribution range minimizes the number of possible migration routes and climate-driven range shifts during glacial times so that eustatic sea-level shifts during the Pleistocene are the only expected factors altering the linear

distribution pattern. In *L. maritima*, however, the existence of inland populations as far as 300 km from the coast poses the question of whether they have the same bioclimatic requirements and what is their phylogeographic history. The study of the intraspecific genetic structure should help to throw light on whether such inland-coastal pattern is determined by historical or ecological factors (Kadereit *et al.*, 2005). Third, occurring on the northern and southern parts of the western Mediterranean basin, this species allows studying biotic exchanges between the western European and North African biotas and particularly the occurrence of refugia in northern Africa, which have not been much studied (Husemann *et al.*, 2014; Veríssimo *et al.*, 2016) compared to the three southern European peninsulas (Hewitt, 2011; Petit *et al.*, 2002).

Our main objective thus is to elucidate the underlying causes for the present geographic distribution patterns and population structure of *L. maritima*, considering ecological, geographical and genetic components. To reconstruct the evolutionary history of *L. maritima* we have used maternally inherited plastid DNA (cpDNA) sequence data, which display stronger inter-population differentiation than biparental markers due to low effective population size and avoid the problems of nuclear markers in polyploids (Twyford *et al.*, 2013). To allow a finer reconstruction of the evolutionary history and an accurate interpretation of the current genetic diversity patterns, we have elaborated species distribution models (SDM) and projected them to past climatic conditions during the Last Glacial Maximum (LGM) and Last Interglacial (LIG). Approaches combining SDM and genetic data peaked eight years ago, have already thrown considerable light on Mediterranean biogeographic history (e.g. Benito Garzón, Sánchez de Dios & Sáinz Ollero, 2007) and are particularly useful when attempting to single out the likely effects of major abiotic forces on particular species.

In sum, we undertake an integrated time-calibrated phylogeographic analysis exploring the genetic association between cpDNA sequence data and the geological and climatic history of the Mediterranean basin. Specifically, in this study we (i) asses the genetic structure of *L. maritima*, (ii) estimate lineage divergence times, (iii) examine the causal interaction between climatic and geological changes and its current genetic structure, and (iv) identify refugia and possible colonization routes.

MATERIALS AND METHODS

Sampling strategy, DNA extraction and plastid DNA sequencing

A total of 120 individuals from 43 populations of *L. maritima* spanning the whole distribution range of the species were sampled in the field preserving fresh leaves in silica gel. In addition, we used herbarium specimens from four different populations collected in Africa: one individual from Tunisia (Jbel Gafsa, MA-77089), two individuals from Algeria (Oran, MA-77088 and the Jbel Milock, MA-841613) and one individual from Morocco (Jbel Ansitten, MA-77086) (Table S1).

Total genomic DNA was extracted from dried leaves using DNeasy Plant Minikit (QIAGEN Inc., California) according to the manufacturer's instructions. After a pilot study, we selected three cpDNA regions that were consistently amplified and variable, including one intergenic spacer region (*trnD-trnT*), one intron (*trnG*) and one gene (*matK*). Three individuals per population were amplified due to the proximity between populations in the most heavily sampled areas and the low level of intrapopulation variability found. The primers and methodology for amplification of these three DNA regions via PCR are described in table S2. Amplified products were purified with ExoSAP-IT (USB Corporation, Ohio) and submitted to STAB Vida Lda (Portugal), Secugen SL (CIB-CSIC, Madrid, Spain) or Macrogen Inc. (Seoul, South Korea) for sequencing. Resulting sequence data were edited, aligned by hand and concatenated using Geneious v8.1.2 (Biomatters Ltd., Auckland, New Zealand).

Phylogeographic and genetic diversity analyses

A 95% statistical parsimony network was constructed in TCS v1.21 (Clement, Posada & Crandall, 2000) to infer genealogical relationships among haplotypes. Gaps resulting from the only informative indel and from mononucleotide repeat units (poly-T and poly-A) were treated as missing data. Additionally, a binary character (A/C) was added to code the indel as a single event.

Geographical structure was assessed applying spatial analysis of molecular variance (SAMOVA 1.0, Dupanloup, Schneider & Excoffier, 2002). This algorithm defines groups of populations that are geographically homogeneous and maximally differentiated

from other such groups and identifies genetic barriers among them. The number of random initial conditions was set to 100 as recommended by Dupanloup, Schneider and Excoffier (2002). The number of groups (K) examined ranged from 2 to 20. The optimal partition of the data, i.e., the number (K) and composition of genetic groups is the one with the highest F_{CT} (proportion of total genetic variance due to differences among population groups). As SAMOVA could not find an optimal K (see results), populations were grouped by a geographical criterion in order to use the resulting groups in subsequent analyses. Eight different regions were considered: W Morocco, N Morocco, Algeria-Tunisia, S Spain, E Spain, N Spain, France, and Islands.

The partition of genetic diversity within and among populations was examined with an analysis of molecular variance (AMOVA) using Arlequin v3.11 (Excoffier, Laval & Schneider, 2005) with 10000 permutations. AMOVA was performed for (i) the whole data set (nonhierarchical AMOVA), and recognizing some hierarchical levels under two hypotheses: (ii) considering the above mentioned eight regions and (iii) grouping those eight regions into two: northern Africa and Europe.

To assess the genetic diversity from cpDNA sequences across the geographical range of *L. maritima*, we calculated the number of haplotypes (*h*), haplotypic diversity (*Hd*), and nucleotide diversity (π) with DnaSP v5.10.01 (Librado & Rozas, 2009).

The existence of phylogeographic structure, i.e., that different haplotypes within the same population are more closely related among them than to haplotypes from other populations, was tested by the permutation test between G_{ST} and N_{ST} (coefficients of genetic differentiation) implemented in PERMUT1.0 (1000 permutations). G_{ST} only considers haplotype frequencies ("unordered" alleles), whereas N_{ST} considers both haplotype frequencies and their genetic relatedness ("ordered" alleles). The test is significant when $N_{ST} > G_{ST}$ (Pons & Petit, 1996).

Divergence times of cpDNA lineages

Lineage divergence time was estimated using two approaches in BEAST v.1.8 (Drummond *et al.*, 2012). First, a higher-level data set spanning the subfamily Malvoideae was used to estimate the crown node age of *L. maritima*. Sequences of two cpDNA regions *(matk*, *ndhF*) were obtained from GenBank for 67 different species of the subfamily (from tribes

Hibisceae, Gossypieae and Malveae, Table S3), including *Lavatera maritima* and its sister species, *L. acerifolia* according to Escobar *et al.* (2009). *Radyera farragei* (F.Muell.) Fryxell & S.H.Hashmi was treated as outgroup following Baum *et al.* (2004). The *Malvaciphyllum macondicus* fossil (Carvalho *et al.*, 2011) was used to assign a minimum age constraint on the stem node of Eumalvoideae (Baum *et al.*, 2004; Fig. 1) using a lognormal distribution with an offset of 56.5 Myr (mean=0.0, SD=0.9). An uncorrelated lognormal clock using a GTR+G+I substitution model, calculated by jModelTest (Posada, 2008), was selected and a Birth-Death process was specified as tree prior. A uniform prior for the ucld.mean with values among $1.0e^{-5}$–$1.0e^{-2}$ substitution/site/Myr (initial value=0.001) and a default exponential prior for ucld.stdev were employed. Two independent Monte Carlo Markov Chains (MCMCs) were run for 10 million generations, sampling every 1000^{th} generation. A second analysis using all haplotypes found in *L. maritima* (*matK, trnD-trnT, trnG*; see results) was run to estimate phylogeographic split ages within *L. maritima*. *Lavatera acerifolia* was used as outgroup. The root node was calibrated with the median crown age of *L. maritima* obtained in the first analysis (Malvoideae, Fig. 1, node A) applying a normal distribution and a strict clock using a uniform prior for clock.rate ($1.0e^{-4}$–$1.0e^{-2}$, initial value= 0.001). Two independent MCMC runs of 200 million generations, sampling every 1000 generations, were conducted using a Coalescent:constant-size process tree prior and the GTR model. MCMC samples from the two analyses were inspected in Tracer v1.5 (Rambaut & Drummond, 2009) to examine convergence of the chains and assure that ESS values were above 200 for all parameters. A maximum clade credibility tree was elaborated with TreeAnnotator v1.8 from the posterior distribution, discarding 10% samples as burn-in. The resulting tree was visualized using FigTree v1.4 (Rambaut, 2012).

Phylogenetic analyses

To estimate relationships among sequences, a phylogenetic tree was built from concatenated cpDNA sequences using *Lavatera acerifolia* as outgroup. To identify an optimal partitioning scheme and corresponding model of sequence evolution for the dataset, we used PartitionFinder v1.1 (Lanfear *et al.*, 2012) using the Akaike information criterion (AIC). We used RaxML v8.1 (Stamatakis, 2014) to infer a maximum-likelihood

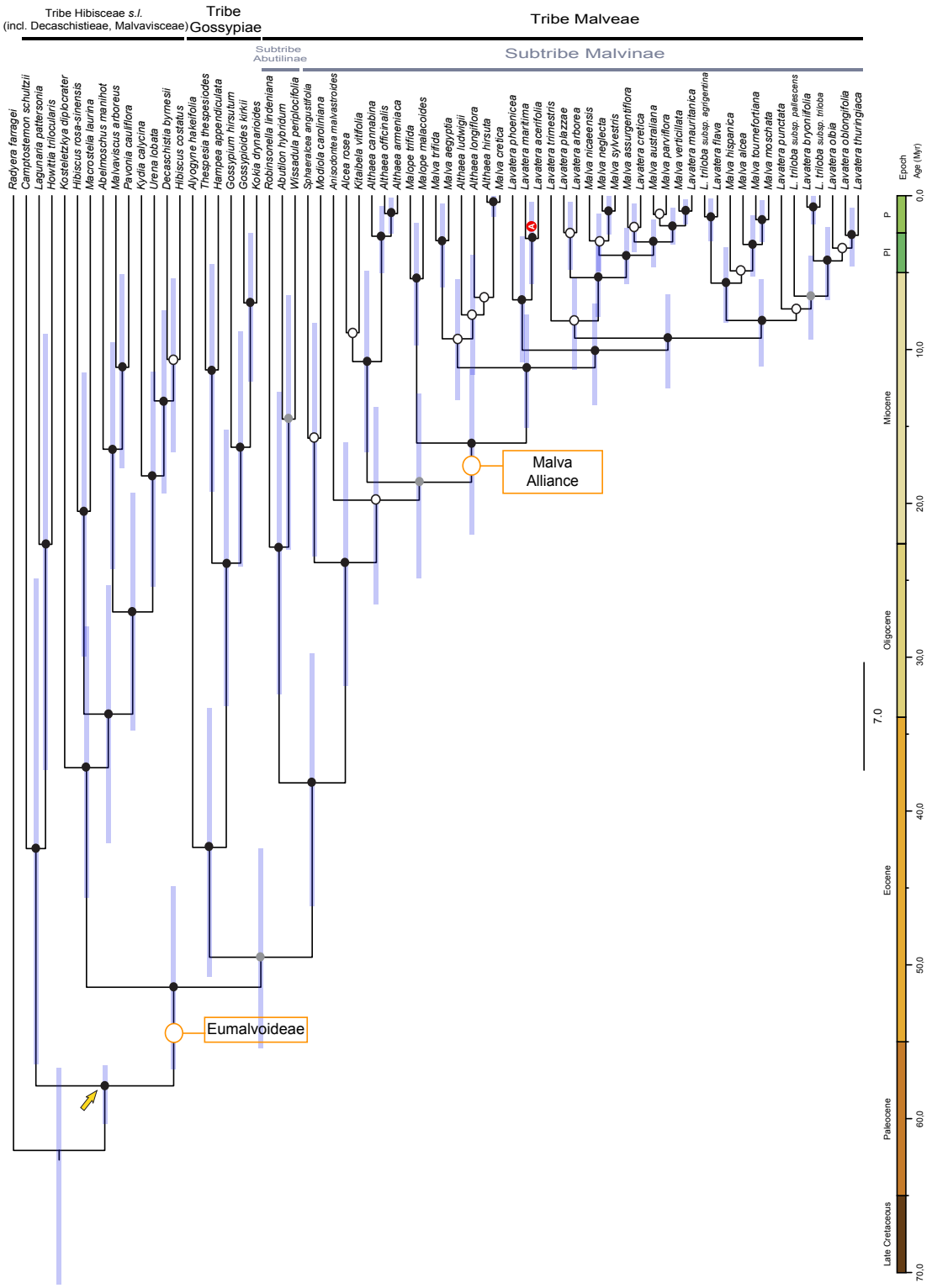


Figure 1. BEAST-derived chronogram of Malvoideae based on cpDNA (*matk*, *ndhF*) sequences using Birth-Death process with calibrations denoted by the yellow arrow. Blue bars indicate the 95% highest posterior density (HPD) credibility for node ages. Nodes with black circles have posterior probabilities (PP) higher than 0.95, nodes with grey circles have PP between 0.90 and 0.949, whereas nodes with white circles have PP lower than 0.89.

(ML) tree. Clade support was assessed with 1000 thorough bootstrap replicates after selecting the best tree from 100 generated runs. Bayesian Inference (BI) was also followed to reconstruct relationships using MrBayes v3.2.3 (Ronquist *et al.*, 2012). TVM and TrN+I were the best models of sequence evolution for the two partitions obtained by PartitionFinder v1.1. Analyses were run for 10 million generations, sampling every 1000th generation. Since branch length values were unrealistically high compared to those obtained by the ML method, the default value λ=10 was increased to λ =100 through the command "brlenspr=unconstrained:exp(100)" (Meseguer, Aldasoro & Sanmartín, 2013; Zamora *et al.*, 2014). Chain convergence was assessed with Tracer v1.5 and trees were visualized using FigTree v1.4.

Species distribution modelling (SDM)

In order to test the potential of bioclimatic layers to explain species presence and choose variables accordingly, we performed a variance partitioning analysis (Borcard, Legendre & Drapeau, 1992) using the 'vegan' library (Oksanen *et al.*, 2012) within R software (R Core Team, 2015). Our predictors were 19 bioclimatic variables (temperature and precipitation) downloaded from the WorldClim website (http: //www. worldclim.org/ bioclim, Hijmans *et al.*, 2005), which are widely used and considered to be biologically meaningful variables for characterizing species range (Buermann *et al.*, 2008), plus 2 topographic variables derived from elevation data: slope and topographic position index (TPI, Guisan, Weiss & Weiss, 1999). TPI measures the relative topographic position of the central point as the difference between the elevation at this point and the mean elevation of its eight surrounding cells (De Reu *et al.*, 2013). Variance partitioning allows calculating the relative contribution of groups of predictors to the explained variance of a response variable (in our case, species presence) (Rocchini *et al.*, 2014). First, predictors were clustered into two groups: climatic (temperature and precipitation) and topographic (slope and TPI) variables. Then, predictors were sorted in three groups: temperature, precipitation and topography.

The distribution of climate suitability was modelled using Maxent v3.3 (Phillips, Anderson & Schapire, 2006) that estimates the optimal potential distribution with a maximum entropy algorithm using presence-only data (Elith *et al.*, 2006). Eighteen

bioclimatic variables were used following the result of our previous statistical analysis after excluding bio09 due to evident distortions when applied to *L. maritima* range. The occurrence data were randomly split into training (75%) and test (25%) data for model evaluation and ten subsample replicates were performed. A jackknife analysis was carried out to measure variable importance (Phillips, Anderson & Schapire, 2006). We used the average prediction from all the model replicates to construct the species distribution maps. In addition, to help reconstructing the evolutionary history of *L. maritima*, the distribution model obtained for the present time was projected to the Last Glacial Maximum (LGM: 20-26.5 kyr; Clark *et al.*, 2009) using paleoclimatic layers simulated under the Model for Interdisciplinary Research on Climate (MIROC, Hasumi & Emori, 2004), and to the Last Interglacial (LIG, 130-115 kyr) envelope using the climatic model of Otto-Bliesner *et al.* (2006). The climate was reconstructed at a scale of 30 arc-seconds (ca. 1 km^2) for current and LIG scenarios, and 2.5 arc-minutes (c. 5 km^2) for LGM scenario (Hijmans *et al.*, 2005). In order to explore the behavior of the most influential environmental variables in these three scenarios, we use probability density plots as a post-hoc method.

Additionally, to explore the climatic differences between inland and coastal populations using SDM, the same analysis with Maxent was run without 10 inland populations (Grand Vallon, Mallos de Riglos, Estopiñán del Castillo, Castillonroy, Embid de la Ribera, Alquézar, Ólvena, Otiñar, Jbel Milock and Jbel Gafsa).

RESULTS

cpDNA variation and haplotype geographic distribution

The alignment of the concatenated plastid regions (*matK, trnD-trnT, trnG*) from 120 individuals of *L. maritima* was 2107 bp long. It included 10 polymorphic sites of which 9 were parsimony informative and one indel (5-bp long) that was also so.

Nine haplotypes (H1-H9) were identified by TCS v1.21 in the sequence data set across the 43 surveyed populations (Fig. 2B). The most common haplotypes, H1 and H2, were separated by two mutational steps. H1 was the most frequent and had the highest number of connections along the network. H1 was broadly distributed in northern

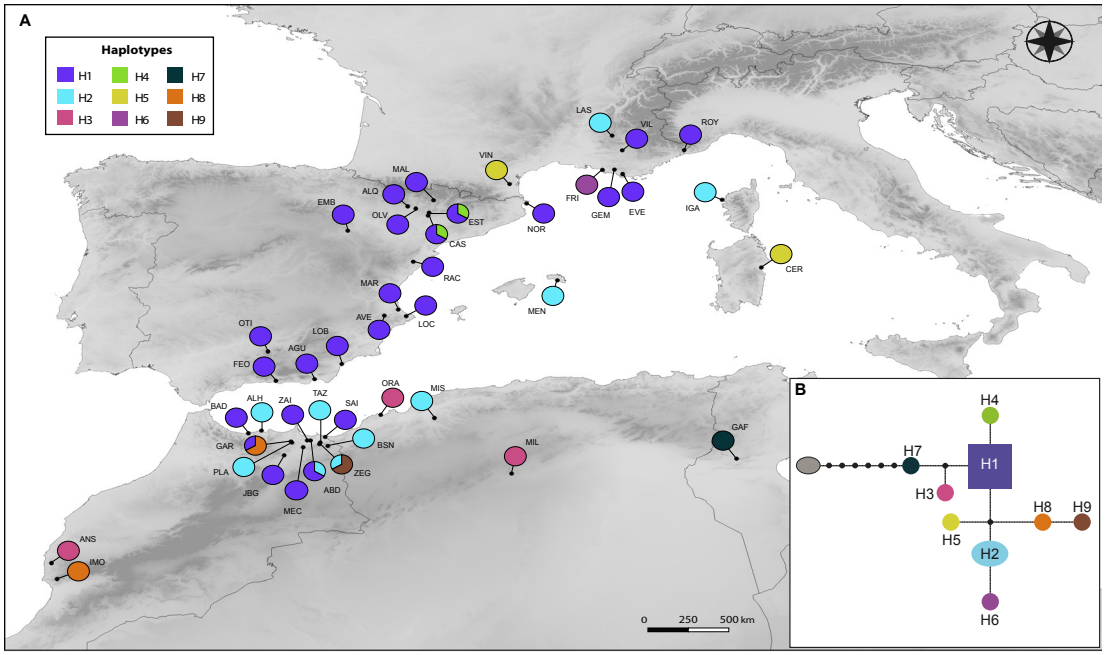


Figure 2. Analysis of cpDNA (*matK*, *trnD-trnT*, *trnG*) haplotypes of *Lavatera maritima*. (a) Geographic distribution of the cpDNA haplotypes (see Table S1 for population codes). (b) Statistical parsimony network of genealogical relationships between the 9 haplotypes of *L. maritima*. Lines represent single nucleotide substitutions and small black circles indicate missing haplotypes. Grey circle represents the outgroup *L. acerifolia*.

Morocco, the Iberian Peninsula and French coastal and subcoastal populations. H2 was found in north Morocco, Mediterranean islands (Minorca and Corsica) and a French population in the Hautes-Alpes department. H3 was detected in herbarium specimens, one from southwest Morocco and two from Algeria (Oran and Jbel Milock). Two populations in Huesca (Estopiñán del Castillo and Castillonroy) shared haplotype H4. H5 grouped Sardinia with one French population in the Pyrénées-Orientales (Vingrau). Two additional unique haplotypes were found: H6 in l'Île de Ratonneau (offshore Marseille, France) and H7 in Jbel Gafsa (Tunisia). H8 and H9 only occurred in Moroccan populations, both in the north and H8 additionally in the southwest (Fig. 2A). Relationships among sequences reconstructed through phylogenetic analyses produced the same topologies under the ML and BI approaches and the majority rule consensus tree from BI showed supports above 95% for all nodes (Fig. S1). Phylogenetic analyses results were congruent with haplotype relationships depicted in the network analysis (Fig. 2).

Table 1. Analysis of molecular variance (AMOVA) of *Lavatera maritima* for cpDNA sequences considering the whole range of distribution (43 populations) and two hierarchical levels. Each level clusters all populations in 8 and 2 geographical groups respectively.

Region analysed	d.f.	SS	Percentage of variation
Nonhierarchical AMOVA			
Whole range (43 populations) (F_{ST}= 0,89)			
Among populations	42	150.83	89.35
Wtihin populations	77	11.333	10.65
Total	119	162.17	
Hierarchical AMOVA			
Eigh geographic regions (F_{ST}= 0,89)			
Among regions	7	46.500	18.45
Among populations within regions	35	104.33	71.22
Within populations	77	11.333	10.33
Total	119	162.17	
Two geographic regions (F_{ST}= 0,89)			
Among populations	1	10.24	8.18
Among populations within regions	41	140.59	81.62
Within populations	77	11.330	10.2
Total	119	162.17	

Genetic diversity and differentiation

Spatial genetic analyses of cpDNA haplotypes using SAMOVA indicated that F_{CT} steadily increased with growing values of K from 2 to 20 groups (Fig. S2). Thus, to explore the partitioning of genetic diversity, the sampling area was divided into eight groups based on geographic distribution for AMOVA analysis. The nonhierarchical AMOVA showed that the highest percentage of variation (89.35%) was explained by differences among populations whereas 10.65% (P<0.00001) was explained by differences within populations. The F_{ST} value (0.893) was significantly different from zero, indicating genetic structure among the 43 populations (Table 1). Hierarchical AMOVA based on eight geographical regions (W Morocco, N Morocco, Algeria-Tunisia, S Spain, E Spain, N Spain, France, and Islands) revealed that differences among populations within groups account for the highest percentage of the genetic variance (71.22%) whereas differences among groups explain 18.45%. Like in the non-hierarchical AMOVA, significant genetic differentiation (F_{ST}= 0.896) was found among the eight geographical regions considered. Analogous results were obtained running a hierarchical AMOVA in which only two geographical regions were considered (N Africa and Europe). The permutation test across the distribution range showed that N_{ST} (0.419) was very similar to G_{ST} (0.486),

Table 2. Genetic diversity parameters from cpDNA sequences across the geographical range of *L. maritima*. *n*, number of sampled individuals; *h*, number of haplotypes; *Hd* (± SD), haplotypic diversity (± standard deviation); π, nucleotide diversity. See Table S1 for population codes.

Region (populations)	cpDNA				
	n	h	Hd (± SD)	π	Haplotype (no. of individuals)
W Morocco (IMO, ANS)	4	2	0.500 (0.265)	0.00096 (0.00051)	H8(3), H3(1)
N Morocco (ABD, BAD, BSN, MEC, GAR, ZEG, JBG, PLA, ALH, SAI, TAZ, ZAI)	36	4	0.610 (0.048)	0.00063 (0.00007)	H1(18), H2(14), H8(2), H9(2)
Algerie and Tunisia (ORA, MIL, GAF, MIS)	6	3	0.733 (0.155)	0.00127 (0.00025)	H2(3), H3(2), H7(1)
S Spain (AGU, LOS, FEO, OTI)	12	1	0.000	0.00000	H1(12)
E Spain (LOC, AVE, MAR, RAC)	12	1	0.000	0.00000	H1(12)
N Spain (ALQ, NOR, CAS, EMB, EST, MAL, OLV, VIN)	24	3	0.366 (0.115)	0.00030 (0.00010)	H1(19), H4(2), H5(3)
France (EVE, GEM, LAS, FRI, ROY, VIL)	18	3	0.529 (0.115)	0.00060 (0.00012)	H1(12), H2(3), H6(3)
Island (IGA, CER, MEN)	8	2	0.429 (0.169)	0.00041 (0.00016)	H2(6), H5(2)
TOTAL	120	8	0.572 (0.042)	0.00058 (0.00005)	

indicating a lack of phylogeographic structure (Pons & Petit, 1996), which is consistent with lack of distinct genetic groups resulting from the SAMOVA. Total nucleotide (π) and haplotype (*H*d) diversity across all populations were 0.00058 and 0.572, respectively. Northern Africa (N Morocco and Algeria-Tunisia) exhibits the highest haplotype diversity (*H*d = 0.610, 0.733, respectively) while eastern and southern Spain have fixed haplotypes (*H*d = 0.000) (Table 2).

Divergence times of cpDNA lineages

The fossil-calibrated cpDNA (*matk, ndhF*) phylogeny of Malvoideae estimated an average divergence time for the split between *L. maritima* and *L. acerifolia* of 2.77 Myr ago (95% HPD: 0.43–5.75 Myr ago; node A, Fig. 1); thus between Messinian (end of the Miocene) and Middle Pleistocene. The second calibrated tree analyzing only *L. maritima* obtained low nodal support values (results not shown) and is not further considered.

Modelling past and present climate suitability

The estimation of the effects of climatic versus topographic variables using variance partitioning produced clear results to explain species presences. The quantitative contribution of climate variables to species presence was 47% and that of topography was below 0% (Fig. 3A). In a second analysis, with three groups of variables, the major contribution was explained by those predictors associated to temperature (18%) and precipitation (16%) (Fig. 3B).

The geographic distribution of the climate suitability for *L. maritima* under current conditions spans its actual distribution in the western Mediterranean basin, including most sampled localities of the species (Fig. 4). The predicted areas encompass regions as southeastern France, eastern and southern Spain, western and northern Africa and

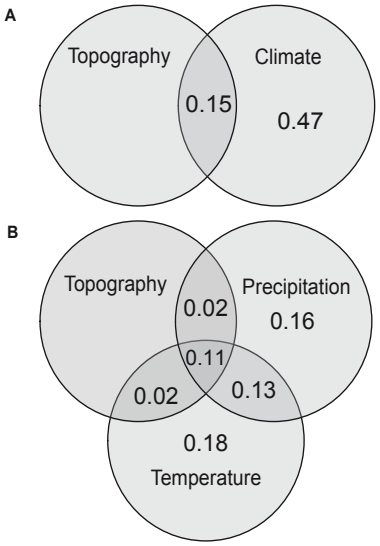


Figure 3. Venn diagram of a variance partitioning analyses showing the percentage of variance that contributes to species presence explained by (A) climatic vs. topographic variables and (B) precipitation, temperature and topographic variables. In both approaches, the total percentage of variance explained was 62% and the residual percentage was 39%. Values <0 not shown.

islands like Minorca or Sardinia. However, the analysis identifies additional areas with moderate probability like southern Portugal, northern Spain (Asturias) or Sicily where, despite the comprehensive floristic knowledge, *L. maritima* has never been recorded. The AUC value was high (0.935), indicating a high predictive power for the model. The predicted geographic distribution of habitat suitability during the LGM (20-26.5 kyr ago) was similar to the present time. But the area with suitable climatic conditions for *L. maritima* in northern Africa was larger during the LGM than in the present time model. The projection to the LIG (115-130 kyr ago) contrasts with the other two previously described scenarios since it shows that the suitable climatic conditions for *L. maritima* spanned a large area along the southwest Moroccan coast and Canary Islands compared to northern Africa, where the area was smaller than in the LGM and present time, and specially to Europe, where the area was extremely reduced. The reduced suitability of environments along the Spanish Mediterranean fringe during the LIG was explored with probability density plots of the most influential variables (bio04, bio12, bio14, bio17, bio19). Precipitation of coldest quarter (bio19) and temperature seasonality (bio04) showed distinctly higher values in the LIG than in the LGM and current time (Fig. S3).

With regard to the modelling without 10 inland populations, the results showed a substantial loss of climate suitability for all interior areas (Fig. S4) both in Africa and Europe, e.g. in southern Portugal, Ebro river basin, northern Spain (Asturias) and inland French areas.

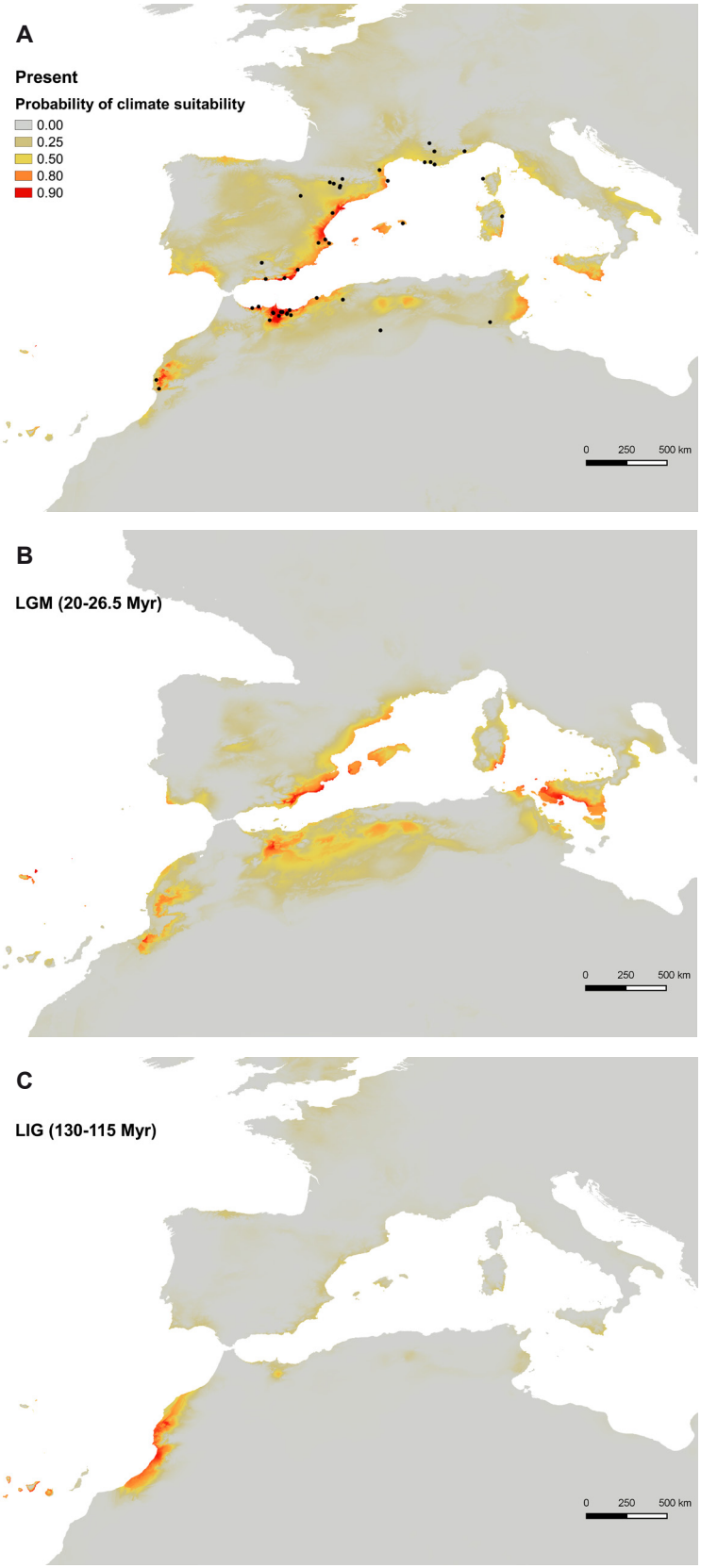


Figure 4. Result of a climate suitability model using Maxent (A) under current conditions, (B) the projection to the Last Glacial Maximum (LGM, 20-26.5 kyr) and (C) the Last Interglacial (LIG, 130-115 kyr). Dots indicate sampled populations for this study.

DISCUSSION

Biogeographic role of North Africa during Pleistocene

Glacial refugia have long been recognized to have played an important role in species range shifts during the Pleistocene in Europe (Cheddadi *et al.*, 2009; Tzedakis, 2007). In the classic phylogeographic scheme for widespread species in this continent, high levels of current genetic diversity are often related with regions that acted as glacial refugia (Avise, 2000; Widmer & Lexer, 2001) whereas low levels are indicative of recently colonized areas (Kadereit *et al.*, 2005). In contrast, the number of glacial refugia identified in North Africa is scarcer, despite the importance of North African Mediterranean refugia first suggested by Battandier (1894). Different factors may have contributed to this lack of recognition of North African refugia including a lower sampling of current genetic data compared to Europe, and substantially lower amounts of available fossil pollen data (Elenga *et al.*, 2000). An underestimation of both plant dispersal capacities across the Mediterranean sea (Guzmán & Vargas, 2009; Piñeiro *et al.*, 2007) and of Pleistocene land bridges (Fernández-Mazuecos & Vargas, 2011; Garnatje *et al.*, 2013; Mayol *et al.*, 2012) has probably led to underrating the connectivity of African and European biotas. On the other hand, some biological factors may have hampered the identification of North African glacial genetic reservoirs, the most significant being the overall aridification trend in that region during the Pleistocene (Zhang *et al.*, 2014), in addition to the effects of Holocene agricultural practices (Cheddadi *et al.*, 2015; Mercuri, Sadori & Ollero, 2011). Despite such scarcity, several studies have identified North African refugia for others plant species (Lumaret *et al.*, 2002; Ortiz *et al.*, 2007; Petit *et al.*, 2002).

Our finding of the highest haplotype and nucleotide diversity for *Lavatera maritima* in North Africa (six of the nine haplotypes, four of them exclusive) suggests this area as the main refugium for this species (Table 2, Fig. 2A). High values of haplotype diversity were also detected in France, which contrasted with a lack of haplotype diversity in populations from southern and eastern Spain. These results support the view that current Iberian populations are the result of post-glacial colonization.

The comparison of the temporal projections of the SDM adds another perspective that has been successfully used in phylogeographic studies over the last decade (Fernández-Mazuecos & Vargas, 2013; Richards, Carstens & Lacey Knowles, 2007; Rodríguez-Sánchez *et al.*, 2010; Waltari *et al.*, 2007). The restriction of areas with the highest probability for the climate model to coastal SW Morocco and the Canary Islands during the LIG (Fig. 4C) points out to a refugium or reservoir of genetic diversity for *L. maritima* in North Africa. But an older occurrence in SW Morocco beyond the LIG --even an origin of the species-- is also feasible provided that its sister species *L. acerifolia* is endemic to the Canary Islands (Fuertes-Aguilar *et al.*, 2002). This pattern is in sharp contrast to the projection of the SDM on the LGM where highly suitable areas were found in eastern Iberia, the islands of Ibiza, Sardinia and Sicily, and across North Africa from the Anti-Atlas to Tunisia (Fig. 4B). The partial lack of parallelism between these suitable areas during the LGM and the cpDNA sequence data, in particular the low diversity in Iberia, requires a historical rather than just an ecological explanation and is discussed below.

Considering the suitable areas for the model through time, only SW Morocco provides favorable environmental conditions in the three temporal scenarios examined. The existence of SW Moroccan lineages in other plant groups (García-Castaño *et al.*, 2014; Hardion *et al.*, 2016), including also the Canaries in *Laurus* (Rodríguez-Sánchez *et al.*, 2009), is consistent with the existence of refugial areas there, which have been specifically proposed in Médail and Diadema (2009). Therefore, evidence combined from the plastid sequences and the SDM suggest that North Africa hosted the bulk of the species genetic diversity at least during the Last Interglacial while european populations were very restricted or even absent. But, what is the evidence about the tempo and ways by which *L. maritima* expanded toward Europe?

Evolutionary history of *L. maritima*

Two lineages can be identified in the haplotype network, separated by two mutational steps (Fig. 2). The first lineage contains the most common and widespread haplotype (H1), occurring from the Rif Mountains to the Maritime Alps and along eastern Iberia, together with a local haplotype restricted to Aragon (NE Spain, H4) and two others

restricted to North Africa: H7 in Tunisia, H3 in SW Morocco, the Rif and Algeria (Fig. 2). The second lineage contains the second most frequent haplotype (H2) together with four others, two of which are singletons (H6, H9). Interestingly, the three non-singleton haplotypes for this lineage occur in geographically distant (Rif-SW Morocco; H8) and even disjunct locations: Rif Mountains-Minorca-Corsica-Provence (H2), Pyrénées-Orientales--Sardinia (H5) (Fig. 2). The fact that SW Morocco harbors two exclusive North African haplotypes each belonging to one of the two lineages (H3, H8) can be an additional indication that such area was an important genetic reservoir. It is likely that subsequent bottleneck events rarified the presence of this species in that area. Field surveys in SW Morocco compared to herbarium records from different times over the 20th century suggest that severe grazing contributed to population decline with many vouchered populations becoming extinct. According to this hypothesis, it is possible that the current haplotype diversity in the Rif Mountains and western Algeria partly reflects the diversity that existed in SW Morocco before the species decline in that area: in addition to those two haplotypes H3, H8, the two most frequent H1, H2, and the singleton H9 from the second lineage (Fig. 2).

On the other hand, PERMUT and SAMOVA analyses indicate a lack of phylogeographic structure. Provided that substantial haplotype variation has been found in this species (9 haplotypes), the absence of an association between genetic data and geography is best explained by an important contribution of long-distance dispersal, which is consistent with the geographically disjunct locations of several haplotypes.

Low support for internal branches within *L. maritima* in the calibrated tree hinders the possibility of describing a temporally detailed scenario where the inferred lineage splits could be associated to colonization events. However, the integration of all the elements here analyzed does allow drawing some conclusions and suggesting a tentative overall scenario for the evolutionary history of *L. maritima* and in particular for the colonization of the European continent. These elements are the estimated age for the stem node of *L. maritima* i.e., the split from *L. acerifolia* (2.77 Myr ago, 95% HPD: 0.43–5.75 Myr ago), the existence of two plastid lineages, the lack of phylogeographic structure, the current haplotype distribution and the SDM data.

Based on the available data, it is feasible that the colonization of European sites from North Africa took place along two different postglacial waves, corresponding to the two plastid lineages identified. The first lineage, containing two exclusive North African haplotypes likely dispersed through eastern Spain northwards to southern France mainly through land at a fast pace. The most frequent haplotype (H1) representing such colonization wave likely followed a leading edge model of colonization. This scenario is compatible with the projection of the SDM on the LIG where high suitable sites are lacking in the European mainland and islands (Fig. 4C). The projection on the LGM, where highly suitable sites occur along eastern Spain together with the Rif Mountains and northern Algeria (Fig. 4B), suggests a suitable spatio-temporal frame for such a colonization wave. However, as already mentioned, we have no support for dating the split of this first lineage. The extreme scarcity of climatically suitable areas along the Spanish Mediterranean fringe during the LIG is puzzling for a cold-sensitive species like *L. maritima*, although it is consistent with the current low diversity for this species in that area. However, the probability density plots on this area along the three temporal projections have allowed identifying two variables (bio04, bio19) whose values in the Spanish Mediterranean fringe during the LIG differed substantially from those in the present time and LGM (Fig. S3). Therefore, possibly in addition to other causes and components of the niche of *L. maritima*, our data suggest that the precipitation of the coldest quarter (bio19) and the temperature seasonality (bio04) prevented the occurrence of suitable sites during the LIG. These variables indicate that during the LIG the coldest quarter was moister than in the current and LGM scenarios, and also warmer. A similar pattern has been found in *Ceratonia siliqua* L. during the LIG (Viruel *et al.*, 2016).

On the contrary, the second plastid lineage likely reached the European mainland and western Mediterranean islands primarily by long-distance dispersal (LDD) events. This hypothesis is based on the current disjunct distribution of three haplotypes belonging to this lineage (Fig. 2). The case of H2 is particularly illustrative of such disjunctions (Rif Mountains-Minorca-Corsica-Provence) but they also occur for H5 and H8. The fact that haplotypes from this second lineage are absent in the Iberian Peninsula, which is the bridging land for African and southern France populations, is consistent with such

a scenario where LDD events were important. We hypothesize that the colonization of European sites by this lineage was also recent and fast. The most influential factor for a successful colonization after LDD events is the suitability of the environment (Piñeiro *et al.*, 2007). According to the SDM projections, climatically suitable areas in southern France and the western Mediterranean islands occurred since the LGM. The exception is Corsica, which lacks suitable sites in the three temporal projections considered but hosts a population in a small offshore island (Gargalu) occurring on volcanic rocks. Phylogeographic lineages mainly composed of North African and western Mediterranean islands populations occur in other groups (García-Castaño *et al.*, 2014; González-Martínez *et al.*, 2010; Migliore *et al.*, 2012). However, underlying causes may be different since those studies focused on Miocene lineages, and therefore their phylogeographic patterns result from processes that operated at different spatial and temporal scales.

Causes for the present phylogeographic pattern

The two-wave historical scenario proposed above poses some additional questions that need to be addressed in order to evaluate its likeliness. How specific is the niche of this species and how well represented it is in the SDM? Why the Iberian populations have just one haplotype (plus a rare closely-related one)? What possible explanations could account for LDD dispersal and for fine and large-scale distribution patterns? How close are the actual environmental conditions of inland populations compared to the coastal and subcoastal ones? Is the pattern detected in *L. maritima* similar to those in other cold-sensitive lowland plants? Jointly addressing these questions leads to some insights into the causes for the present-day patterns of genetic diversity in this species.

Comparing the phylogeographic pattern of *Lavatera maritima* to those of other non-narrowly distributed cold-sensitive Mediterranean plants throws some light on possible underlying causes. When species showing a very low genetic diversity are set aside (*Rosmarinus officinalis* L., *Nerium oleander* L.; Mateu-Andrés *et al.*, 2013; Mateu-Andrés *et al.*, 2015), the main contrast with other non-annual species such as *Smilax aspera* L. *, Chamaerops humilis* L.*, Myrtus communis* L. and *L. nobilis L. / L. azorica* (Seub.) Franco (Chen *et al.*, 2014; García-Castaño *et al.*, 2014; Migliore *et al.*, 2012; Rodríguez-Sánchez *et al.*, 2009) seems to be the lack of phylogeographic structure in *L. maritima*. As

discussed above, frequent LDD may counteract spatial structuring of plastid haplotype variation compared to situations in which ranges are expanded by a gradual isolation by distance model. This could explain the fragmented distribution of haplotypes in distant populations, but the patchy distribution of the species seems to require other underlying causes. There are important components of the ecological niche of *L. maritima* that are not represented in our SDM. These elements are explicitly acknowledged by a phytosociologically-defined vegetation association known as *Lavaterion maritimae* that accounts for shrubs growing in nitrophilous rock crevices and cliffs, due to animal dejections, in dry warm coastal or continental western Mediterranean sites (Rivas-Martínez, 2003). Such a specific niche points to birds playing a role in fruit dispersal and edaphic conditions. The stochasticity involved in such dispersal mode could explain both the patchy distribution displayed and the lack of phylogeographic structure found in *L. maritima*. We hypothesize that at least two groups of birds can be involved. Seagulls (*Larus* spp.) are abundant in some populations (e.g., Gargalu, Corsica; F. Médail, pers. comm.) and could be suitable fruit dispersers between mainland and islands, and among islands (Fig. 2). A second potential group of dispersers is vultures possibly affecting inland populations, where feathers have been found and constitute circumstantial evidence of their presence (J. Fuertes, pers. observ.).

Like the absence of climatically suitable sites in Mediterranean Iberian peninsula during the LIG, the current occurrence of inland continental populations of *L. maritima* is a bit puzzling. The most abundant ones are surrounding the Ebro river basin (NE Spain), but there are also other inland locations in southern Spain (Otiñar, Jaén), the Grand Vallon (Hautes-Alpes) in France and Algeria (Jbel Milock) and Tunisia (Jbel Gafsa) (Fig. 2). The sites where they occur fit the ecological niche described above, and it can be explained by bird migrations habits. However, one would think that climatic conditions differ substantially from those in coastal and subcoastal populations. The SDM constructed excluding the 10 inland continental sites corroborated such prediction since the suitable inland areas were markedly reduced compared to the model with all presence data (Fig. S4). The fact that climatic conditions in those inland areas do differ from those in coastal and subcoastal areas could indicate a potential differentiation of inland populations

involving a wider tolerance to some degree of continentality. This hypothesis could be addressed with a study involving more variable molecular markers. An alternative hypothesis is a transient occurrence in those inland localities following bird-mediated dispersal events, which that might not be successful in the long term. The fact that the most frequent haplotype (H1) predominates in these populations is consistent with this alternative hypothesis; this scenario would be somehow analogue to the population of the Gargalu island offshore Corsica.

The lack of genetic diversity in the Iberian populations is another relevant finding. Since there are many suitable sites along eastern Spain, the lack of diversity and specifically the absence of haplotypes from the second lineage could require both a fluid south-north dispersion following a leading edge model and a low probability of Iberia being involved in stochastic bird-mediated dispersal events. The visible pattern, however, is that the Alboran sea, separating close southern Spanish and northern African coasts, is an effective barrier, unlike what happens in other groups (Silva *et al.*, 2015).

In summary, the phylogeographic patterns found in *L. maritima* can be due to a combination of elements such as a specific niche that is effectively reached by bird-mediated fruit dispersal, a Plio-Pleistocene origin, the availability of North African genetic reservoirs, the scarcity of climatically suitable areas for this species in Europe during the LIG, the importance of LDD and a two-wave colonization of Europe.

ACKNOWLEDGEMENTS

The authors are grateful to J.C. Moreno, J.C. Zamora, F. Durán, I. Álvarez, M. Kaid-Harche, R. Riina, S. Manzano, D. Orgaz, A. Baumel, F. Médail, M. Pires for providing samples or assistance with fieldwork; to F. Durán, A. Herrero, E. Cano for lab assistance; to I. Liberal, M. Fernández-Mazuecos, J.L. Blanco-Pastor, P. Jiménez-Mejías, J.V. Sandoval-Sierra, J. Fernández-López, M. Mairal, I. Sanmartin for their help in different analyses; and to J. Viruel for sharing results and discussion from his work. Funding was provided by research grants from the Spanish Ministry of Science, Technology and Innovation and European Fund of Regional Development (CGL2010-16138) to JFA and Ministry of Economy and Competitiveness (CGL2013-49097-C2-1-P) to GNF and through the FPI fellowship BES-2011-049852 (co-financed by European Social Fund) to I.V-M.

REFERENCES

Avise JC. 2000. *Phylogeography: the history and formation of species*. Cambridge, MA: Harvard University Press.

Bates DM. 1968. Generic Relationships in the Malvaceae, tribe Malveae. *Gentes Herbariorum*. 10: 117–135.

Battandier JA. 1894. Considérations sur les plantes réfugiées, rares ou en voie d'extinction de la flore algerienne *Association Française pour l'Avancement des Sciences*. Congrès de Caen, Paris.

Baum DA, DeWitt Smith S, Yen A, Alverson WS, Nyffeler R, Whitlock BA, Oldham RL. 2004. Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. *American Journal of Botany* 91: 1863-1871.

Bayer C, Kubitzki K. 2003. Malvaceae. *The Families and Genera of Vascular Plants*. Berlin: Springer. 5: 225-311.

Benito-Garzón M, Ruiz-Benito P, Zavala MA. 2013. Interspecific differences in tree growth and mortality responses to environmental drivers determine potential species distributional limits in Iberian forests. *Global Ecology and Biogeography* 22: 1141-1151.

Benito Garzón M, Sánchez de Dios R, Sáinz Ollero H. 2007. Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. *Ecography* 30: 120-134.

Bocquet G, Widler B, Kiefer H. 1978. The Messinian model-A new outlook for the floristics and systematics of the Mediterranean area. *Candollea* 33: 269-287.

Borcard D, Legendre P, Drapeau P. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73: 1045-1055.

Buermann W, Saatchi S, Smith TB, Zutta BR, Chaves JA, Milá B, Graham CH. 2008. Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography* 35: 1160-1176.

Carvalho MR, Herrera FA, Jaramillo CA, Wing SL, Callejas R. 2011. Paleocene Malvaceae from northern South America and their biogeographical implications. *American Journal of Botany* 98: 1337-1355.

Caujapé-Castells J, Jansen RK. 2003. The influence of the Miocene Mediterranean desiccation on the geographical expansion and genetic variation of *Androcymbium gramineum* (Cav.) McBride (Colchicaceae). *Molecular Ecology* 12: 1515-1525.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. 2009. The Last Glacial Maximum. *Science* 325: 710-714.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1659.

Cheddadi R, Fady B, François L, Hajar L, Suc JP, Huang K, Demarteau M, Vendramin GG, Ortu E. 2009. Putative glacial refugia of *Cedrus atlantica* deduced from Quaternary pollen records and modern genetic diversity. *Journal of Biogeography* 36: 1361-1371.

Cheddadi R, Nourelbait M, Bouaissa O, Tabel J, Rhoujjati A, López-Sáez JA, Alba-Sánchez F, Khater C, Ballouche A, Dezileau L. 2015. A history of human impact on Moroccan mountain landscapes. *African Archaeological Review* 32: 233-248.

Chen C, Qi ZC, Xu XH, Comes HP, Koch MA, Jin XJ, Fu CX, Qiu YX. 2014. Understanding the formation of Mediterranean-African-Asian disjunctions: evidence for Miocene climate-driven vicariance and recent long-distance dispersal in the Tertiary relict *Smilax aspera* (Smilacaceae). *New Phytologist* 204: 243-255.

De Reu J, Bourgeois J, Bats M, Zwertvaegher A, Gelorini V, De Smedt P, Chu W, Antrop M, De Maeyer P, Finke P, Van Meirvenne M, Verniers J, Crombé P. 2013. Application of the topographic position index to heterogeneous landscapes. *Geomorphology* 186: 39-49.

Dercourt J, Gaetani M, Vrielynck B, Barrier E, Biju-Duval B, Brunet M, Cadet J, Crasquin S, Sandulescu M. 2000. Peri-Tethys Palaeogeographical Atlas 2000. *Paris: Commission de la Carte Géologique du Monde/Commission for the Geologic Map of the World*.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969-1973.

Dupanloup I, Schneider S, Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11: 2571-2581.

Elenga H, Peyron O, Bonnefille R, Jolly D, Cheddadi R, Guiot J, Andrieu V, Bottema S, Buchet G, De Beaulieu JL. 2000. Pollen-based biome reconstruction for southern Europe and Africa 18,000 yr bp. *Journal of Biogeography* 27: 621-634.

Elith J, H. Graham C, P. Anderson R, Dudík M, Ferrier S, Guisan A, J. Hijmans R, Huettmann F, R. Leathwick J, Lehmann A, Li J, G. Lohmann L, A. Loiselle B, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC. M. Overton J, Townsend Peterson A, J. Phillips S, Richardson K, Scachetti-Pereira R, E. Schapire R, Soberón J, Williams S, S. Wisz M, E. Zimmermann N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.

Escobar P, Schönswetter P, Fuertes Aguilar J, Nieto Feliner G, Schneeweiss GM. 2009. Five molecular markers reveal extensive morphological homoplasy and reticulate evolution in the Malva alliance (Malvaceae). *Molecular Phylogenetics and Evolution* 50: 226-239.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47-50.

Fernandes R. 1993. *Lavatera* L. In: Castroviejo S, Aedo C, Cirujano S, Laínz M, Montserrat P, Morales R, Muñoz Garmendia F, Navarro C, Paiva J and Soriano C, eds. *Flora iberica*. Madrid: Real Jardín Botánico, CSIC. 3: 232-243.

Fernández-Mazuecos M, Vargas P. 2011. Historical isolation versus recent long-distance connections between Europe and Africa in bifid toadflaxes (*Linaria* sect. *Versicolores*). *PLoS One* 6: e22234.

Fernández-Mazuecos M, Vargas P. 2013. Congruence between distribution modelling and phylogeographical analyses reveals Quaternary survival of a toadflax species (*Linaria elegans*) in oceanic climate areas of a mountain ring range. *New Phytologist* 198: 1274-1289.

Fernández-Mazuecos M, Vargas P. 2010. Ecological rather than geographical isolation dominates Quaternary formation of Mediterranean *Cistus* species. *Molecular Ecology* 19: 1381-1395.

Fiz-Palacios O, Valcárcel V. 2013. From Messinian crisis to Mediterranean climate: a temporal gap of diversification recovered from multiple plant phylogenies. *Perspectives in Plant Ecology, Evolution and Systematics* 15: 130-137.

Fuertes-Aguilar J, Ray MF, Francisco-Ortega J, Santos-Guerra A, Jansen RK. 2002. Molecular evidence from chloroplast and nuclear markers for multiple colonizations of *Lavatera* (Malvaceae) in the Canary Islands. *Systematic Botany* 27: 74-83.

García-Castaño JL, Terrab A, Ortiz MÁ, Stuessy TF, Talavera S. 2014. Patterns of phylogeography and vicariance of *Chamaerops humilis* L. (Palmae). *Turkish Journal of Botany* 38: 1132-1146.

Garnatje T, Pérez-Collazos E, Pellicer J, Catalán P. 2013. Balearic insular isolation and large continental spread framed the phylogeography of the western Mediterranean *Cheirolophus intybaceus* s.l. (Asteraceae). *Plant Biology* 15: 166-175.

Ghermaoui M, Hassaine K, MoulaÏ R. 2016. Influence du Goéland leucophée Larus michahellis sur les formations végétales ouvertes du littoral de Rachgoun (Ouest Oranie, Algérie). *Revue d'écologie* 71: 250-265.

González-Martínez SC, Dubreuil M, Riba M, Vendramin G, Sebastiani F, Mayol M. 2010. Spatial genetic structure of *Taxus baccata* L. in the western Mediterranean Basin: past and present limits to gene movement over a broad geographic scale. *Molecular Phylogenetics and Evolution* 55: 805-815.

Guisan A, Weiss SB, Weiss AD. 1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology* 143: 107-122.

Guzmán B, Vargas P. 2009. Long-distance colonization of the Western Mediterranean by *Cistus ladanifer* (Cistaceae) despite the absence of special dispersal mechanisms. *Journal of Biogeography* 36: 954-968.

Hardion L, Dumas P-J, Abdel-Samad F, Kharrat MBD, Surina B, Affre L, Médail F, Bacchetta G, Baumel A. 2016. Geographical isolation caused the diversification of the Mediterranean thorny cushion-like *Astragalus* L. sect. *Tragacantha* DC.(Fabaceae). *Molecular phylogenetics and evolution* 97: 187-195.

Hasumi H, Emori S. 2004. K-1 coupled model (MIROC) description. K-1 Technical Report 1. *Center for Climate System Research, University of Tokyo, Tokyo*.

Hewitt GM. 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica* 139: 617-638.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.

Husemann M, Schmitt T, Zachos FE, Ulrich W, Habel JC, Riddle B. 2014. Palaearctic biogeography revisited: evidence for the existence of a North African refugium for Western Palaearctic biota. *Journal of Biogeography* 41: 81-94.

Kadereit JW, Arafeh R, Somogyi G, Westberg E. 2005. Terrestrial growth and marine dispersal? Comparative phylogeography of five coastal plant species at a European scale. *Taxon* 54: 861-876.

Krijgsman W. 2002. The Mediterranean: *Mare Nostrum* of earth sciences. *Earth and Planetary Science Letters* 205: 1-12.

Krijgsman W, Hilgen F, Raffi I, Sierro F, Wilson D. 1999. Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400: 652-655.

Lanfear R, Calcott B, Ho SY, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution* 29: 1695-1701.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Lumaret R, Mir C, Michaud H, Raynal V. 2002. Phylogeographical variation of chloroplast DNA in holm oak (*Quercus ilex* L.). *Molecular Ecology* 11: 2327-2336.

Mateu-Andrés I, Aguilella A, Boisset F, Currás R, Guara M, Laguna E, Marzo A, Puche M, Pedrola J. 2013. Geographical patterns of genetic variation in rosemary (*Rosmarinus officinalis*) in the Mediterranean basin. *Botanical Journal of the Linnean Society* 171: 700-712.

Mateu-Andrés I, Ciurana M-J, Aguilella A, Boisset F, Guara M, Laguna E, Currás R, Ferrer P, Vela E, Puche MF. 2015. Plastid DNA Homogeneity in *Celtis australis* L. (Cannabaceae) and *Nerium oleander* L. (Apocynaceae) throughout the Mediterranean Basin. *International Journal of Plant Sciences* 176: 421-432.

Mayol M, Palau C, Rosselló JA, González-Martínez SC, Molins A, Riba M. 2012. Patterns of genetic variability and habitat occupancy in *Crepis triasii* (Asteraceae) at different spatial scales: insights on evolutionary processes leading to diversification in continental islands. *Annals of Botany* 109: 429-441.

Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* 36: 1333-1345.

Médail F, Quezel P. 1997. Hot-spots analysis for conservation of plant biodiversity in the Mediterranean Basin. *Annals of the Missouri Botanical Garden*: 112-127.

Mercuri AM, Sadori L, Ollero PU. 2011. Mediterranean and north-African cultural adaptations to mid-Holocene environmental and climatic changes. *The Holocene* 21: 189-206.

Meseguer AS, Aldasoro JJ, Sanmartín I. 2013. Bayesian inference of phylogeny, morphology and range evolution reveals a complex evolutionary history in St. John's wort (*Hypericum*). *Molecular phylogenetics and evolution* 67: 379-403.

Migliore J, Baumel A, Juin M, Médail F. 2012. From Mediterranean shores to central Saharan mountains: key phylogeographical insights from the genus *Myrtus*. *Journal of Biogeography* 39: 942-956.

Molins A, Mayol M, Rosselló JA. 2009. Phylogeographical structure in the coastal species *Senecio rodriguezii* (Asteraceae), a narrowly distributed endemic Mediterranean plant. *Journal of Biogeography* 36: 1372-1383.

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858.

Naciri Y, Cavat F, Jeanmonod D. 2010. *Silene patula* (Siphonomorpha, Caryophyllaceae) in North Africa: A test of colonisation routes using chloroplast markers. *Molecular phylogenetics and evolution* 54: 922-932.

Nieto Feliner G. 2014. Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics* 16: 265-278.

Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Simpson G, Solymos P, Henry M, Stevens H. 2012. Vegan: community ecology package, v2.0-3.

Ortiz MÁ, Tremetsberger K, Talavera S, Stuessy T, GarcÍA-CastaÑO JL. 2007. Population structure of *Hypochaeris salzmanniana* DC. (Asteraceae), an endemic species to the Atlantic coast on both sides of the Strait of Gibraltar, in relation to Quaternary sea level changes. *Molecular Ecology* 16: 541-552.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A. 2006. Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *science* 311: 1751-1753.

Papadopoulou A, Knowles LL. 2015. Species-specific responses to island connectivity cycles: refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Molecular ecology* 24: 4252-4268.

Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu J-L, Kremer A. 2002. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* 156: 49-74.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.

Piñeiro R, Aguilar JF, Munt DD, Feliner GN. 2007. Ecology matters: Atlantic–Mediterranean disjunction in the sand-dune shrub *Armeria pungens* (Plumbaginaceae). *Molecular Ecology* 16: 2155-2171.

Pons O, Petit RJ. 1996. Measwring and Testing Genetic Differentiation With Ordered Versus Unordered Alleles. *Genetics* 144: 1237-1245.

Posada D. 2008. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution* 25: 1253-1256.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rambaut A. 2012. FigTree v1. 4. Available at: http://tree.bio.ed.ac.uk/software/figtree.

Rambaut A, Drummond A. 2009. Tracer v1.5, Available at http://beast.bio.ed.ac.uk/Tracer.

Richards CL, Carstens BC, Lacey Knowles L. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography* 34: 1833-1845.

Rivas-Martínez S. 2003. Parietarietea Rivas-Martínez ex Rivas-Goday 1964 es un nombre válido. *Fitosociologia* 40: 33-34.

Rocchini D, Dadalt L, Delucchi L, Neteler M, Palmer MW. 2014. Disentangling the role of remotely sensed spectral heterogeneity as a proxy for North American plant species richness. *Community Ecology* 15: 37-43.

Rodríguez-Sánchez F, Hampe A, Jordano P, Arroyo J. 2010. Past tree range dynamics in the Iberian Peninsula inferred through phylogeography and palaeodistribution modelling: a review. *Review of Palaeobotany and Palynology* 162: 507-521.

Rodríguez-Sánchez F, Guzmán B, Valido A, Vargas P, Arroyo J. 2009. Late Neogene history of the laurel tree (*Laurus* L., Lauraceae) based on phylogeographical analyses of Mediterranean and Macaronesian populations. *Journal of Biogeography* 36: 1270-1281.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539-542.

Salvi D, Bisconti R, Canestrelli D. 2016. High phylogeographical complexity within Mediterranean islands: insights from the Corsican fire salamander. *Journal of Biogeography* 43: 192-203.

Salvo G, Ho SY, Rosenbaum G, Ree R, Conti E. 2010. Tracing the temporal and spatial origins of island endemics in the Mediterranean region: a case study from the citrus family (*Ruta* L., Rutaceae). *Systematic Biology* 59: 705-722.

Sanmartín I. 2003. Dispersal vs. vicariance in the Mediterranean: historical biogeography of the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *Journal of Biogeography* 30: 1883-1897.

Santiso X, Lopez L, Retuerto R, Barreiro R. 2016. Phylogeography of a widespread species: pre-glacial vicariance, refugia, occasional blocking straits and long-distance migrations. *AoB Plants* 8.

Sfenthourakis S, Svenning JC. 2011. Mediterranean biogeography: where history meets ecology across scales. *Frontiers of Biogeography* 3: 7-9.

Silva JL, Lim S-Y, Kim S-C, Mejías JA. 2015. Phylogeography of cliff-dwelling relicts with a highly narrow and disjunct distribution in the western Mediterranean. *American Journal of Botany* 102: 1538-1551.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.

Suc JP. 1984. Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature* 307: 429-432.

Tate JA, Aguilar JF, Wagstaff SJ, La Duke JC, Slotta TAB, Simpson BB. 2005. Phylogenetic relationships within the tribe Malveae (Malvaceae, subfamily Malvoideae) as inferred from ITS sequence data. *American Journal of Botany* 92: 584-602.

Thompson JD, Lavergne S, Affre L, Gaudeul M, Debussche M. 2005. Ecological differentiation of Mediterranean endemic plants. *Taxon* 54: 967–976.

Twyford AD, Kidner CA, Harrison N, Ennos RA. 2013. Population history and seed dispersal in widespread Central American Begonia species (Begoniaceae) inferred from plastome-derived microsatellite markers. *Botanical Journal of the Linnean Society* 171: 260-276.

Tzedakis P. 2007. Seven ambiguities in the Mediterranean palaeoenvironmental narrative. *Quaternary Science Reviews* 26: 2042-2066.

Valente LM, Vargas P. 2013. Contrasting evolutionary hypotheses between two mediterranean-climate floristic hotspots: the Cape of southern Africa and the Mediterranean Basin. *Journal of Biogeography* 40: 2032-2046.

Veríssimo J, Znari M, Stuckas H, Fritz U, Pereira P, Teixeira J, Arculeo M, Marrone F, Sacco F, Naimi M, Kehlmaier C, Velo-Antón G. 2016. Pleistocene diversification in Morocco and recent demographic expansion in the Mediterranean pond turtle *Mauremys leprosa*. *Biological Journal of the Linnean Society* 119: 943-959.

Viruel J, Médail F, Juin M, Haguenauer A, Nieto Feliner G, Dagher-Kharrat MB, La Malfa S, Ouahmane L, Sanguin H, Baumel A. 2016. Mediterranean carob populations, native or naturalized? A continuing riddle. *International Conference of Ecological Sciences*. Marseille: doi: 10.13140/RG.2.2.33681.84328.

Waltari E, Hijmans RJ, Peterson AT, Nyári ÁS, Perkins SL, Guralnick RP. 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS ONE* 2: e563.

Widmer A, Lexer C. 2001. Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology & Evolution* 16: 267-269.

Zamora JC, de Diego Calonge F, Hosaka K, Martín MP. 2014. Systematics of the genus *Geastrum* (Fungi: Basidiomycota) revisited. *Taxon* 63: 477-497.

Zhang Z, Ramstein G, Schuster M, Li C, Contoux C, Yan Q. 2014. Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene. *Nature* 513: 401-404.

SUPPORTING INFORMATION

Table S1. Populations of *Lavatera maritima* sampled for this study.

Population	Codes	Latitude	Longitude	Locality	Colector
Oran	ORA	-0.70405	35.73708	Argelia, Orán	A. Faure (MA-77088)
Jbel Milock	MIL	2.84500	33.92804	Argelia, Laghouat	L. Faurel (MA-841613)
Misserghin	MIS	0.75178	35.64034	Argelia, Orán	I. Álvarez & M. Kaid-Harche (MA-910937)
Jbel Gafsa	GAF	8.94496	34.38727	Tunisia, Gafsa	CJ Pitard (MA-77089)
Vingrau	VIN	2.79025	42.85441	France, Pyrrénnées Orientales	F. Médail (AIX)
Evenos	EVE	5.85515	43.16281	France, Var	A. Baumel (AIX)
Gémenos	GEM	5.63338	43.30002	France, Bouches-du-Rhone	A. Baumel (AIX)
Grand Vallon	LAS	5.57096	44.34828	France, Bouches du Rhône	A. Baumel (AIX)
Île de Ratonneau	FRI	5.30689	43.28809	France, Bouches du Rhône	A. Baumel (AIX)
Col de l'Arma	ROY	7.52514	43.90100	France, Alpes maritimes	M. Pires (AIX)
Roche amère	VIL	5.84486	43.89434	France, Alpes de Haute Provence	A. Baumel (AIX)
Île de Gargalu	IGA	8.55339	42.36927	France, Corsica	F. Médail (AIX)
Cala Gonone	CER	9.61732	40.28000	Italy, Sardinia	P. Escobar (MA-709504)
Imouzzer des Ida-Outanane	IMO	-9.48204	30.67605	Morocco, Agadir	J. Fuertes & G. Nieto (MA-853378)
Jbel Ansitten	ANS	-9.63333	31.16667	Morocco, Agadir	E. Jahandiez (MA-77086)
Abdadgadel	ABD	-2.60819	34.94898	Morocco, Zaio	A. Gonzalez & I. Villa (MA-910940)
Badés (Peñon de la Gomera)	BAD	-4.29381	35.17097	Morocco, Al-Hoceima	A. Gonzalez & I. Villa (IVM 24)
Beni Snassen Monts	BSN	-2.13904	34.77612	Morocco, Berkane	A. Gonzalez & I. Villa (MA-910944)
Embalse Mechrá-Homadi	MEC	-2.80688	34.73757	Morocco, Selouanne	A. Gonzalez & I. Villa (MA-910942)
Gareb	GAR	-3.13889	34.91750	Morocco, Tiztoutine	A. Gonzalez & I. Villa (MA-910941)
Gorges du Zegzel	ZEG	-2.36846	34.83502	Morocco, Berkane	A. Gonzalez & I. Villa (MA-910946)
Jbel Guilliz	JBG	-3.32375	34.49419	Morocco, Guercif	A. Gonzalez & I. Villa (IVM 25)
Plaine du Gareb	PLA	-3.11258	34.88392	Morocco, Tiztoutine	A. Gonzalez & I. Villa (MA-910947)
Puerto de Alhucemas	ALH	-3.93890	35.24999	Morocco, Al-Hoceima	A. Gonzalez & I. Villa (MA-910939)
Saidia	SAI	-2.21197	35.05932	Morocco, Saidia	A. Gonzalez & I. Villa (MA-910945)
Tazaguine	TAZ	-2.34392	34.88567	Morocco, Berkane	A. Gonzalez & I. Villa (MA-910943)
Zaio	ZAI	-2.69856	34.95003	Morocco, Zaio	A. Gonzalez & I. Villa (IVM 26)
Aguadulce	AGU	-2.49856	36.83790	Spain, Andalucía, Almería	F. Durán, I. Villa & J.C. Zamora (MA-910953)
Los Lobos	LOB	-1.76064	37.30581	Spain, Andalucía, Almería	F. Durán, I. Villa & J.C. Zamora (MA-910954)
Lobres (río Guadalfeo)	FEO	-3.54506	36.78714	Spain, Andalucía, Granada	F. Durán, I. Villa & J.C. Zamora (MA-910951)
Otiñar	OTI	-3.76291	37.69118	Spain, Andalucía, Jaén	F. Durán, I. Villa & J.C. Zamora (MA-910955)
L'ocaive	LOC	-0.01412	38.77895	Spain, Com. Valenciana, Alicante	A. Gonzalez (MA)
Barranco del Averno	AVE	-0.60751	38.79714	Spain, Com. Valenciana, Alicante	A. Gonzalez (MA)
Marxuquera	MAR	-0.23001	38.98666	Spain, Com. Valenciana, Valencia	A. Gonzalez (MA)
Racó del Frare	RAC	0.17813	40.46188	Spain, Com. Valenciana, Castellón	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Alquezar	ALQ	0.02758	42.16807	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Cap de Norfeu	NOR	3.25050	42.25360	Spain, Cataluña, Girona	A. Gonzalez (MA)
Castillonroy	CAS	0.57896	41.88214	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Embid de la Ribera	EMB	-1.59905	41.41685	Spain, Aragón, Zaragoza	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Estopiñán del Castillo	EST	0.60577	41.97638	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Mallos de Riglos	MAL	0.72723	42.35529225	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Ólvena	OLV	0.24546	42.09971667	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Cala de Sant Llorenç	MEN	4.08759	39.8864	Spain, Islas Baleares, Menorca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)

Table S2. Primers and methodology used for amplification of three DNA regions via PCR. Ta, annealing temperature.

Primer F	Secuencia (5'-3')	Primer R	Secuencia (5'-3')	Referencia	Ta (°C)	Elongation time (min)	Cycles (n)
matKF2	AGC CAT GAA TGT GTA GAA GAA GC	*matKRint*	TTC TAG ATG GAT GGG ATG AGG	Cronn *et al.* 2002; this study	55	1'3	35
trnT_Fw	CCG CTA GAC GAT GGG GGC	*trnT*	CTA CCA CTG AGT TAA AAG GG	This study; Grivet *et al.*,2001	60	1'3	35
trnG ^{umc}	GTA GCG GGA ATC GAA CCC GCA	*trnG2G*	GCG GGT ATA GTT TAG TGG TAA	Shaw *et al.*, 2005	60	1'3	35

Table S3. Species downloaded from GenBank and accession numbers for both markers.

Specie	ndhF sequence	matk sequence
Radyera farragei	AY589078 (Baum *et al*., 2004)	AY589063 (Baum *et al*., 2004)
Camptostemon schultzii	AF111727 (Alverson *et al*., 1999)	AY321162 (Nyffeler *et al*., unpublished)
Lagunaria pattersonia	AY589084 (Baum *et al*., 2004)	AY589064 (Baum *et al*., 2004)
Howittia trilocularis	AY589085 (Baum *et al*., 2004)	AY589065 (Baum *et al*., 2004)
Kosteletzkya diplocrater	EF207307 (Koopman and Baum, 2008)	EF207276 (Koopman and Baum, 2008)
Hibiscus rosa-sinensis	AY589075 (Baum *et al*., 2004)	AY321160 (Nyffeler *et al*., unpublished)
Macrostelia laurina	EF207299 (Koopman and Baum, 2008)	EF207267 (Koopman and Baum, 2008)
Abelmoschus manihot	AF384639 (Pfeil *et al*., 2002)	EF562457 (Koopman and Baum, 2008)
Malvaviscus arboreus	AF111718 (Alverson *et al*., 1999)	AY589061 (Baum *et al*., 2004)
Pavonia cauliflora	--	AY589056 (Baum *et al*., 2004)
Kydia calycina	EF207293 (Koopman and Baum, 2008)	EF207261 (Koopman and Baum, 2008)
Urena lobata	EF207291 (Koopman and Baum, 2008)	EF207260 (Koopman and Baum, 2008)
Decaschistia byrnesii	AY589079 (Baum *et al*., 2004)	AY589066 (Baum *et al*., 2004)
Hibiscus costatus	U55323 (Seelanan *et al*., 1997)	AY589057 (Baum *et al*., 2004)
Alyogyne hakeifolia	AY589083 (Pfeil *et al*., 2002)	AY589059 (Baum *et al*., 2004)
Thespesia thespesiodes	U55326 (Seelanan *et al*., 1997)	AY321161 (Nyffeler *et al*., unpublished)
Hampea appendiculata	AY589077 (Baum *et al*., 2004)	AY589062 (Baum *et al*., 2004)
Gossypium hirsutum	U55340 (Seelanan *et al*., 1997)	AY321158 (Nyffeler *et al*., unpublished)
Gossypioides kirkii	U55329 (Seelanan *et al*., 1997)	AF403563 (Seelanan *et al*., 1997)
Kokia drynarioides	U55330 (Seelanan *et al*., 1997)	AF403564 (Seelanan *et al*., 1997)
Robinsonella lindeniana	FJ204750 (Tate, 2011)	FJ204711 (Tate, 2011)
Abutilon hybridum	AF111716 (Alverson *et al*., 1999)	AY589058 (Baum *et al*., 2004)
Wissadula periplocifolia	FJ204717 (Tate, 2011)	FJ204713 (Tate, 2011)
Sphaeralcea angustifolia	EF207286 (Koopman and Baum, 2008)	EF207255 (Koopman and Baum, 2008)
Modiola caroliniana	EF207287 (Koopman and Baum, 2008)	EF207256 (Koopman and Baum, 2008)
Anisodontea malvastroides	EU346848 (Escobar Garcia, P., unpublished)	EU346803 (Escobar Garcia, P., unpublished)
Alcea rosea	EU346847 (Escobar Garcia, P., unpublished)	EU346805 (Escobar Garcia, P., unpublished)
Kitaibela vitifolia	EU346849 (Escobar Garcia, P., unpublished)	EU346804 (Escobar Garcia, P., unpublished)
Althaea cannabina	EU346810 (Escobar Garcia, P., unpublished)	EU346764 (Escobar Garcia, P., unpublished)
Althaea officinalis	EU346811 (Escobar Garcia, P., unpublished)	EU346765 (Escobar Garcia, P., unpublished)
Althaea armeniaca	EU346807 (Escobar Garcia, P., unpublished)	EU346763 (Escobar Garcia, P., unpublished)
Malope trifida	EU346834 (Escobar Garcia, P., unpublished)	EU346801 (Escobar Garcia, P., unpublished)
Malope malacoides	EU346833 (Escobar Garcia, P., unpublished)	EU346800 (Escobar Garcia, P., unpublished)
Malva trifida	EU346836 (Escobar Garcia, P., unpublished)	EU346799 (Escobar Garcia, P., unpublished)
Malva aegyptia	EU346835 (Escobar Garcia, P., unpublished)	EU346798 (Escobar Garcia, P., unpublished)
Althaea ludwigii	EU346812 (Escobar Garcia, P., unpublished)	EU346796 (Escobar Garcia, P., unpublished)
Althaea longiflora	EU346809 (Escobar Garcia, P., unpublished)	EU346795 (Escobar Garcia, P., unpublished)
Althaea hirsuta	EU346808 (Escobar Garcia, P., unpublished)	EU346794 (Escobar Garcia, P., unpublished)
Malva cretica	EU346837 (Escobar Garcia, P., unpublished)	EU346797 (Escobar Garcia, P., unpublished)
Lavatera phoenicea	EU346828 (Escobar Garcia, P., unpublished)	EU346802 (Escobar Garcia, P., unpublished)
Lavatera maritima	this study	this study
Lavatera acerifolia	this study	this study
Lavatera trimestris	EU346832 (Escobar Garcia, P., unpublished)	EU346774 (Escobar Garcia, P., unpublished)
Lavatera plazzae	EU346829 (Escobar Garcia, P., unpublished)	EU346773 (Escobar Garcia, P., unpublished)
Lavatera arborea	EU346821 (Escobar Garcia, P., unpublished)	EU346779 (Escobar Garcia, P., unpublished)
Malva nicaeensis	EU346843 (Escobar Garcia, P., unpublished)	EU346785 (Escobar Garcia, P., unpublished)
Malva neglecta	EU346842 (Escobar Garcia, P., unpublished)	EU346788 (Escobar Garcia, P., unpublished)
Malva sylvestris	EU346845 (Escobar Garcia, P., unpublished)	EU346787 (Escobar Garcia, P., unpublished)
Malva assurgentiflora	EU346819 (Escobar Garcia, P., unpublished)	EU346780 (Escobar Garcia, P., unpublished)
Lavatera cretica	EU346813 (Escobar Garcia, P., unpublished)	EU346783 (Escobar Garcia, P., unpublished)
Malva australiana	EU346827 (Escobar Garcia, P., unpublished)	EU346784 (Escobar Garcia, P., unpublished)
Malva parviflora	EU346844 (Escobar Garcia, P., unpublished)	EU346786 (Escobar Garcia, P., unpublished)
Malva verticillata	EU346846 (Escobar Garcia, P., unpublished)	EU346789 (Escobar Garcia, P., unpublished)
Lavatera mauritanica	EU346824 (Escobar Garcia, P., unpublished)	EU346782 (Escobar Garcia, P., unpublished)
Lavatera triloba subsp. *agrigentina*	EU346814 (Escobar Garcia, P., unpublished)	EU346769 (Escobar Garcia, P., unpublished)
Lavatera flava	EU346818 (Escobar Garcia, P., unpublished)	EU346772 (Escobar Garcia, P., unpublished)
Malva hispanica	EU346838 (Escobar Garcia, P., unpublished)	EU346793 (Escobar Garcia, P., unpublished)
Malva alcea	EU346840 (Escobar Garcia, P., unpublished)	EU346790 (Escobar Garcia, P., unpublished)
Malva tournefortiana	EU346839 (Escobar Garcia, P., unpublished)	EU346791 (Escobar Garcia, P., unpublished)
Malva moschata	EU346841 (Escobar Garcia, P., unpublished)	EU346792 (Escobar Garcia, P., unpublished)
Lavatera punctata	EU346830 (Escobar Garcia, P., unpublished)	EU346776 (Escobar Garcia, P., unpublished)
Lavatera triloba subsp. *pallescens*	EU346817 (Escobar Garcia, P., unpublished)	EU346770 (Escobar Garcia, P., unpublished)
Lavatera bryoniifolia	EU346815 (Escobar Garcia, P., unpublished)	EU346768 (Escobar Garcia, P., unpublished)
Lavatera triloba subsp. *triloba*	EU346816 (Escobar Garcia, P., unpublished)	EU346771 (Escobar Garcia, P., unpublished)
Lavatera olbia	EU346826 (Escobar Garcia, P., unpublished)	EU346766 (Escobar Garcia, P., unpublished)
Lavatera oblongifolia	EU346825 (Escobar Garcia, P., unpublished)	EU346767 (Escobar Garcia, P., unpublished)
Lavatera thuringiaca	EU346831 (Escobar Garcia, P., unpublished)	EU346775 (Escobar Garcia, P., unpublished)

References

Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. 1999. Phylogeny of the core Malvales: evidence from ndhF sequence data. American Journal of Botany 86: 1474-1486.

Cronn RC, Small RL, Haselkorn T, Wendel JF. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. American Journal of Botany 89: 707-725.

Grivet D, Heinze B, Vendramin G, Petit R. 2001. Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. Molecular Ecology Notes 1: 345-349.

Koopman MM, Baum DA. 2008. Phylogeny and biogeography of tribe Hibisceae (Malvaceae) on Madagascar. Systematic Botany 33: 364-374.

Nyffeler R, Bayer C, Alverson WS, Yen A, Whitlock BA, Chase MW, Baum DA. 2005. Phylogenetic analysis of the Malvadendrina clade (Malvaceae sl) based on plastid DNA sequences. Organisms Diversity & Evolution 5: 109-123.

Pfeil B, Brubaker CL, Craven LA, Crisp M. 2002. Phylogeny of Hibiscus and the tribe Hibisceae (Malvaceae) using chloroplast DNA sequences of ndhF and the rpl16 intron. Systematic Botany 27: 333-350.

Seelanan T, Schnabel A, Wendel JF. 1997. Congruence and consensus in the cotton tribe (Malvaceae). Systematic Botany 22: 259-290.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 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.

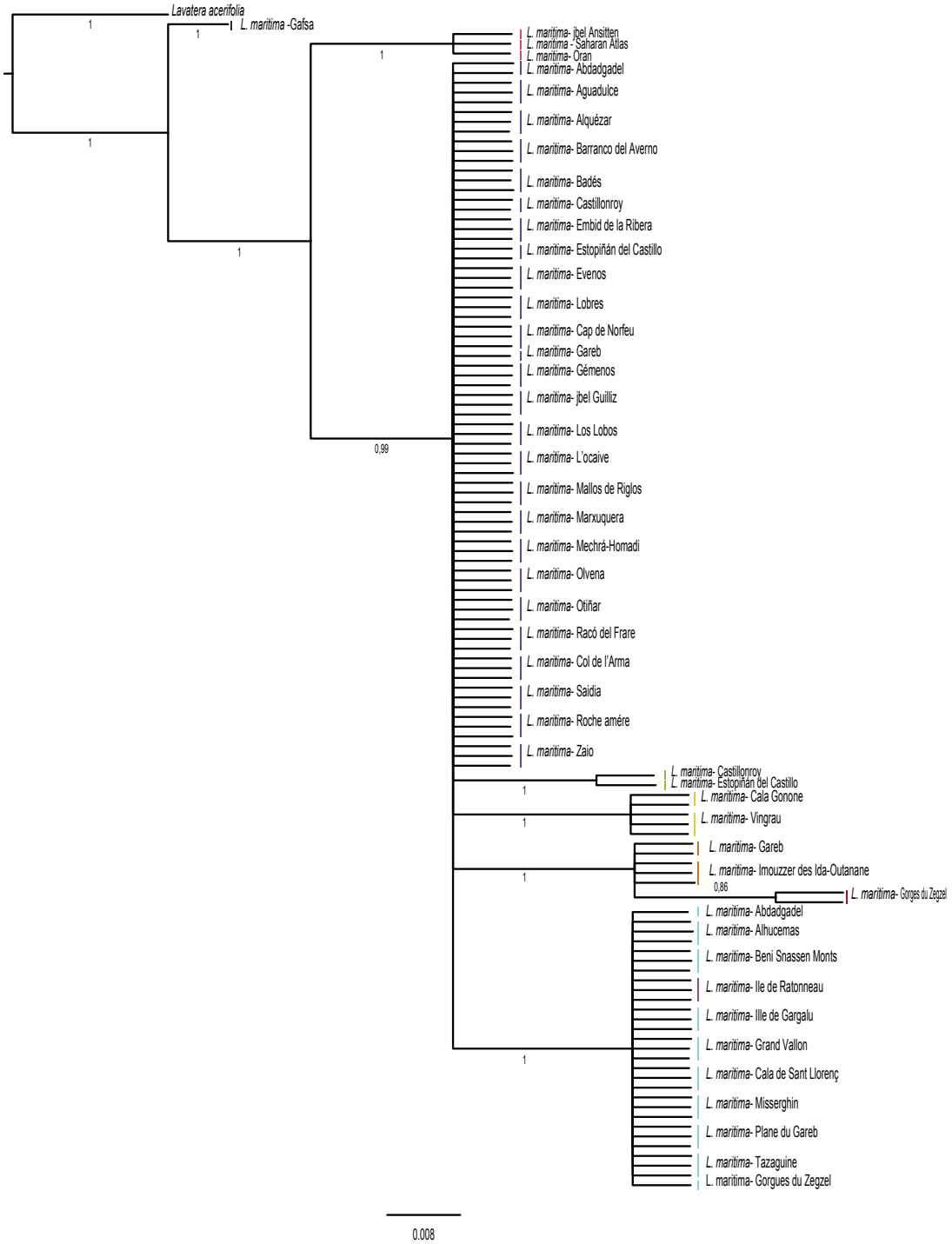


Figure S1. Fifty percent majority rule Bayesian consensus tree. Values below branches represent Bayesian posterior probability (PP). Branches without PP have values higher than 0.99. Color codes in vertical bars identify haplotypes following those in Fig. 2. Species and population names are indicated.

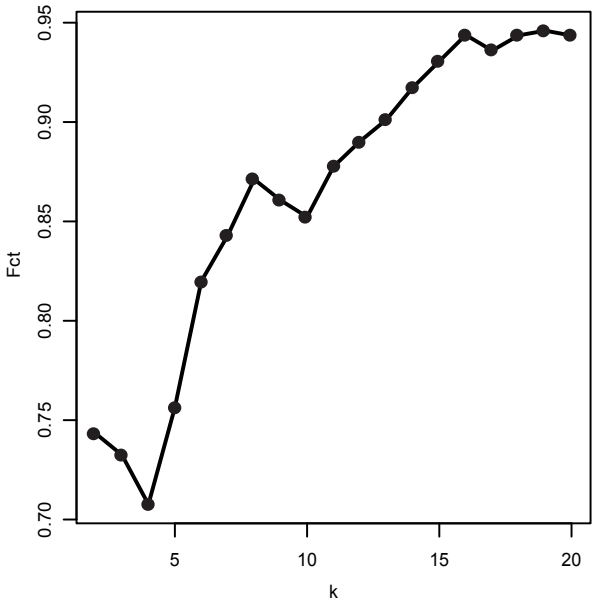


Figure S2. Spatial analysis of molecular variance (SAMOVA). Correlation between the number of groups (K) and FCT using cpDNA sequences.

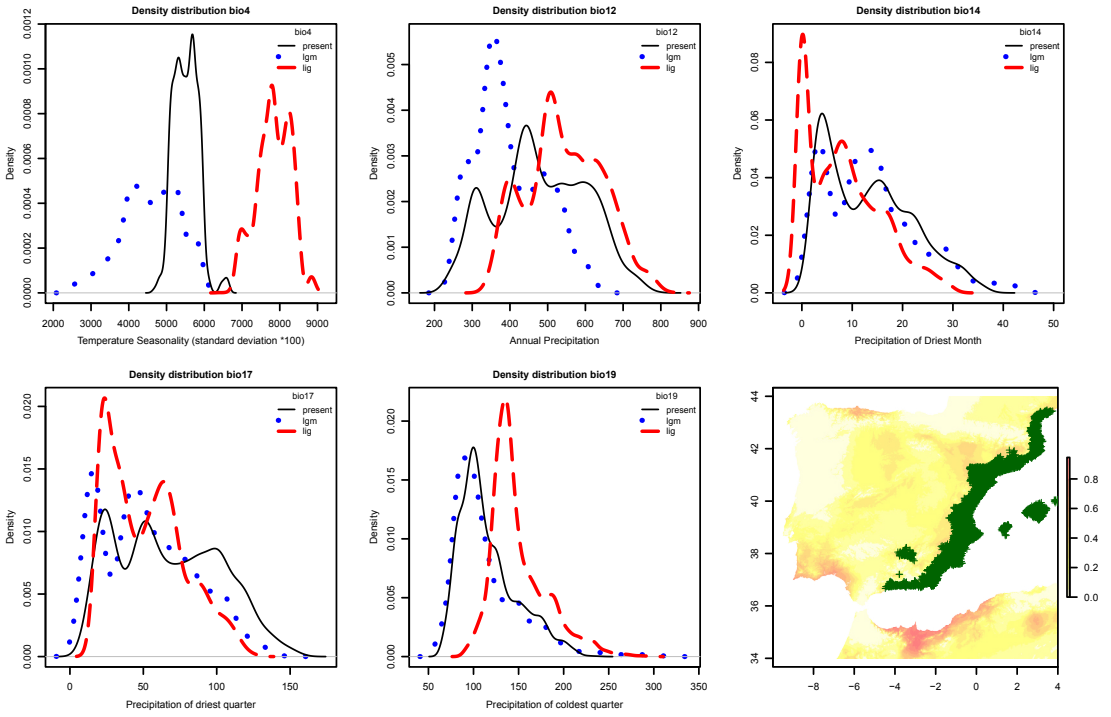


Figure S3. Probability density plots of environmental variables in the Spanish Mediterranean fringe. The five most important variables have been analyzed. Black line represents background obtained of the present time period, blue dotted line represents background of LGM and red dashed line shows background of LIG.

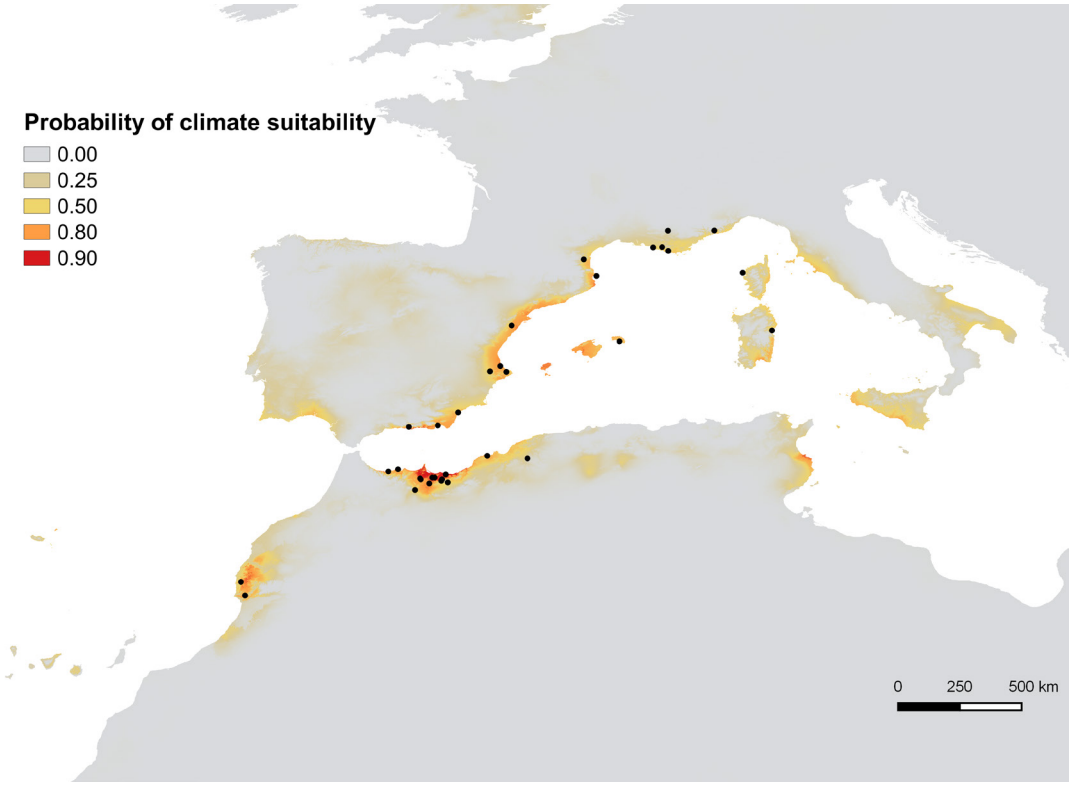


Figure S4. Result of a climate suitability model for *Lavatera maritima* without 10 inland populations. Dots indicate sampled populations for this study.

CAPÍTULO 2

Genotyping SNPs of hexaploid genomes: GBS data and niche modeling unveils the colonization history of the Canarian endemic shrub *Lavatera acerifolia* (Malvaceae)

Villa-Machío, I., Fernández de Castro, A. G., Fuertes-Aguilar, J., Nieto Feliner, G.
Submitted to *Molecular Ecology*

ABSTRACT

The Canarian archipelago is one of the two most studied oceanic areas from a biogeographical standpoint. Topics such as adaptive radiation, successful colonization events per lineage, insular woodiness, and phylogenetic relationships with mainland lineages, have been addressed. However, there is a scarcity of studies using a substantial representation of the genome, which has sometimes precluded a fine-scale reconstruction of recent evolutionary events. NGS-associated techniques have begun to alleviate this data shortage and GBS has proved a useful technique. Nonetheless, applying this method to polyploid genomes has been a challenge. We performed a phylogeographic study of the hexaploid Canarian endemic *Lavatera acerifolia* using SNPs and species distribution modelling. SNPs were identified from GBS data comparing different parameters using three bioinformatics approaches. The aims were to unveil the colonization history of this species and assess whether it fits the classical island biogeography models. Genetic groups fitting an east-west pattern were identified by Bayesian clustering methods, coalescent-based tree analyses and GLM estimations of the influence of proximity to the mainland on heterozygosity identify. Results are consistent with the palaeo-island hypothesis on the origin of diversity in Tenerife, which gathers the highest number of groups. Since the suitability per island increases with distance to the mainland, the occurrence of only one population in the western islands (Gomera, La Palma) suggests that colonization front is taking place. We find that the utility of a non-closely related true reference genome for SNP identification is questionable, while the use of mock-reference genomes appears as a reliable signal.

INTRODUCTION

Polyploidy, or whole genome duplication (WGD), is a fundamental evolutionary mechanism in plants. It contributes to diversity by providing new versatile genetic material for evolution, particularly when it involves hybridization (allopolyploidy) and thus gathers differentiated genomes (Soltis *et al.* 2014; Soltis & Soltis 2009). Because of the recurrence of WGD events, followed by concomitant genomic changes and adjustments, evolution of polyploid lineages can be seen as a succession of cycles (Wendel 2015). As a result, unraveling the detailed evolutionary history of WGD lineages is challenging. The development of restriction site-associated DNA sequencing approaches (RADseq), using Next Generation Sequencing (NGS) techniques, provides an opportunity to discover a large amount of single nucleotide polymorphisms (SNPs). These approaches allow the genotyping of non-model organisms for ecological, evolutionary and phylogenetic studies (Andrews *et al.* 2016). However, in polyploid genomes, detecting true polymorphisms from short-read data and alternative alleles at a single locus, while avoiding paralogous loci, is also challenging.

Lavatera acerifolia Cav. (Malvaceae) is a hexaploid (2n=44, 6x, Escobar *et al.* 2009) species with a small genome size (3,582 pg., Escobar 2007). This species belongs to the *Malva* alliance in which all but one species are polyploid. The most frequent ploidy level is hexaploid but there are species up to 16-ploid (Devesa Alcaraz & Luque 1986). *Lavatera acerifolia*, one of the two endemic species in the Canary Islands from the *Malva* lineage, is sister to the western Mediterranean *Lavatera maritima* Gouan, also a hexaploid (2n=44, Escobar *et al.* 2009), which reaches southwestern Morocco. To date, there are no explicit hypotheses on how many times and when polyploidy has occurred in the *Malva* alliance (Escobar *et al.* 2009). However, the fact that both *L. acerifolia* and the lineage sister to the *L. maritima-L. acerifolia* clade are hexaploids reinforces the idea that the most recent WGD event is not associated with the speciation event splitting *L. acerifolia* and *L. maritima*. If this is the case, one could expect that, in *L. acerifolia*, disomic inheritance was acquired some time ago at the meiotic level such that only chromosomes coming from the same parental origin pair and thus gametes are fertile. However, this does not imply diploidization at the gene level and thus, being hexaploid, *L. acerifolia*

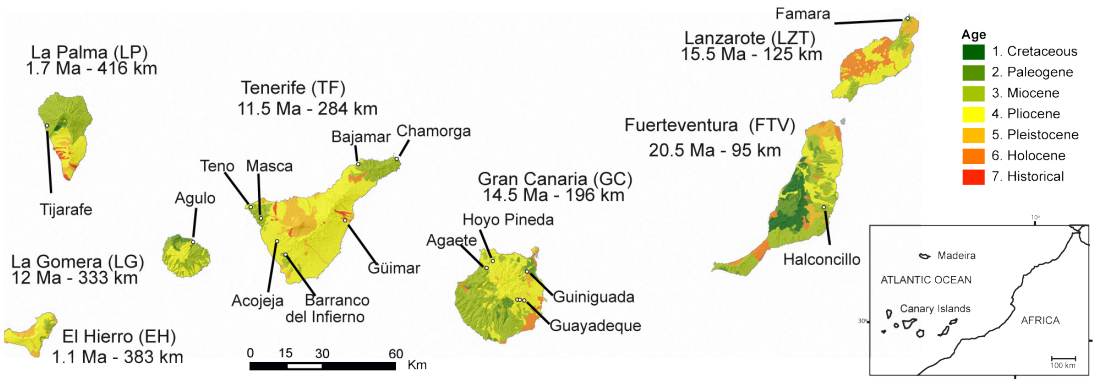


Figure 1. Geographical location of the sampled populations of the Canarian endemism *Lavatera acerifolia*. The oldest recorded age is given for each island along with the shortest distance to the closest point on the continent (Cape Juby, Tarfaya, Morocco). The color coding reflects the geological period of the different units within each island.

is expected to have large numbers of homologous loci rendering paralogous sequences, which should be filtered when searching for SNPs.

The Canary archipelago is formed by seven islands and several islets that are aligned from east to west and have emerged sequentially within the last 20 Ma. Thus, the oldest island is the easternmost (Fuerteventura) and the youngest is the westernmost (El Hierro, Fig. 1, Fernández-Palacios *et al.* 2011). *Lavatera acerifolia* occurs on all islands of the archipelago except El Hierro. Since islands allow genetic groups to be related to discrete geographic areas, the Canary archipelago has been subjected to numerous studies on evolutionary processes and phylogeography of plants and animals (Husemann *et al.* 2014; Valtueña *et al.* 2016). The well-known geological history of the islands is an additional asset for investigating the origin and timing of colonization. Two major sources of colonization have been targeted for the Canary Islands biota: neighboring North Africa and the Mediterranean basin (Juan *et al.* 2000). The most frequent pattern found by molecular phylogenetic analyses of different plant and animal species is a stepwise colonization from the eastern-oldest islands to the western-youngest ones (Hess *et al.* 2000; Juan *et al.* 2000; Talavera *et al.* 2013; Thorpe *et al.* 1994). This colonization pattern is associated with a decrease in genetic diversity within populations according to an increasing distance to the mainland due to founder events as well as dispersal limitations from the source areas (García-Verdugo *et al.* 2015; Yamada & Maki 2012).

While molecular markers have been extensively used to assess patterns of genetic structure and colonization in oceanic islands, only a few studies have employed the discovery of SNPs from Next Generation Sequencing (NGS) techniques for elucidating the population structure of island endemics. Among all RADseq methods available (RADseq, GBS, CRoPS, 2bRAD, RRLs and ddRAD among others, Andrews *et al.* 2016), we chose Genotyping-by-Sequencing (GBS) to detect and recover SNPs in *L. acerifolia.* This approach generates a reduced representation library of the genome with restriction enzymes offering several advantages for this study. It is cost-effective in providing a considerable amount of data (Elshire *et al.* 2011), has been previously used in other polyploid species, and has been reported to be a suitable technique for generating SNPs (McAllister & Miller 2016; Qi *et al.* 2015; Tyler *et al.* 2016). Despite this, there remains insufficient empirical knowledge on how to design the best filtering of the raw data to generate reliable SNPs in polyploid genomes.

The aims of this paper are twofold. On the one hand, we aim to assess and refine the most reliable bioinformatics filtering approaches for a polyploid organism, potentially presenting substantial amounts of paralogy, a serious but insufficiently addressed problem when using NGS techniques (Limborg *et al.* 2016). On the other hand, we focus on how the evolutionary history of an oceanic island endemic species, and particularly its colonization pattern, fits the general island theory, using a genomic high-throughput technique capable of producing high amounts of SNPs combined with species distribution modelling (SDM) projected onto past conditions.

The specific goals of this study are: i) developing SNPs from GBS data in populations of *L. acerifolia*, ii) assessing the reliability of SNPs discovered using three different available bioinformatics approaches, namely *de novo* assembly, by using a reference genome, and by constructing a "mock-reference" genome, iii) estimating the genetic structure of this endemic species, iv) assessing the genetic diversity patterns along its geographic range, v) elucidating the migration routes within the Canary archipelago, vi) constructing a SDM for this species, and vii) exploring how all the evidence fits the general island theory.

MATERIAL AND METHODS

Plant material and sampling

A total of 15 populations spanning the natural distribution of *L. acerifolia* in the Canary Islands (Lanzarote, Fuerteventura, Gran Canaria, Tenerife, La Gomera and La Palma) was sampled. All populations were georeferenced during the fieldwork (Table 1) and leaves collected and dried in silica gel.

Table 1. Origin of the populations of *Lavatera acerifolia* sampled in this study.

Island	Locality	Code	Colector	Latitude	Longitude
Fuerteventura	Morro del Halconcillo	HAL	S. Scholz, I. Villa	28.35689	-13.92653
Gran Canaria	Vecindad de Enfrente (Agaete)	AGA	G. Nieto, I. Villa	28.08228	-15.67106
Gran Canaria	Barranco Guiniguada	GUIN	G. Nieto, I. Villa	28.06603	-15.46278
Gran Canaria	Barranco de Guayadeque 1	GUA	G. Nieto, I. Villa	27.93675	-15.51142
Gran Canaria	Barranco de Guayadeque 2	GUA	G. Nieto, I. Villa	27.93597	-15.49903
Gran Canaria	Barranco de Guayadeque 3	GUA	G. Nieto, I. Villa	27.93353	-15.47608
Gran Canaria	Hoyo de Pineda (Barranco Anzo)	HP	G. Nieto, I. Villa	28.11314	-15.63947
La Gomera	Agulo	AGU	J. Fuertes, A. Gonzalez	28.18478	-17.19017
La Palma	Barranco Jorado (Tijarafe)	TIJ	J. Fuertes, A. Gonzalez	28.70344	-17.96049
Lanzarote	Barranco Famara	FAM	Jose D. Naranjo	29.21831	-13.47834
Tenerife	Barranco de los Infiernos	INF	J. Fuertes, A. Gonzalez	28.13375	-16.71158
Tenerife	Masca	MAS	J. Fuertes, A. Gonzalez	28.29856	-16.84128
Tenerife	Barranco Guaria (Acojeja)	ACO	J. Fuertes, A. Gonzalez	28.19489	-16.75581
Tenerife	Güimar	GI	J. Fuertes, A. Gonzalez	28.29383	-16.40582
Tenerife	Bajamar	BAJ	J. Fuertes, A. Gonzalez	28.55209	-16.34079
Tenerife	Chamorga	CHA	J. Fuertes, A. Gonzalez	28.57828	-16.14081
Tenerife	Punta de Teno (Buenavista)	TENO	J. Fuertes, A. Gonzalez	28.34908	-16.89486

Genotyping-by-sequencing

Genotyping-by-sequencing data were generated for 28 individuals of *L. acerifolia* from different populations, plus one individual of its sister species, *Lavatera maritima*. To cover a wide and even geographic representation of the species range, we only sampled two individuals per population except for Bajamar and Famara, each of which was sampled with a single individual. Total DNA was extracted from leaves using DNeasy Plant Minikits (QIAGEN Inc., California) and concentrated using a precipitation protocol described in Sambrook *et al.* (1989). After quantification by Qubit fluorimeter and quality control with gel electrophoresis, DNA samples were processed to obtain pair-end GBS libraries according to Elshire *et al.* (2011) with the following modifications: 100 ng of DNA from each sample were used for restriction digestion with PstI (New England Biolabs), and the adapters were ligated to the DNA fragments introducing an inline index

with the barcoded adapters. Paired-end (2x101 bp) Illumina sequencing was performed on a HiSeq2000 platform at Centro Nacional de Análisis Genómico (CNAG, Barcelona). Demultiplexed reads were visualized with FastQC v.0.11.5 (Andrews 2010) for quality control, and edited and filtered with Trimmomatic v.0.36 (Bolger *et al.* 2014). These genetic data were processed for downstream bioinformatics analyses.

Bioinformatics analyses

To achieve a reliable and efficient SNP calling despite the difficulties related to paralogy in this polyploid species (2n=44, 6x) for which there is no close reference genome available, three different approaches were used in working with paired-end data: (1) a *de novo* assembly, a commonly used procedure when a reference genome is absent (Elshire *et al.* 2011; Escudero *et al.* 2014; Mastretta-Yanes *et al.* 2015), (2) an assembly with a reduced reference sequence generated from our own data from *L. acerifolia*, known as a mock reference (Melo *et al.* 2016) and (3) an assembly using a reference genome of the closest relative available (Penjor *et al.* 2014; Schröder *et al.* 2016). All analyses were performed at the SGAI-CSIC (cluster Trueno).

Currently, there are several software packages available for analyzing *de novo* and assembled GBS data. GBS-SNP-CROP (Melo *et al.* 2016) is one of the few pipelines designed to discover SNPs that considers polyploid species without a close relative reference genome. PyRAD (Eaton 2014), along with TASSEL-UNEAK (Lu *et al.* 2013) and Stacks (Catchen *et al.* 2011), are some of the most commonly used pipelines for *de novo* assembly. However, the latter two present drawbacks, such as their ability to analyze only single-ends (TASSEL-UNEAK) and the computational limitations due to the requirement of considerable amounts of memory that substantially increases computational time (Stacks). We thus employed PyRAD, a pipeline designed to cluster samples based on similarity, because it is easy to use and it allowed us to develop the analysis employing paired-end data. The *de novo* method implemented in PyRAD has been used largely for phylogenetic studies comprising different species that share low levels of similarity. However, PyRAD can also be applied in population-level studies because it allows the specification of a high clustering threshold. By contrast, the other software we used, GBS-SNP-CROP, is a reference-optional method created for paired-end genotyping-

by-sequencing data. The pipeline was originally developed with a polyploid organism (*Actinidia arguta*) and it offers the great advantage of building your own reference genome, a useful option when working with non-model organisms. These features make it a convenient candidate for analyzing our data from *L. acerifolia*.

The presence of a hexaploid genome with three putative sets of genes requires additional steps in order to discriminate among paralogs. To account for the potential paralogy, two crucial aspects were considered in all workflows: the depth of sequence coverage used in the SNP calling process (Andrews *et al.* 2016; Clevenger *et al.* 2015) and the degree of variation detected as indicative of a possible mixture of paralogs in the same loci. Accordingly, our data were subjected to stringent parameters such as restricting the search only to potential biallelic SNPs, high average read depths, both combined with different clustering thresholds (80, 85, 90, 95) as detailed in Appendix S1 where the three workflows are described.

Genetic structure analysis and relatedness

The genetic structure of *L. acerifolia* was explored under different methodological approaches. To infer population structure, two independent Bayesian model-based approaches were implemented: Bayesian Analysis of Population Structure (BAPS v.6.0, Corander *et al.* 2013) and STRUCTURE v.2.3.4 (Pritchard *et al.* 2000). In BAPS, a mixture analysis was conducted to estimate the number of genetically diverse populations, the maximum being 15. Then, an admixture analysis was carried out with a minimum population size of 2, the number of iterations for estimating the admixture coefficient for the individuals was set to 100, the number of reference individuals from each population was 200 and 20 iterations were used for estimating the admixture coefficient for the reference individuals. For the *de novo* approach, five matrices with different levels of missing data were used (m indicating the minimum number of samples with data for a particular locus to be included in the final dataset): novo-c90m11, novo-c90m14, novo-c90m17, novo-c90m21 and novo-c90m25. The results from all of them were generally congruent (Table 2). However, some differences were observed with respect to the number of SNPs obtained, which decreased as the amount of missing data diminished. Since a compromise between these two inversely proportional parameters had to be reached,

the matrix chosen for subsequent intraspecific population structure analyses from the *de novo* workflow required SNPs to be present in 75% of the samples (novo-c90m21). The same procedure was performed with the five matrices using the mock reference approach (mock3, mock4, mock5, mock20x, mock50x). Mock3, mock4 and mock5 showed similar results regarding population structure, but differed in the number of SNPs recovered. For comparative purposes, we selected mock3 for two reasons: first, to match this parameter with *de novo* workflow 1 (maximum number of SNPs allowed in a consensus sequence) and second, because we were examining a hexaploid genome with three putative sets of genes. This mock3 matrix was subjected to the next filter for the average read depth (20, 50). The results were also congruent, but they showed differences with respect to the number of SNPs, which were higher in the mock20x matrix. We estimated that an average read depth of 20 is enough coverage for working with a polyploid genome. Consequently, the mock20x, constructed with an average read depth of 20 and a maximum of 3 SNPs per cluster, was selected for subsequent analyses. In summary, novo-c90m21, from the *de novo* approach, and mock20x, from the mock reference approach, were the two selected datasets for the genetic and phylogeographic analyses in our study. Both matrices were analyzed in STRUCTURE for estimating the genetic groups under an admixture model with allele frequencies correlated among populations. Each run consisted of 10^6 replicates with a burn-in period of 10^5. Ten replicates were carried out for each k value (from 1 to 16, the number of populations plus 1). The number of genetic clusters, i.e., the optimal partition of the genetic dataset, was estimated applying the Evanno criterion (Evanno *et al.* 2005) implemented in Structure Harvester (Earl 2012) and the multiple iterations of each K were combined in the online application CLUMPAK (http://clumpak.tau.ac.il/). The overall distribution of genetic variation of both datasets was evaluated by a principal component analysis (PCA) using the package adegenet (Jombart & Ahmed 2011) in the R environment (R Core Team 2015). In addition, to examine the effect of non-natural admixture, we replicate the STRUCTURE analysis without the individual from the Guiniguada population, which had likely undergone introgression from cultivated accessions from other populations (see discussion).

Further, to infer the genetic relationships among populations, we used these same matrices obtained from workflows 1 and 2, with the inclusion of *L. maritima* as an outgroup. A species tree was estimated with the SVDquartets method (Chifman & Kubatko 2014) implemented in PAUP v4 (Swofford 2003). This method infers the topology among randomly sampled quartets of taxa under the coalescent model. All possible random quartets were sampled with 1000 replicates of nonparametric bootstrapping to measure the uncertainty in the relationships.

Island genetic diversity estimates

Standard genetic diversity indices were estimated with the PopGenome package (Pfeifer *et al.* 2014) in the R environment (R Core Team 2015). Nucleotide diversity (π) and the fixation index (F_{st}) within and between groups were calculated considering different groupings: populations, islands and distance classes grouped along geographic longitude. In the case of populations, Bajamar and Famara samples were discarded since they only hold one individual. The third grouping consists of four groups in which the populations from the easternmost (Famara and Halconcillo) and westernmost (Tijarafe and Agulo) islands were clustered together since they present a limited number of populations.

In order to assess the influence of environmental and geographic factors on the genetic diversity across the whole range of the species, a generalized linear model (GLM) was performed in the R environment. This analysis estimates the effect of the distance to the mainland, the relative topoclimatic suitability (see next section) of the niche and the geological substrate age on the proportion of heterozygous sites present in the genome of each population whose values were calculated using vcflib (Garrison 2012). The response variable was the amount of heterozygous sites per population, while the predictor variables were the distance to the mainland, the relative topoclimatic suitability and the geological substrate age. The relative topoclimatic suitability was calculated as follows: the total area of the archipelago was divided applying the Voronoi tessellation using population coordinates to calculate the polygons, such that each population was limited by one polygon. Then, relative suitability attributed to each population was calculated as the ratio between suitable area and total area for each polygon corresponding to an island. Several models of GLM were conducted with different combinations of predictor

variables. Finally, the model with the lowest Akaike information criterion (AIC) score was selected (MuMIn package in R, Barton 2013).

Species distribution model (SDM)

Development of bioclimatic variables

The reduced scale of the distribution of insular species requires that SDM be performed at a higher resolution than in widespread continental species to better reflect local ecological factors (Austin & Van Niel 2011; Lassueur *et al.* 2006). Therefore, to generate an accurate model of the environmental distribution of *L. acerifolia*, we developed a set of spatial climate layers at a 50-meter resolution, based on the network of meteorological stations of the archipelago (data provided by AEMET, www.aemet.es). Only stations with 10 or more years of climate records were considered. For the monthly variables of minimum and maximum temperature and precipitation, we developed a stepwise generalized additive model (GAM) with altitude, northness, latitude and longitude as predictor variables. Models were selected based on AIC scores and then projected to the whole archipelago, including El Hierro despite the absence of *L. acerifolia* on this island. To account for spatial biases of the models, residuals of each model in each meteorological station were used to develop an interpolated map of residuals for each variable by kriging. This interpolated layer was added to the predicted value of the GAM model to obtain the final layers of each monthly variable. The final dataset of monthly variables was used to develop the bioclimatic variables described by Hijmans *et al.* (2005), using the dismo package in R (Hijmans *et al.* 2015). To assess the importance of topography, we also incorporated two topographic predictors: slope and topographic index (TPI), derived from the digital elevation model (DEM) of the archipelago. We also derived these climate variables for past conditions by using the climate anomalies developed for the MIROC model for the mid-Holocene (6 kya) and Last Glacial Maximum (22 kya). Although climate anomalies have also been projected to the Last Interglacial Maximum (120-140 kya), we decided not to use this period because, due to the dynamic nature of island geology, the topography of the archipelago and coast line in that period differed considerably from the current one. We downscaled the climate anomaly to a 50-meter

resolution for the study area using the Delta method. The resulting anomaly for each monthly variable was summed to the variables calculated for the present.

Model calibration

With our 50-meter working resolution, 34 presence cells were recorded in total. To calibrate niche models, we used biomod2 (Thuiller *et al.* 2009) for which wee developed 5 datasets containing presence points and 200 random generated pseudoabsence cells, which were weighted to account for the same importance as presences.

To select climatic predictor variables to model the topoclimatic niche we conducted a correlation analysis for the values of the variables in the cells where the species was present with the R package ecospat (Broennimann *et al.* 2014), which returned a recommended value of 8 predictor variables. To select those 8 variables, we first performed a PCA in which 11 predictors obtained the highest scores along the first three axes. These were nine bioclimatic variables (1, 2, 5, 6, 8, 10, 11, 15, 17) plus TPI and slope. Finally, we conducted a hierarchical partitioning approach (Chevan & Sutherland 1991), as implemented in the R package hier.part (Walsh & Mac Nally 2013), to select among those 11 variables, the 8 showing the highest independent contributions (TPI, slope, bio1: mean annual temperature, bio6: minimum temperature of the coldest month, bio7:temperature annual range, bio8: mean temperature of the wettest quarter, bio11: mean temperature of the coldest quarter, and bio15: precipitation seasonality).

We used five algorithms available in biomod2 for niche modeling: GLM and GBM with stepwise selection, MARS, ANN, and RF. For each of the algorithms chosen, we conducted 10 runs of each presence-pseudoabsence dataset. In each run, 85% of data was randomly selected for calibration and the rest for model evaluation. To evaluate the models, we used TSS and ROC scores. Models with scores below 0.8 for any of the two criteria were excluded. The remaining models were retained to build an ensemble model based on the contribution of each individual model, which was weighted according to the TSS score. Finally, this ensemble model was projected to the past climate conditions developed for the archipelago.

RESULTS

GBS output

Sequencing of twenty-eight individuals of *L.acerifolia* from 15 populations generated a total of 25 201 642 identified reads and an average of 900 058 reads per sample. The results of each pipeline using different parameters are shown in Table 2.

The selected matrix from the *de novo* assembly (novo-c90m21) produced 1566 loci (a minimum of 1514 from Guayadeque and a maximum of 1553 from Agulo) including a total of 2099 SNPs (717 unlinked). In the case of the selected matrix from the mock reference pipeline (mock20x), the number of mapped alignments, excluding those marked as secondary or supplementary, varied from 2 054 836 in Halconcillo to 345 233 in Chamorga. This matrix consisted of 1140 clusters with 1485 bi-allelic and independent SNPs. We obtained an average value of recovered SNPs per sample of 1111 (57-1485). The highest percentage of missing data was found in individuals from Chamorga and Güimar. The GBS-SNP-CROP pipeline using the reference genome obtained a number of mapped alignments from *L. acerifolia* to *G. arboreum* that ranged from 1 367 384 (Halconcillo) to 226 895 (Chamorga). After applying the filters recommended by the authors (Appendix S1) for working with hexaploid genomes, the analysis resulted in a

Table 2. Number of SNPs obtained in *Lavatera acerifolia* through the processing of genotyping-by-sequence (GBS) data under different parameters with two bioinformatics workflows. For the *de novo* assembly without a reference genome (workflow 1), only differences based on the minimum number of samples showing a SNP (and associated % of missing data) are shown. For the assembly using a mock reference genome (workflow 2), the number of SNPs under different combinations of average read depth and number of SNPs per cluster are shown.

Workflow 1 - *De novo*	**Minimum number of samples represented in a final locus (% of samples)**	**Loci**	**Unlinked SNPs**	**total SNPs**	
novo-c90m25	25 (89.29%)	1476	675	1950	
novo-c90m21	21 (75%)	1566	717	2099	
novo-c90m17	17 (60.72%)	1650	769	2330	
novo-c90m14	14 (50%)	1709	804	2504	
novo-c90m11	11 (39.29%)	1782	846	2722	
Workflow 2 - Mock reference	**Minimum number of samples represented in a final SNP (% of samples)**	**Clusters**		**SNPs (biallelic positions)**	**Average read depth**
Number of SNPs / cluster					
up to 7	21 (75%)	1233		1687	4
mock5	21 (75%)	1231		1673	4
mock4	21 (75%)	1262		1643	4
mock3	21 (75%)	1205		1563	4
Average read depth					
mock10x	21 (75%)	1205		1563	10
mock20x	21 (75%)	1140		1485	20
mock50x	21 (75%)	668		870	50

very low number of recovered SNPs (164). For this reason, this workflow 3 was excluded in posterior analyses.

The downstream analyses were conducted on both datasets (novo-c90m21 and mock20x, Table 2), but we only present results based on the mock reference pipeline (mock20x matrix). The results from the *de novo* workflow 1 (novo-c90m21 matrix) are explained in the discussion section (see below).

Population genetic structure and relatedness

The 1485 retained SNPs from the mock reference pipeline (workflow2) were used for elucidating the genetic structure of *L. acerifolia*. The genetic variation was visualized with a PCA (Fig. S2a, b), which revealed four distinct geographic groups, i.e., eastern (E group: Famara, Halconcillo), central-eastern (CE group: Chamorga, Güimar, Agaete, Hoyo Pineda, Guayadeque), central-western (CW group: Teno, Bajamar, Guiniguada) and western populations (W group: Tijarafe, Agulo, Infierno, Acojeja, Masca), distributed along a longitudinal geographic gradient. The variance accumulated in the first three principal components explained 48.24% of the total genetic variance. Similarly, the BAPS analysis identified the same four genetic clusters (Fig. 2a) and only one individual from the Guiniguada population (CW) exhibited admixture from the CE group. STRUCTURE recognized an optimum number (K) of two groups (E+CE vs. CW+E), with populations from Teno, Bajamar, Guiniguada, Guayadeque and Famara exhibiting admixture (Fig. 2a). The second most likely partition for STRUCTURE was K=4 and the four groups detected coincided with those identified by BAPS and the PCA (E, CE, CW, W), but 11 of the 15 populations displayed admixture (Fig. 2a).

The topology of the coalescent-based tree inferred by the SVDquartets method is consistent with the genetic groups identified by BAPS and STRUCTURE (Fig. 2a). The eastern group (86% BS) is sister to the remaining populations, which are split into two clades of similar size. The first one includes the CE group (91% BS) sister to Guiniguada (without support). The second clade includes the W group (100%BS) sister to two of the populations from the CW group (Teno and Bajamar, 86% BS).

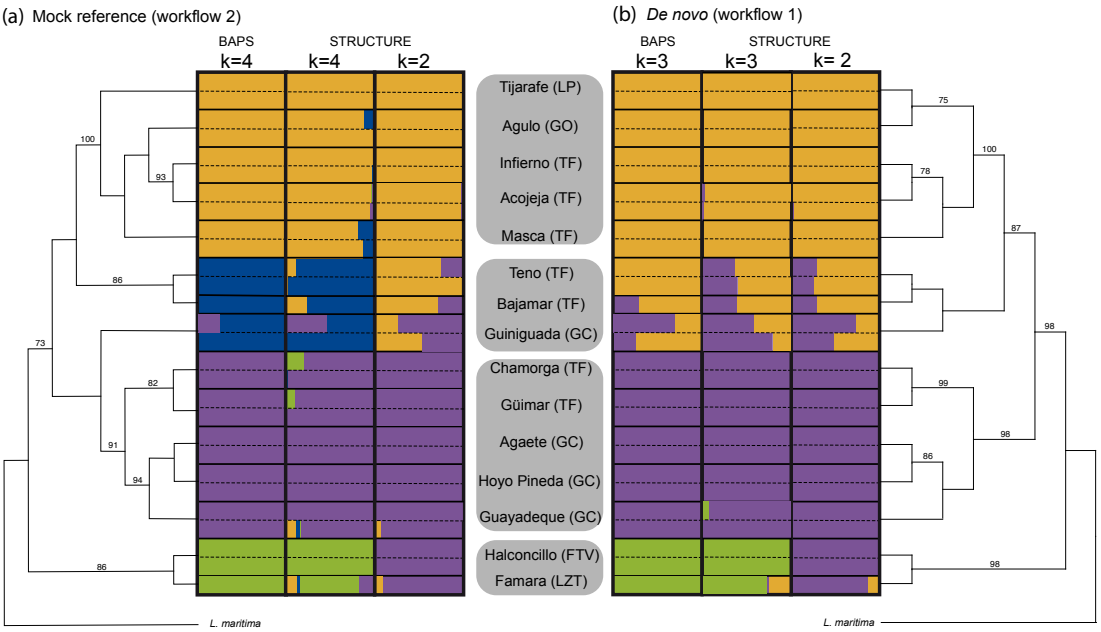


Figure 2. Bayesian analysis of genetic groups and coalescence-based tree of 15 populations from *Lavatera acerifolia* based on SNPs identified from genotyping-by-sequence (GBS) data generated using two bioinformatics workflows: mock reference genome (a) and *de novo* assembly without a reference genome (b). Species tree showing topological relationships among populations constructed with SVDquartets using *L. maritima* as an outgroup; bootstrap support values ≥ 75% are shown on the branches. Bayesian clustering of populations performed using two algorithms (STRUCTURE, BAPS) for each of the two datasets. Samples are represented by rectangles where the color indicates the probability of each sample belonging to each of the genetic groups. Two groups (K=2) is the optimal partition according to STRUCTURE for the datasets obtained under the two bioinformatics workflows (a and b). The second most likely partition is K=4 for the mock reference dataset and K=3 for the *de novo* dataset, coinciding with the partition identified to BAPS. Populations sampled are arranged according to proximity to the mainland, from east (bottom) to west (top) with shades indicating four genetic groups (K=4): western, central-western, central-eastern, eastern.

Genetic diversity

The distribution of nucleotide diversity among different islands and geographic groups showed a decreasing pattern from eastern to western islands (Table S1b). The analysis of nucleotide diversity per population showed that the samples from Guayadeque (GC, central-eastern group) exhibited the highest value of nucleotide diversity, whilst the lowest was found in Tijarafe and Masca (western group). With respect to the islands, an increase in nucleotide diversity was observed from La Palma to Lanzarote. However, Fuerteventura was an exception to this east-west trend since it only harbors two threatened populations containing a few living individuals. The same pattern for the levels of nucleotide diversity was found when examining classes grouped along geographic longitude.

The F_{st} index, measuring the degree of genetic differentiation within and between different groups, ranged from 0.103 (Agulo) to 0.210 (Masca; Table S1b) in the population grouping. The pairwise comparison among populations showed that the highest differentiation occurred between Tijarafe and Güimar, in Tenerife (0.312) and the lowest between Hoyo Pineda and Guayadeque, in Gran Canaria (Table S2b). Within islands, the F_{st} value varied between 0.092 (Tenerife) and 0.190 (La Palma; Table S1b). Between islands, the La Palma-Fuerteventura pair showed the highest differentiation (0.268) while La Gomera-Tenerife showed the lowest (0.008; Table S2b). For the grouping of distance classes along geographic longitude, the highest F_{st} value was found in the isolated populations from the Eastern islands (Famara and Halconcillo, 0.147) and the lowest in the CW group (Tenerife, 0.081; Table S1b). The W (Tijarafe and Agulo) and E (Famara and Halconcillo) groups showed the highest genetic differentiation (0.167; Table S2b).

The proportion of heterozygous sites in the genome of each population ranged from 0.14 to 0.30 (Fig. 3). For this analysis, we excluded the individuals from Guiniguada since we detected evidence of non-natural introgression in this population (see discussion).

For the analysis of the influence of external factors on the genetic diversity distribution, we compared heterozygosity with a set of geographical, geological and ecological variables relevant to island phylogeography: population distance to the continent, percent of suitable areas as defined by the SDM, and geological substrate age. The result of the best GLM model showed that the proportion of heterozygous sites

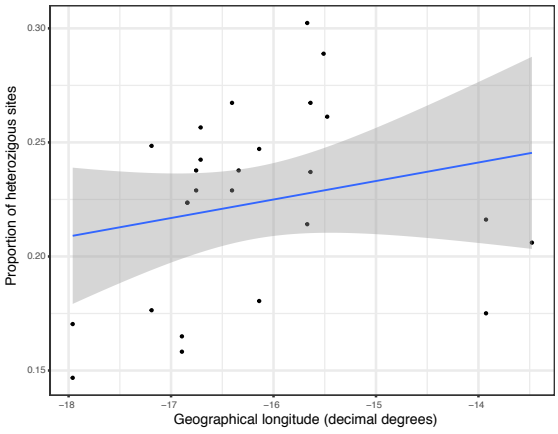


Figure 3. Proportion of heterozygous sites found in the genome of 26 individuals from 14 native populations of *Lavatera acerifolia* based on SNPs data identified through genotyping-by-sequence (GBS) using a mock reference genome (mock20x matrix; see text). The solid line indicates values of heterozygous sites predicted by a generalized linear model (GLM). The x axis indicates the minimum distance from each island to the mainland, measured in decimal degrees.

was negatively related to the distance of the populations to the mainland (estimated coefficient= -5.10e-04 ± 1.14e-04 SD, p<0.001). The analysis showed that the proportion of heterozygous sites was significantly related to the substrate age. However, the estimated coefficient (-1.74e-8 ± 4.25e-09 SD, p<0.001) was much smaller than the one for distance and consequently, we consider the relationship between the proportion of heterozygous sites and substrate age as negligible. The proportion of heterozygous sites was not related to the relative topographic suitability of the niche.

Species distribution modelling

More than 90% of the model runs (229 over 250) were retained for the ensemble model (Fig. S3). TSS scores ranged from 0.383 to 1 and ROC from 0.25 to 1. There were no significant differences in any of the two scores, either between presence-absence datasets or runs ($F_{4,48}$=1.445, P=0.233).

The ensemble model was explained mainly by slope with an independent contribution of 0.77. TPI showed a contribution of 0.076. The contribution to the model of climatic variables was limited, as it ranged from 0.05 for the mean temperature of the coldest quarter (biovariable 11) to 0.01 for the annual temperature range (bio7). The suitability threshold was 0.79 for ROC and 0.81 for the TSS score. The projection of the model assembled with the TSS score is shown in Fig 4a. In the eastern islands (Lanzarote and Fuerteventura), suitable areas were restricted to only a few steep areas. The central and western islands (Gran Canaria, Tenerife, La Palma, La Gomera and El Hierro) showed a higher proportion of suitable areas for *L. acerifolia*, which mainly corresponded to ravines running across wide altitudinal ranges. In Gran Canaria and Tenerife, the most suitable areas follow two different patterns (Fig. 4a). In Gran Canaria, suitability was higher in the western part of the island whereas in Tenerife, the amount of suitable areas was higher in the northeastern and northwestern parts, which are the oldest mountain massifs on the island, than in younger central areas. Suitability per island significantly increased with distance to the continent (GAM, $F_{1,440685}$=19660, P<0.001). The relative suitable surface per island plotted against distance to the continent (Fig. S4) showed that, considering Tenerife as three separate geological units instead of a single one, the potential area for *L. acerifolia* reached its maximum at the center part of the archipelago,

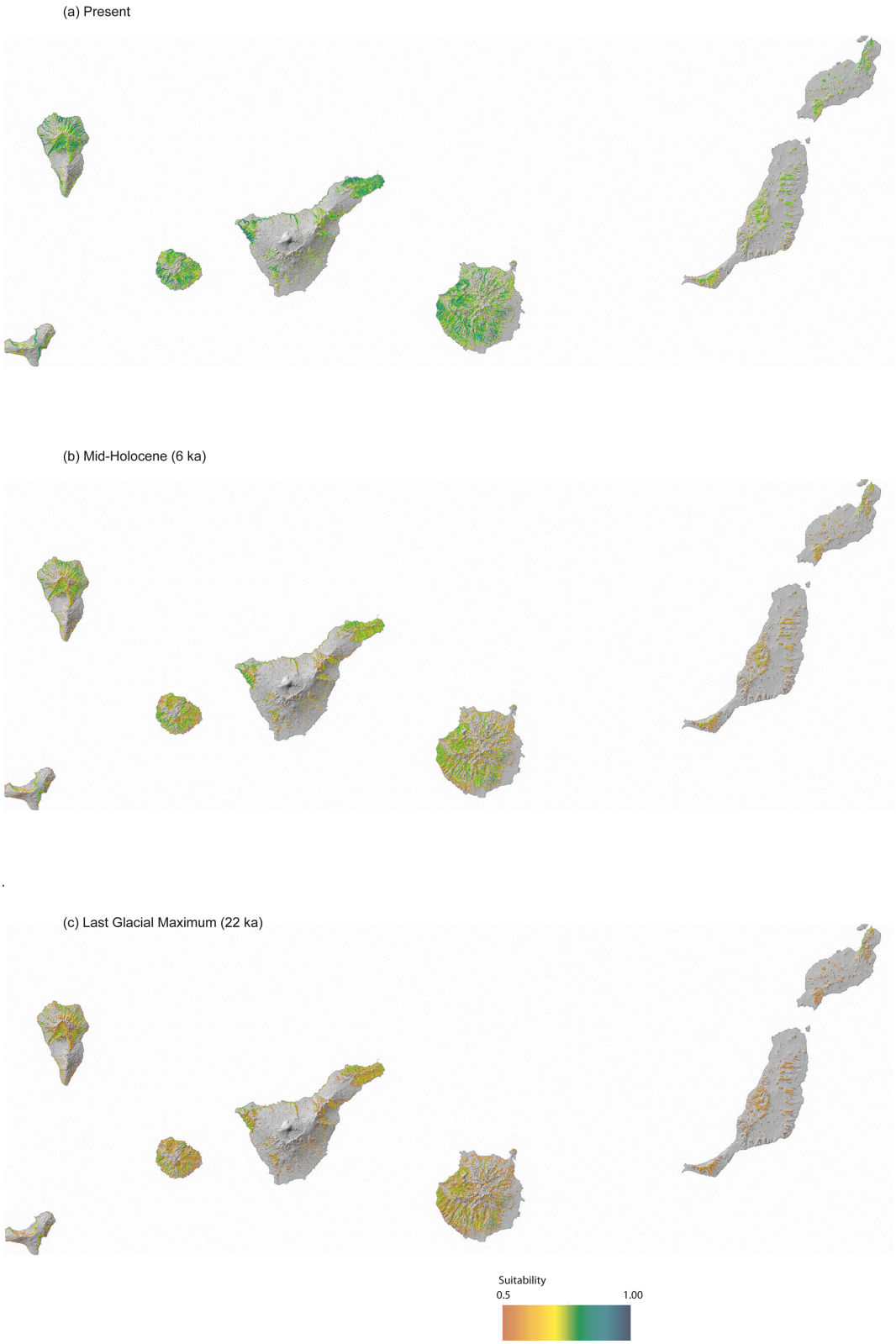


Figure 4. Niche suitability for an ensemble species distribution model of *Lavatera acerifolia* constructed using five algorithms and 7 variables, and based on the TSS evaluation score: a) current climate conditions; b) projection of the ensemble model for the mid-Holocene (6 ka) climate conditions; c) projection for the Last Glacial Maximum (22 ka). Values of niche suitability <0.5 are not colored. Presence-absence threshold is set up by the ensemble model in 0.65.

coinciding with the location of Tenerife. Similarly, topoclimatic suitability showed a significant relationship with age, peaking in geological units dating from the Miocene period (GAM, $F_{1,326}$=41.24,P=4.8*e^{-10}, Fig. S4).

The projected suitability of the MIROC climate model to the mid-Holocene and Last-Glacial Maximum followed a spatial pattern that is similar to the present (Fig. 4b, 4c). All the islands maintained potential suitable areas in both periods in the same regions and the altitudinal range of these potential areas was also conserved over time. However, the overall suitability decreased with age in both temporal scenarios.

DISCUSSION

The thousands of polymorphisms screened throughout genomes with GBS methods offer major advantages for genomic studies (Narum *et al.* 2013). In fact, a considerable number of articles has demonstrated the utility of GBS for species phylogenetic reconstruction (Wong *et al.* 2015) as well as for phylogeographic and population genetic studies (Nicotra *et al.* 2016; Pellegrino *et al.* 2016). As in other methods based on data generated through high-throughput sequencing techniques, the filtering process is fundamental to maximizing signal and minimizing noise, a task that is particularly challenging for polyploid species. For the discussion of the phylogeographic and biogeographic questions, we rely primarily on the results generated using filtering under workflow 2 (mock reference genome), but a comparison between workflows 1 and 2 is provided below at the end of this section.

Population structure in *L. acerifolia*

Our results detected a significant population genetic structure in *L. acerifolia*, consisting of four main genetic groups distributed across six of the seven main islands of the Canaries. Populations from Lanzarote and Fuerteventura form the eastern genetic cluster, which is strongly separated from the rest of the islands. Populations from Gran Canaria all belong to the central-eastern genetic cluster except for Guiniguada, which falls in the central-western group with two populations from northern Tenerife (Bajamar and Teno), but presents admixture in all analyses regardless of whether it forms a separate group with Bajamar and Teno (K=4) or not (K=2, 3, Fig. 2). The population from Guiniguada is likely

to have suffered introgression under cultivation since it is located where the Jardín Botánico Canario Viera y Clavijo was established in the 1950s, in which several other individuals collected on other islands including Tenerife have since been cultivated (J. Naranjo, JBCVC, comm. pers.). In fact, the exclusion of one individual from Guiniguada from the analyses (the one exhibiting the highest admixture) produced significant changes (see below). Tenerife shows the highest number of distinct genetic clusters: the northern populations (Teno and Bajamar) are part of the central-western genetic cluster, the eastern and north-eastern populations (Chamorga and Güimar) are part of the central-eastern genetic cluster along with Agaete, Hoyo Pineda and Guayadeque from Gran Canaria, and the western populations (Masca, Acojeja and Infierno) form the western genetic cluster along with populations from La Palma (Tijarafe) and Gomera (Agulo). The split of the Tenerife populations into three different clusters and the geographic distribution of these clusters is consistent with the palaeo-island hypothesis, which proposes that the geological origin of this island has a strong influence on the phylogeography of plant and animal species (Cox *et al.* 2010; Emerson *et al.* 2000; Juan *et al.* 1996; Macías-Hernández *et al.* 2013; Mairal *et al.* 2015). This hypothesis is supported by (1) the finding that populations of some species are structured into three genetic groups that geographically coincide with the three palaeo-islands upon which the island was formed (Carracedo 1994; Fernández-Palacios *et al.* 2011) and (2) the hypothesis that frequent volcanic activity and landslides in areas connecting those palaeo-islands would have provided isolation. Our finding of distinct niche suitability values between central areas and the three palaeo-islands is also consistent with the possibility that the three genetic clusters had long persisted in situ before the central part of the island was formed. The temporal frame, with the volcanic activity that built the central part of the island starting c. 3.5 Ma ago (Ancochea *et al.* 1990), is not contradictory with our estimation for the divergence of *L. acerifolia* from its sister species, *L. maritima*, c. 2.7 Ma ago (Villa-Machío et al., submitted), provided that we consider the wide confidence intervals associated with these types of estimations. However, for both the palaeo-island explanation as well as for the existence of genetic clusters including populations from several islands, which also occurs in other species from the Canary Islands and other

archipelagoes (Trusty *et al.* 2005), there are other possible biogeographic scenarios (see below).

Inferring the evolutionary history of *L. acerifolia* requires an interpretation of all the evidence in the frame of the general colonization of the archipelago and thus of island biogeographic theory (MacArthur & Wilson 1967; Warren *et al.* 2015; Whittaker & Fernández-Palacios 2007). In addition to identifying genetic clusters based on the GBS data and their distribution along the islands, discussed above, we believe that *L. acerifolia* fits a classical stepping-stone pattern of colonization. This hypothesis is congruent with the results provided by the species coalescent tree (SVDquartets), the estimates of genetic (nucleotide) diversity, differentiation (F_{st}) and heterozygosity levels, and the species distribution modeling.

Island migration model and phylogeography of *L. acerifolia*

The combination of molecular markers and SDM provides several sources of evidence that are supportive of a general east-west migration pattern. Several features, such as genetic structure, differentiation or nucleotide diversity, are consistent with such a pattern. Within the archipelago, the singularity of Tenerife is also noteworthy.

The topology of the SVDquartets tree is consistent with the composition of the genetic clusters obtained with STRUCTURE and BAPS, the only difference concerning the placement of the branch subtending the Guiniguada population depending on the dataset (mock reference vs. *de novo* approaches; Fig. 2a, b). In both trees, the earliest split is represented by Lanzarote - Fuerteventura vs. the remaining islands. The distribution of genetic diversity across the species range shows the same pattern using the datasets from workflows 1 and 2, although the nucleotide diversity (π) values resulting from the *de novo* approach (workflow 1) were greater (Table S1). This consists of a decline in π in populations as we move away from the mainland, which is consistent with species that colonized the islands following a stepping-stone model. With regards to differentiation, F_{st} values based on the dataset from workflow 2 (mock reference) are below 0.15 indicating low genetic differentiation between populations, between islands and between groups of islands along geographic longitude (Table S1b; Frankham *et al.* 2010; Jimenez *et al.* 2017). The strongest between-island differentiation is found between La Palma and

Fuerteventura (Table S2b). For this parameter there is a difference with the results based on the other dataset (*de novo*, workflow 1; Table S2a), since F_{st} values around 0.15 for the *de novo* dataset were achieved in all cases (populations, islands and groups of islands along geographic longitude). This difference could be due to the contribution to F_{st} of a higher number of SNPs obtained from this dataset (2099 vs. 1485), some of which could be paralogs (see below).

The pattern found in the proportion of heterozygous sites in each population is similar to that of nucleotide diversity (π). It varies according to the distance from the mainland. The distant populations, which are those occurring on younger islands, showed low proportions of heterozygous sites, likely resulting from a recent colonization event of *L. acerifolia* from older islands. In contrast, populations from the oldest islands, which are the closest to the mainland, present higher proportions of heterozygous sites, indicating old established populations. This pattern had been described for other Canary plant species such as *Phoenix canariensis* (Saro *et al.* 2015) and *Rumex bucephalophorus* (Talavera *et al.* 2013).

The conclusion that *L.acerifolia* shows a clear east-west colonization pattern is thus supported by several findings: (1) the higher genetic diversity and heterozygosity in the populations closer to the mainland; (2) the earlier split of populations from the oldest and easternmost islands, Lanzarote and Fuerteventura, in the SVDquartets tree; (3) the distinct genetic group formed by populations from these two islands; and (4) the African origin of its sister species (Villa-Machío et al., submitted). Around 25% of plants that colonized Macaronesia used this route (Caujapé-Castells 2011; Talavera *et al.* 2013).

However, such an east-west colonization pattern is compatible with different specific biogeographic scenarios. SDM provides very substantial independent evidence that is consistent with the phylogeographic reconstruction inferred from the GBS data and adds significant information. First, projections of the model to the latest glacial period show stability in areas with a suitable niche, even if the predicted areas are smaller, suggesting that *L. acerifolia* could have persisted within the same areas. This minimizes the possibility of extinction on islands and recolonization from other islands. Second, our SDM shows that the most influential variable in the potential distribution

of *L. acerifolia* is the slope, i.e., island topography. Recent studies on SDM from a close Canarian species (*Navaea phoenicea*; G. Fernández de Castro 2016) and observations in *L. maritima* show the importance of steep habitats, where organic matter accumulates, on shaping the species ranges. Topoclimatic suitability for *L. acerifolia* is significantly related to age, especially in areas of the Miocene period. More recent eruption events have smoothened the abrupt topography creating less complex habitats that are unsuitable for the species (Figs. 1, 4) according to the model. This is visible in southern La Palma, eastern Gran Canaria and especially in Tenerife. Third, the distribution of niche suitability along the archipelago and among different geological areas also poses some questions. Suitable areas are reduced in the eastern islands. This is consistent with the importance of topographic variables for SDM and suggests that not only recent eruptions but also flat senescent islands contribute to fewer opportunities for the occurrence of the species. The exceptions are remnant mountain enclaves, where genetic diversity is unexpectedly maintained. On the contrary, topoclimatic suitability tends to increase towards the western and younger islands. However, despite the availability of extensive areas that could be colonized, mainly on rocky slopes and cliffs that are not affected by historical eruptions, the westernmost islands only host one population per island or, in the case of El Hierro, none. This pattern may be indicative of the existence of a colonization front taking place at a slow rate, given that the species lacks evident dispersal mechanisms.

The genetic structure, the pattern of colonization inferred through molecular analyses and the supporting evidence of SDMs allow us to match the evolutionary history of *L. acerifolia* to the classical conception of island biogeography in which colonization is dependent on distance (or age) and size (or suitable areas) (Whittaker & Fernández-Palacios 2007). These sources of evidence also allow us to pin down and refine the interpretation of this general pattern of colonization by incorporating the dynamic ontogeny of the archipelago.

Selecting pipelines for polyploid species

Previous studies have shown that phylogenetic results using GBS can differ when using different sequencing lines for the same samples, different taxon sampling or even different software (Qi *et al.* 2015). The hexaploid nature of our study species represents

another source of uncertainty for the analysis of GBS data due to the potential confusion of paralogous loci resulting from polyploidization event(s) with orthologous loci. To minimize the uncertainty involved in the whole process of converting the raw GBS data into a set of SNPs, we compared the results from three different workflows using two different software programs: *de novo* assembly (workflow 1) using PyRAD, and two reference-based methods using GBS-SNP-CROP, a mock reference (workflow 2) and a real reference genome (workflow 3). The comparison of the three approaches discussed below is illustrative given the importance of polyploidy in the evolution of angiosperm species and the scarce number of studies using GBS with polyploids (Shafer *et al.* 2016). To allow comparison among the workflows, we tried to use the same settings and level of stringency, when possible, particularly in parameters such as the minimum number of samples (21, 75%). However, there are several caveats for this. For instance, this parameter did not represent exactly the same selection criterion in both pipelines because they are not completely parallel: in workflow 1, it selected the minimum number of samples with data for a particular locus to be included in the final dataset, while in workflow 2 the minimum number of samples operates in a later stage at which only SNPs occurring on a minimum number of samples are selected. In addition, the same percent (75%) does not imply the same final levels of missing data since the number of SNPs ultimately discovered differed between workflows. In addition, the two softwares have different capabilities, e.g., pyRAD estimates *a priori* error rate (E) and heterozygosity (H) for creating the consensus sequences within each sample. These estimates of E and H use the maximum likelihood equation of Lynch (2008), based on diploid organisms and therefore such estimates need not be accurate for a hexaploid species.

Our results show that the advantages for SNP calling when using a real reference genome (Clevenger *et al.* 2015) are greatly counteracted when it is not closely related to the study system (workflow 3). In our case, using the closest available genome, i.e., *Gossypium arboreum* belonging to a different tribe *Gossypieae,* led to a poorly informative SNP dataset for the questions of interest. The resulting matrix contained only 164 SNPs, representing 8% and 11% of the number of SNPs obtained from workflows 1 and 2, respectively. In addition, the genetic structure obtained from analyzing those 164 SNPs

with STRUCTURE is very poor (K=9; results not shown). This could be due to the low signal recovered when focusing on orthologous loci between distant species (estimated divergence time for *G. arboreum* and *L. acerifolia*: 45 Ma, range: 40-60 Ma; www.timetree.org/). It is likely that in addition to a distant phylogenetic relationship, the diploid nature of *G. arboretum* compared to the hexaploid nature of *L. acerifolia* contributed to rendering such an uninformative dataset. Our results are consistent with the recommendation to avoid non-close reference genomes by Shafer *et al.* (2016).

The two remaining workflows, 1 and 2, which provided a comparable number of SNPs (2099 in the *de novo*, 1485 in the mock reference approach), rendered congruent results. For instance, two genetic groups (K=2; Fig 2a, b) was the optimal partition of the dataset with both workflows using STRUCTURE. There were differences arising in the secondmost likely partition (K=3 for the *de novo* approach vs. K=4 with the mock reference), which coincided with the most likely partition identified by BAPS based on the two matrices (Fig. 2a, b). A major concern affecting all clustering approaches is the risk of inferring artefactual discrete groups in populations where genetic diversity is distributed continuously. Such spurious clusters are more likely to arise under spatially heterogeneous sampling of populations. However, this apparent incongruence disappears when the recently introgressed individual, from Guiniguada, is excluded from the analysis (results not shown) since the optimal number of groups identified is the same for both pipelines (K=2), but the secondmost likely partition by STRUCTURE in the *de novo* dataset increases to 4, as in the mock reference approach. The trees obtained with the SVDquartets approach also show some differences between the two approaches (Fig. 2a, b).

Apparently, this pattern is not the most expected: workflow 1, which recovers a higher number of SNPs (2099), recognizes fewer groups than workflow 2, which led to a lower number of SNPs (1485). If all the SNPs generated in the two approaches were orthologous across all samples, the higher the number of SNPs the better the sampling of the genomes and thus the more accurate estimation of the number of genetic groups. In this case, we could expect that the *de novo* workflow reflected the genetic structure more accurately (secondmost likely K, and BAPS partition). However, a higher number

of SNPs in the *de novo* workflow compared to the mock reference must be partly related, first, to relaxing the maximum number of shared heterozygous sites across all samples (maxSharedH=28), and it is conceivable that some of the SNPs despite other rigorous parameters (e.g., number of alleles) are not truly orthologous. The absence of a fourth genetic group in the *de novo* approach suggests that this cluster was genetically poorly differentiated and consequently cannot be readily detected by this workflow. Moreover, the SNP matrix generated from the *de novo* workflow 1 contains a significantly higher number of missing data resulting from the higher number of SNPs. Secondly, even if the mock reference is a pseudo-reference genome, calling our SNPs to a reference might be more effective for a polyploid genome than using a *de novo* constructed assembly. Indeed, workflow 2 minimizes the error due to false SNPs calls by controlling the minimum depth of the less-frequent allele (Melo *et al.* 2016). Further, the fact that the removal of an artificially introgressed individual (Guiniguada) causes the secondmost likely partition based on the *de novo* workflow to equal the mock reference result (K=4), suggests that the mock reference workflow is using more reliable SNPs.

To our knowledge, this is the first GBS population-level study describing plant phylogeography on oceanic islands, and certainly the first with island polyploids using a mock reference approach. The results obtained in this study, as well as the methods applied, could be further explored in other groups. In particular, we would like to call attention to the unquestionable convenience of using a reference genome from a closely related species when conducting phylogeographic studies based on GBS data. Nevertheless, when working with non-model organisms without a real reference genome, as with *L. acerifolia*, building a pseudoreference genome is the best option. Furthermore, although the results obtained from both pipelines were similar, PyRAD, which was recommended for phylogenetic analyses rather than phylogeographic studies with polyploid species, seems to be less sensitive to describing within-species and among-population variation. We believe that using GBS-SNP-CROP is a more advisable option when conducting polyploid genomic studies of non-model organisms at the intraspecific level.

ACKNOWLEDGEMENTS

The authors are grateful to Juli Caujapé Castells, Mario Mairal, Águedo Marrero, José Naranjo, Alfredo Reyes Betancort, Arnoldo Santos Guerra, Ángel B. Fernández López, and Stephan Scholz for providing samples, information on localities, or assistance with fieldwork; José Naranjo for information on possible sources of introgression within the Jardín Botánico Canario for the population of Guiniguada; Yurena Arjona, Mario Fernández-Mazuecos and Isabel Sanmartín for help in some analyses; Joaquina M. García-Martín for helpful advice . The Spanish State Meteorological Agency (AEMET) provided climate station data within the frame of the Dissertation project to I. V-M. Funding was provided by the Spanish ministries of Science, Technology and Innovation (grant CGL2010-16138) and Economy and Competitivity (grant CGL2013-49097-C2-1-P) and through the FPI fellowship BES-2011-049852 (co-financed by European Social Fund) to I.V-M.

REFERENCES

Ancochea E, Fuster J, Ibarrola E*, et al.* (1990) Volcanic evolution of the island of Tenerife (Canary Islands) in the light of new K-Ar data. *Journal of Volcanology and Geothermal Research* 44, 231-249.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17, 81-92.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Austin MP, Van Niel KP (2011) Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography* 38, 1-8.

Barton K (2013) MuMIn: multi-model inference. *R package version* 1.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Broennimann O, Petitpierre B, Randin C*, et al.* (2014) ecospat: Spatial ecology miscellaneous methods. *R package version* 1.

Carracedo J (1994) The Canary Islands: an example of structural control on the growth of large oceanic-island volcanoes. *Journal of Volcanology and Geothermal Research* 60, 225-241.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* 1, 171-182.

Caujapé-Castells J (2011) Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. *The biology of island floras*, 284-324.

Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA (2015) Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations. *Molecular plant* 8, 831-846.

Corander J, Cheng L, Marttinen P, Sirén J, Tang J (2013) BAPS: Bayesian analysis of population structure V. 6.0. *Department of Mathematics and statistics. University of Helsinki, Finland*.

Cox SC, Carranza S, Brown RP (2010) Divergence times and colonization of the Canary Islands by Gallotia lizards. *Molecular Phylogenetics and Evolution* 56, 747-757.

Chevan A, Sutherland M (1991) Hierarchical partitioning. *The American Statistician* 45, 90-96.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30, 3317-3324.

Devesa Alcaraz JA, Luque T (1986) Contribución al estudio citotaxonómico del género Lavatera (Malvaceae) en España.

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources* 4, 359-361.

Eaton DA (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, btu121.

Elshire RJ, Glaubitz JC, Sun Q*, et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6, e19379.

Emerson BC, Oromí P, Hewitt GM (2000) Interpreting colonization of the *Calathus* (Coleoptera: Carabidae) on the Canary Islands and Madeira through the application of the parametric bootstrap. *Evolution* 54, 2081-2090.

Escobar P (2007) *Filogenia de la alianza genérica de Malva: un enfoque molecular.* PhD Thesis, Universidad Autónoma de Madrid.

Escobar P, Schönswetter P, Fuertes Aguilar J, Nieto Feliner G, Schneeweiss GM (2009) Five molecular markers reveal extensive morphological homoplasy and reticulate evolution in the *Malva* alliance (Malvaceae). *Molecular Phylogenetics and Evolution* 50, 226-239.

Escudero M, Eaton DAR, Hahn M, Hipp AL (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution* 79, 359-367.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14, 2611-2620.

Fernández-Palacios JM, de Nascimento L, Otto R*, et al.* (2011) A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography* 38, 226-246.

Frankham R, Ballou JD, Briscoe DA (2010) *Introduction to Conservation Genetics* Cambridge University Press.

G. Fernández de Castro A (2016) *Islas dentro de islas: biología y conservación del paleoendemismo macaronésico Navaea phoenicea (Vent) Webb & Berthel* PhD Thesis, Universidad Autónoma de Madrid.

García-Verdugo C, Sajeva M, La Mantia T*, et al.* (2015) Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. *Molecular Ecology* 24, 726-741.

Garrison E (2012) Vcflib: A C++ library for parsing and manipulating VCF files.

Hess J, Kadereit J, Vargas P (2000) The colonization history of *Olea europaea* L. in Macaronesia based on internal transcribed spacer 1 (ITS-1) sequences, randomly amplified polymorphic DNAs (RAPD), and intersimple sequence repeats (ISSR). *Molecular Ecology* 9, 857-868.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology* 25, 1965-1978.

Hijmans RJ, Phillips S, Leathwick J, Elith J (2015) dismo: Species distribution modeling. R package version 1.0-12. *The R Foundation for Statistical Computing, Vienna http:// cran. r-project. org*.

Husemann M, Deppermann J, Hochkirch A (2014) Multiple independent colonization of the Canary Islands by the winged grasshopper genus *Sphingonotus* Fieber, 1852. *Molecular Phylogenetics and Evolution* 81, 174-181.

Jimenez A, Weigelt B, Santos-Guerra A*, et al.* (2017) Surviving in isolation: genetic variation, bottlenecks and reproductive strategies in the Canarian endemic *Limonium macrophyllum* (Plumbaginaceae). *Genetica* 145, 91-104.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070-3071.

Juan C, Emerson BC, Oromí P, Hewitt GM (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology & Evolution* 15, 104-109.

Juan C, Ibrahim KM, Oromi P, Hewitt GM (1996) Mitochondrial DNA sequence variation and phylogeography of Pimelia darkling beetles on the Island of Tenerife (Canary Islands). *Heredity* 77, 589-598.

Lassueur T, Joost S, Randin CF (2006) Very high resolution digital elevation models: Do they improve models of plant species distribution? *Ecological modelling* 198, 139-153.

Limborg MT, Seeb LW, Seeb JE (2016) Sorting duplicated loci disentangles complexities of polyploid genomes masked by genotyping by sequencing. *Molecular Ecology*.

Lu F, Lipka AE, Glaubitz J*, et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet* 9, e1003215.

Lynch M (2008) Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Molecular Biology and Evolution* 25, 2409-2419.

MacArthur RH, Wilson EO (1967) The theory of island biogeography. *Princeton, NJ*.

Macías-Hernández N, Bidegaray-Batista L, Emerson BC, Oromí P, Arnedo M (2013) The imprint of geologic history on within-island diversification of woodlouse-hunter spiders (Araneae, Dysderidae) in the Canary Islands. *Journal of Heredity* 104, 341-356.

Mairal M, Sanmartin I, Aldasoro JJ*, et al.* (2015) Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic *Canarina canariensis* (Campanulaceae). *Mol Ecol* 24, 3944-3963.

Mastretta-Yanes A, Arrigo N, Alvarez N*, et al.* (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* 15, 28-41.

McAllister CA, Miller AJ (2016) Single nucleotide polymorphism discovery via genotyping by sequencing to assess population genetic structure and recurrent polyploidization in *Andropogon gerardii*. *American Journal of Botany* 103, 1314-1325.

Melo AT, Bartaula R, Hale I (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC bioinformatics* 17, 29.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* 22, 2841-2847.

Nicotra AB, Chong C, Bragg JG*, et al.* (2016) Population and phylogenomic decomposition via genotyping-by-sequencing in Australian *Pelargonium*. *Molecular Ecology* 25, 2000-2014.

Pellegrino I, Boatti L, Cucco M*, et al.* (2016) Development of SNP markers for population structure and phylogeography characterization in little owl (*Athene noctua*) using a genotyping- by-sequencing approach. *Conservation genetics resources* 8, 13-16.

Penjor T, Mimura T, Matsumoto R, Yamamoto M, Nagano Y (2014) Characterization of limes (*Citrus aurantifolia*) grown in Bhutan and Indonesia using high-throughput sequencing. *Scientific reports* 4, 4853.

Pfeifer B, Wittelsbürger U, Onsins SER, Lercher MJ (2014) PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, msu136.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.

Qi ZC, Yu Y, Liu X*, et al.* (2015) Phylogenomics of polyploid *Fothergilla* (Hamamelidaceae) by RAD-tag based GBS—insights into species origin and effects of software pipelines. *Journal of Systematics and Evolution* 53, 432-447.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: A laboratory manual*, 2 edn. Cold Spring Harbor laboratory press New York.

Saro I, González-Pérez MA, García-Verdugo C, Sosa PA (2015) Patterns of genetic diversity in *Phoenix canariensis*, a widespread oceanic palm (species) endemic from the Canarian archipelago. *Tree genetics & genomes* 11, 1-13.

Schröder S, Mamidi S, Lee R*, et al.* (2016) Optimization of genotyping by sequencing (GBS) data in common bean (*Phaseolus vulgaris* L.). *Molecular Breeding* 36, 6.

Shafer ABA, Peart CR, Tusso S*, et al.* (2016) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, n/a-n/a.

Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then…and now: Stebbins revisited. *American Journal of Botany* 101, 1057-1078.

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. *Annual review of plant biology* 60, 561-588.

Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4.

Talavera M, Navarro-Sampedro L, Ortiz PL, Arista M (2013) Phylogeography and seed dispersal in islands: the case of *Rumex bucephalophorus* subsp. *canariensis* (Polygonaceae). *Annals of botany* 111, 249-260.

Thorpe RS, McGregor DP, Cumming AM, Jordan WC (1994) DNA evolution and colonization sequence of island lizards in relation to geological history: mtDNA RFLP, cytochrome b, cytochrome oxidase, 12s rRNA sequence, and nuclear RAPD analysis. *Evolution*, 230-240.

Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography* 32, 369-373.

Trusty JL, Olmstead RG, Santos-Guerra A, SÁ-Fontinha S, Francisco-Ortega J (2005) Molecular phylogenetics of the Macaronesian-endemic genus *Bystropogon* (Lamiaceae): palaeo-islands, ecological shifts and interisland colonizations. *Molecular Ecology* 14, 1177-1189.

Tyler L, Lee SJ, Young ND*, et al.* (2016) Population Structure in the Model Grass Is Highly Correlated with Flowering Differences across Broad Geographic Areas. *The Plant Genome* 9.

Valtueña FJ, López J, Álvarez J, Rodríguez-Riaño T, Ortega-Olivencia A (2016) *Scrophularia arguta*, a widespread annual plant in the Canary Islands: a single recent colonization event or a more complex phylogeographic pattern? *Ecology and Evolution* 6, 4258-4273.

Walsh C, Mac Nally R (2013) The hier. part package. Hierarchical partitioning. R package version 1.0-4.

Warren BH, Simberloff D, Ricklefs RE*, et al.* (2015) Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecology Letters* 18, 200-217.

Wendel JF (2015) The wondrous cycles of polyploidy in plants. *American Journal of Botany* 102, 1753-1756.

Whittaker RJ, Fernández-Palacios JM (2007) *Island biogeography: ecology, evolution, and conservation* Oxford University Press.

Wong MML, Gujaria-Verma N, Ramsay L*, et al.* (2015) Classification and Characterization of Species within the Genus *Lens* Using Genotyping-by-Sequencing (GBS). *PLoS ONE* 10, e0122025.

Yamada T, Maki M (2012) Impact of geographical isolation on genetic differentiation in insular and mainland populations of *Weigela coraeensis* (Caprifoliaceae) on Honshu and the Izu Islands. *Journal of Biogeography* 39, 901-917.

Author Contributions

All authors contributed to the design of the study, conducted the fieldwork and wrote the manuscript. In addition, G.N.F. helped with in depth interpretation of results, A.G.F. carried out the SDM and I.V.M. performed the laboratory work, filtered raw data by using different bioinformatics pipelines and analyzed all genetic data.

SUPPORTING INFORMATION

Appendix S1

Workflow1- De novo. Conducted with the software PyRAD v3.0.65 (Eaton 2014), this pipeline processes the data by filtering it through various procedures for aligning and assembling reads. The first filter discarded reads with more than 4 sites for which Phred quality scored below 20. Subsequently, different thresholds were applied for sequence clustering (c = 80, 85, 90, 95) using VSEARCH v.1.1.3 (https://github.com/torognes/vsearch) and the remaining reads were aligned with MUSCLE v.3.8.31 (Edgar 2004) (Fig. S1). After applying this filter, we selected a 90% threshold for clustering reads within and across samples since this recovered the highest number of SNPs (Table 2). A 90% threshold might seem slightly low for an intraspecific study since it could increase the number of paralogous loci aggregated into a single cluster (Harvey et al., 2015). To counteract this effect, we combined this threshold with the exclusion of loci with more than 3 heterozygous sites in consensus sequences and a maximum number of 20 SNPs per final loci, also discarding highly repetitive genome regions. The minimum coverage for a cluster was set to 10. Because we are analyzing intraspecific samples, we would expect a higher number of shared polymorphic sites across individuals compared to phylogenetic studies that analyze individuals from different species. For this parameter, we have deviated from the recommendations in PyRAD (Eaton 2014) and avoided a stringent threshold. In order to consider a locus for the output files, an additional threshold was considered, the minimum number of samples represented in a final locus (m = 11, 14, 17, 21 and 25), which also determines the amount of missing data. After performing this filter, several datasets were obtained from this workflow: novo-c90m11, novo-c90m14, novo-c90m17, novo-c90m21 and novo-c90m25. We kept these five matrices for comparative purposes.

Workflow 2- Mock reference. This approach was carried out with GBS-SNP-CROP v.2.0 (Melo *et al.* 2016) software which included four main steps: (1) processing the raw data; (2) building a mock reference, that is, a reduced representative reference to allow GBS reads to be mapped and SNPs discovered; (3) mapping processed reads to the mock reference and generating alignments files; and (4) SNP calling and genotyping. Since our

data were already processed, we started the workflow directly in step 2 by building a mock reference from quality-filtered reads belonging to 5 samples (AGU, FTV, GUIN, TENO, TIJ; Table 1). These samples were selected to represent the genetic groups recovered from the *de novo* approach. This second step started by merging paired-end reads into single reads using the PEAR v.0.9.10 software package (Zhang *et al.* 2014). To build the mock reference, GBS-SNP-CROP concatenated into a single FASTQ file not only the "assembled" file, which contains only successfully merged reads, but also the "stitched" file, containing R1 and R2 unmerged reads joined by a high-quality poly-A stretch. As in the *de novo* workflow 1, the clustering of these assembled-reads was carried out applying a 90% similarity threshold both within and among samples with the USEARCH software v9.1.13 (http://www.drive5.com/usearch/). This second step concludes by creating a representative reference sequence for the whole set of GBS fragments. We also built a mock reference based on just one sample (AGU), a possibility that is also suggested by Melo *et al.* (2016). But we discarded it since it produced a far shorter sequence than the one generated with five samples (4.557.193 vs. 18.644.945 bp) which we expected to imply reducing substantially the number of recovered SNPs. In the third step, the quality-filtered reads from all samples were aligned to this mock reference using BWA v.0.7.12 (Li & Durbin 2009) and SAMtools v.1.2 (Li *et al.* 2009). The fourth step, SNP calling and its filtering, was conducted following the authors' guidelines (Melo et al., 2016) for hexaploid species: minimum depth required for calling a homozygote when the alternative allele depth=0 (mnHoDepth0 =17); minimum depth required for calling a homozygote when the alternative allele depth=1 (mnHoDepth1 =76); minimum depth required for each allele when calling a heterozygote (mnHetDepth =3); allele frequency filter (altStrength=0.862); minimum required ratio of the depth of the secondary allele to that of the primary (mnAlleleRatio=0.063); minimum proportion of genotyped individuals to retain a SNP (MnCall=0.75); and minimum and maximum average depth of an acceptable SNP (mnAvgDepth=4, mxAvgDepth=200). The resulting default matrix produced clusters with up to 7 SNPs and an average depth of 10 reads per cluster. We filtered this matrix and produced three files reducing the maximum number of SNPs to 3, 4 and 5 (mock3, mock4 and mock5 files, respectively) and two others imposing a

minimum depth of 20 and 50 (mock20x and mock50x files, respectively). We kept these five matrices for comparative purposes.

Some of the criteria were applied under the two workflows to allow comparison between them. For instance, the number of minimum number of samples in which a SNP should be present (21, i.e., 75% of the samples) was applied to obtain both final matrices. In addition, we limited the number of alleles per SNP to two in the *de novo* workflow 1 despite the fact that PyRAD can handle more because GBS-SNP-CROP only retains potential bi-allelic SNPs.

Workflow3- Reference genome. We used the closest available reference genome to *Lavatera acerifolia*, which corresponded to the diploid species *Gossypium arboreum* (GenBank accession number GCA_000612285.2), belonging to the tribe Gossypieae, and sister to tribe Malveae, in the family Malvaceae. This approach was also performed using GBS-SNP-CROP v.2.0 (Melo *et al.* 2016), which provides an efficient pipeline for species with or without a reference genome for SNPs discovery. We applied strict parameters for hexaploid species following the authors' recommendations under workflow 2.

In addition, these workflows were rerun including one sample of its sister species (*L.maritima*) for phylogenetic analyses.

References

Eaton DA (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, btu121.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760.

Li H, Handsaker B, Wysoker A*, et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Melo AT, Bartaula R, Hale I (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC bioinformatics* **17**, 29.

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620.

Table S1. Genetic diversity parameters estimated for *Lavatera acerifolia* across its geographical range using SNPs extracted from GBS data using two different bioinformatics workflows (*de novo* assembly and mock reference genome). Nucleotide diversity (π) and F_{ST} values were calculated for different clusters (populations, islands and longitudinal). Populations with one individual were excluded for the population level estimation. See population codes in Table 1.

a) **Workflow 1 - *De novo***

Populations

	TIJ	**AGU**	**TENO**	**MAS**	**ACO**	**INF**	**GI**	**CHA**	**HP**	**AGA**	**GUA**	**GUIN**	**HAL**
π	0.059	0.101	0.099	0.076	0.091	0.085	0.097	0.081	0.106	0.105	0.126	0.103	0.077
FST	0.476	0.345	0.352	0.403	0.387	0.361	0.402	0.437	0.328	0.344	0.306	0.311	0.485

Islands

	LP	**GO**	**TF**	**GC**	**LZT**	**FTV**
π	0.059	0.101	0.138	0.127	0.112	0.077
FST	0.459	0.352	0.256	0.311	0.427	0.452

Longitudinal groups

	W	**CW**	**CE**	**E**
π	0.101	0.138	0.127	0.124
FST	0.277	0.175	0.222	0.301

b) **Workflow 2 -Mock reference**

Populations

	TIJ	**AGU**	**TENO**	**MAS**	**ACO**	**INF**	**GI**	**CHA**	**HP**	**AGA**	**GUA**	**GUIN**	**HAL**
π	0.040	0.055	0.052	0.047	0.059	0.053	0.057	0.056	0.061	0.062	0.071	0.060	0.051
FST	0.200	0.103	0.127	0.210	0.177	0.169	0.195	0.148	0.108	0.139	0.109	0.123	0.198

Islands

	LP	**GO**	**TF**	**GC**	**LZT**	**FTV**
π	0.040	0.055	0.065	0.064	0.067	0.051
FST	0.190	0.108	0.092	0.139	0.127	0.163

Longitudinal groups

	W	**CW**	**CE**	**E**
π	0.051	0.065	0.064	0.058
FST	0.115	0.081	0.121	0.147

Table S2. F_{st} pairwise values estimated for *Lavatera acerifolia* employing SNPs extracted from GBS data using two different bioinformatics workflows (*de novo* assembly and mock reference genome). F_{ST} values were calculated for different clusters (populations, islands and longitudinal). See populations codes in Table 1.

a) **Workflow 1 - *De novo***

Populations

	AGU	TENO	MAS	ACO	INF	GI	CHA	HP	AGA	GUA	GUIN	HAL
TIJ	0.317	0.432	0.414	0.411	0.388	0.539	0.538	0.489	0.492	0.454	0.445	0.608
AGU		0.308	0.247	0.242	0.182	0.414	0.414	0.358	0.366	0.328	0.307	0.493
TENO			0.350	0.353	0.303	0.395	0.438	0.326	0.337	0.313	0.179	0.503
MAS				0.307	0.248	0.475	0.521	0.422	0.435	0.379	0.350	0.551
ACO					0.165	0.463	0.494	0.393	0.407	0.364	0.349	0.527
INF						0.451	0.494	0.393	0.372	0.367	0.299	0.510
GI							0.280	0.272	0.314	0.230	0.338	0.506
CHA								0.337	0.367	0.299	0.387	0.540
HP									0.140	0.065	0.188	0.408
AGA										0.153	0.227	0.421
GUA											0.192	0.386
GUIN												0.440

Islands

	GO	TF	GC	LZT	FTV
LP	0.317	0.280	0.417	0.568	0.608
GO		0.130	0.287	0.457	0.493
TF			0.118	0.331	0.377
GC				0.356	0.360
LZT					0.412

Longitudinal groups

	CW	CE	E
W	0.127	0.285	0.382
CW		0.118	0.261
CE			0.254

b) **Workflow 2 -Mock reference**

Populations

	AGU	TENO	MAS	ACO	INF	GI	CHA	HP	AGA	GUA	GUIN	HAL
TIJ	0.107	0.150	0.195	0.236	0.155	0.313	0.216	0.211	0.230	0.205	0.199	0.269
AGU		0.098	0.118	0.069	0.013	0.203	0.110	0.122	0.138	0.118	0.103	0.156
TENO			0.182	0.188	0.159	0.210	0.144	0.097	0.125	0.127	0.016	0.218
MAS				0.178	0.157	0.310	0.259	0.223	0.264	0.199	0.207	0.308
ACO					0.003	0.260	0.229	0.191	0.243	0.194	0.193	0.257
INF						0.276	0.248	0.211	0.233	0.196	0.164	0.281
GI							0.043	0.123	0.172	0.083	0.188	0.235
CHA								0.074	0.133	0.071	0.168	0.128
HP									-0.031	-0.036	0.042	0.162
AGA										-0.014	0.040	0.202
GUA											0.066	0.154
GUIN												0.233

Islands

	GO	TF	GC	LZT	FTV
LP	0.107	0.119	0.206	0.231	0.269
GO		0.009	0.116	0.145	0.156
TF			0.077	0.092	0.157
GC				0.116	0.184
LZT					0.044

Longitudinal groups

	CW	CE	E
W	0.036	0.136	0.167
CW		0.077	0.125
CE			0.150

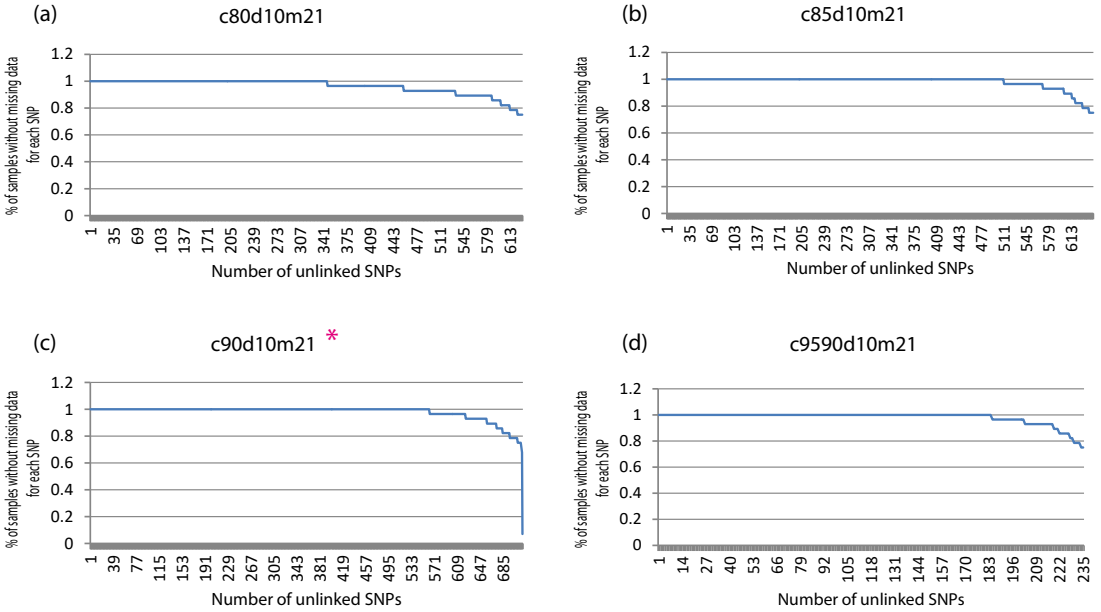


Figure S1. Graphic representation of the effects of different threshold filters applied in workflow 1 (*de novo*) on the percent of missing data in the construction of a SNPs dataset from GBS data in *Lavatera acerifolia*: a) 80% threshold filtering for clustering reads within samples, b) 85% threshold for clustering reads within samples, c) 90% threshold for clustering reads within samples, and d) 95% threshold for clustering reads within samples and 90% for clustering across samples.

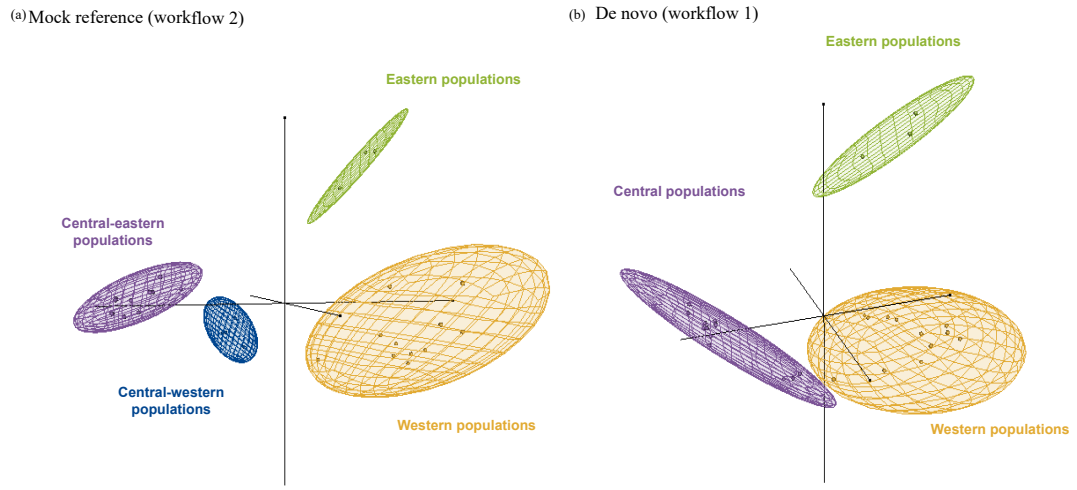


Figure S2. Principal component analysis (PCA) of 28 samples of *Lavatera acerifolia* based on SNPs data generated from GBS using two different bioinformatics workflows. Scatter diagrams of the samples against the first three axes based on: a) 1485 SNPs extracted from a matrix (mock20x) under workflow 2 (mock reference genome) explaining 48.24% of variance; b) 1318 variant sites records from a matrix (novo-c90m21) under workflow 2 (*de novo* assembly), explaining 32.15% of variance. Ellipses enclosing sample points constructed with 95% confidence, matching genetic groups discussed in the main text.

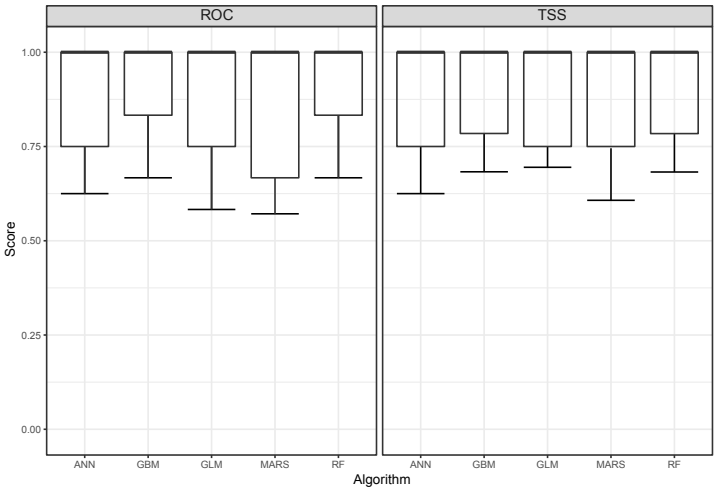


Figure S3. Boxplots of the scores of 50 evaluation runs for testing the species distribution model developed for *Lavatera acerifolia* using five algorithms (ANN, GBM, GLM, MARS, RF): a) TSS scores; b) AUC scores (see text).

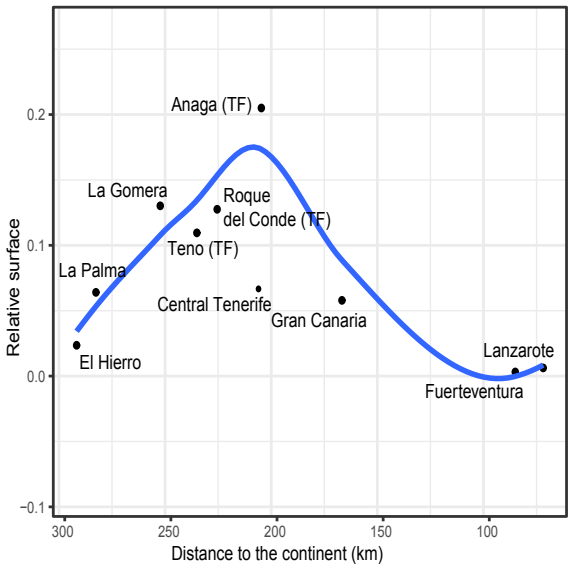


Figure S4. Scatterplot of the relative suitable surface (total suitable surface/total surface) in each island for the species distribution model developed for *Lavatera acerifolia* represented against the distance to the mainland. Tenerife is divided into four groups (the three palaeo-islands, Anaga, Teno, Roque del Conde, and the central area). The blue line is fitted with a loess function, excluding central Tenerife.

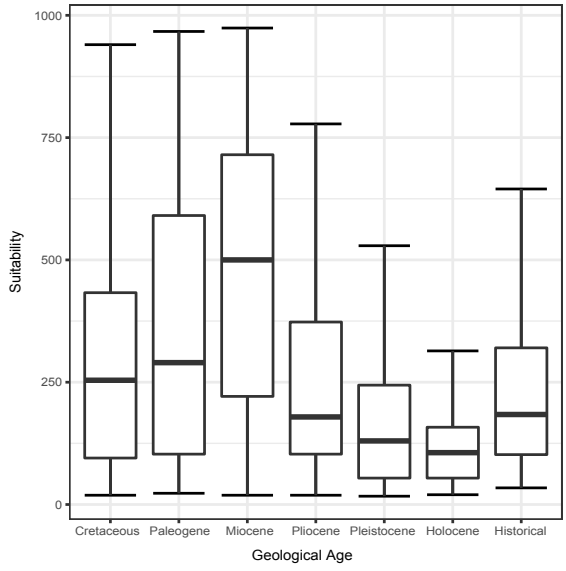


Figure S5. Relationship between relative topoclimatic suitability (total suitable surface/ total surface) for the species distribution model developed for *Lavatera acerifolia* and geological age.

CAPÍTULO 3

Niche segregation between Mediterranean and Macaronesian lineages: the *Lavatera maritima–Lavatera acerifolia* divergence

Villa-Machío, I., Fuertes-Aguilar, J., Nieto Feliner, G.
Manuscript in preparation

INTRODUCTION

The geographical distributions of species are determined by their ecological requirements along with the interacting historical factors. These two components represent the major visions of biogeography (ecological and historical). The same components are fundamental in diversification processes (Soberón, 2007; Wiens, 2004). One of the goals of evolutionary biology is to understand the relative roles of geography and ecology in speciation. Studying recent speciation events in which the involved species –ideally sister species— are known, may facilitate gaining knowledge on the relative importance of geographical and ecological speciation, which is important because ecological determinants of speciation are difficult to account for. In particular, current niches of the sister species and their degree of overlap may provide clues for inferring the mode of speciation that split them and the role of geography in the process. Similar niche characteristics are often observed in close relatives suggesting that niche conservatism operated across speciation processes (Burns & Strauss, 2011; Peterson, Soberón & Sánchez-Cordero, 1999; Wiens & Graham, 2005). When the niche is very similar, ecological differentiation is excluded and the reproductive isolation required for differentiation and speciation is inferred to have occurred by geographic isolation, e.g., in a vicariance scenario. But closely related species may show different environmental niches (Nakazato, Warren & Moyle, 2010; Warren *et al.*, 2008). In this case, niche differentiation may have occurred during the speciation process itself via the adaptation to new ecological conditions, i.e., under the ecological speciation model (Schluter, 2009). However, niche differentiation may also occur after a speciation process that was driven by geographic isolation, that is, subsequently to allopatric speciation (Butlin, Galindo & Grahame, 2008; Mayr, 1942). The chances for niche differentiation to occur subsequently to allopatric speciation are greater in cases of peripatric speciation, i.e., when population sizes of the differentiated species are markedly different (Coyne & Orr, 2004). Good sources of examples of this mode of peripatric speciation are oceanic islands, containing endemics that are peripheral isolates of a continental widely distributed species. Because of this, reconstructing niche evolution over time can help us understand current species distribution and their origin. The availability of species distribution datasets along with

the developments of geographical information systems and species distribution modeling allow addressing questions on niche evolution of sister species and even ancestors (Warren *et al.*, 2008). Combining niche modelling and genetic clustering techniques may allow discovering undetected or cryptic hints for current intraspecific differentiation whenever different genetic groups show some degree of niche differentiation (Piñeiro *et al.*, 2007). However, using such a combined approach, we would like to explore if we can also identify traces of ancient niche differentiation, e.g., in a pair of sister species one of which has evolved as a peripheral isolate of the other. For this, we propose to analyze the genetic data under a phylogenetic approach so that we can place the identified genetic groups in an evolutionary frame and try hindcasting ancient niches of their ancestors up to the common ancestor of the two species.

The Canary archipelago and the North African mainland are a suitable area for examining the above questions because of the biogeographic relationships between their biotas (García-Aloy *et al.*, 2017; Mairal *et al.*, 2015a; Weingartner, Wahlberg & Nylin, 2006). The Macaronesian region consists of five volcanic archipelagos in the Atlantic region, but the Canary Islands present the largest surface and the closest proximity to the continent. In addition, this region includes diverse seamounts that, in the glacial periods, may have facilitated stepping stone colonization of the archipelagoes from the mainland (Fernández-Palacios *et al.*, 2011). As a consequence of such biogeographic connections, the Canary Islands have acted as refugia for continental linages from the Late Cenozoic (Mairal *et al.*, 2015b). But a significant number of molecular studies suggested a general dispersal pattern from the mainland followed by insular speciation (Carine *et al.*, 2004; Díaz-Pérez *et al.*, 2008; Ferreira *et al.*, 2015; Garnatje, Garcia & Canela, 2007) although it has been proposed that such basic scenario should be completed to include multiple colonization events and subsequent admixture instead of single colonization events (Caujapé-Castells *et al.*, 2017).

Our study system involves two sister species: *Lavatera maritima,* which is a continental species widely distributed along the western Mediterranean basin, and *L. acerifolia*, an endemic from the Canary Islands. Their divergence time, based on plastid DNA sequences, was estimated to the Plio-Pleistocene, c. 2,7 Ma (Chapter 1). The

phylogeographic study of *L. maritima* based on plastid DNA sequences, presented in that same Chapter, strongly supported a North African origin and suggested more than one dispersal events to Europe. The projection of its species distribution model, based on bioclimatic variables, onto the Last Interglacial (130-115 ka) also highlighted the feasibility of a colonization of the Canaries since there were large areas of high suitability for *L. maritima* in the archipelago not only during the LIG but also along the LGM. The evolutionary history of *L. acerifolia* inferred based on GBS data revealed a clear pattern of east-west colonization along the Canary Islands (Chapter 2). This route is consistent with the species distribution model, based on bioclimatic and topographic variables, for the current time as well as for the projection up to the Last Glacial Maximum (20-26.5 ka). Based on the adequate knowledge of the phylogeographic history of these two species and their sister group relationship (Escobar et al., 2009), this species pair constitutes a suitable system for addressing niche evolution over time up to their splitting. There are two possible basic scenarios for the divergence between the two species, either it occurred in North Africa or *L. acerifolia* diverged in the Canary archipelago. Therefore, we use phylogenetic relationships based on GBS data and environmental niche overlap analyses combined with the species distribution modelling (SDM) at both species and genetic-group levels, for assessing the relative importance of geographic and ecological factors in speciation.

The specific goals are: 1) to calculate the bioclimatic niche overlap between the two sister species, 2) to identify the most influential bioclimatic variable(s) in the niche of each species, 3) to estimate the phylogenetic relationships of the genetic groups of *L. maritima* and *L. acerifolia*, 4) to examine the association of the different genetic groups and their environmental niche spaces, and 5) to explore the cross-similarity of climate suitability between species in both present and past scenarios.

MATERIAL AND METHODS

Study system and sampling

The study area includes the native distribution of *L. maritima* and *L. acerifolia*, that is, the western Mediterranean basin including north-western Morocco (France, Italy, Spain, Morocco and Algeria) and the Canary Islands (Fuerteventura, Lanzarote, Gran Canaria, Tenerife, La Gomera, El Hierro and La Palma), respectively. For contructing the environmental models of the two species, we collected 54 presence data of L. *maritima* during the field sampling, plus 3 more records from three African populations, Jbel Gafsa (MA-77089), Oran (MA-77088) and Jbel Ansitten, MA-77086), from which there were specimen records at the Herbarium of Real Jardín Botánico-CSIC (Madrid). Because of these records lacked exact geographic coordinates, they were georeferenced using Google Earth (https://www.google.com/intl/es/earth/). Seventeen occurrences of *L. acerifolia,* distributed along the species range (six out of the seven islands), were recorded. In total, 57 populations of *L. maritima* and 17 of *L. acerifolia* were georeferenced (Table 1).

DNA extraction and GBS data

For discovering the genetic groups across the two species and inferring the coalescence-based relationships as well as for constructing the environmental models of those genetic groups, we used 15 populations from *L. acerifolia* and 33 populations from *L. maritima*, which represent the natural distribution of both species. All populations consisted of two individuals, except for Gourges du Zegzel, Menorca, Maxuquera and Estopiñán del Castillo populations, which comprised three individual of *L. maritima*, and Bajamar and Famara populations, which included only one individual of *L. acerifolia.* In total, 98 individuals belonging to 48 populations were analyzed, plus one individual from the Canary endemic *Naveae phoenicea*, sister to the *L. maritima-L. acerifolia* clade (Escobar *et al.*, 2009). Total DNA was extracted from leaves using a DNeasy Plant Minikit (QIAGEN Inc., California) and concentrated using a precipitation protocol described in Sambrook, Fritsch and Maniatis (1989). DNA samples were processed to obtain pair-end GBS libraries following the same protocol explained in Chapter 2.

Table 1. Populations of *Lavatera maritima* and *Lavatera acerifolia* used for this study. Shaded rows indicate populations for which genetic data have been newly generated.

Species	Population	Codes	Longitude	Latitude	Locality	Colector
Lavatera maritima	Oran	ORA	-0.7041	35.7371	Argelia, Orán	A. Faure (MA-77088)
Lavatera maritima	Jbel Milock	MIL	2.8450	33.9280	Argelia, Laghouat	L. Faurel (MA-841613)
Lavatera maritima	Misserghin	MIS	0.7518	35.6403	Argelia, Orán	I. Álvarez & M. Kaid-Harche (MA-910937)
Lavatera maritima	Jbel Gafsa	GAF	8.9450	34.3873	Tunisia, Gafsa	CJ Pitard (MA-77089)
Lavatera maritima	Vingrau	VIN	2.7902	42.8544	France, Pyrrénnées Orientales	F. Médail (AIX)
Lavatera maritima	Evenos	EVE	5.8552	43.1628	France, Var	A. Baumel (AIX)
Lavatera maritima	Cabasse	CAB	5.8927	43.1787	France, Var	A. Baumel (AIX)
Lavatera maritima	Gémenos	GEM	5.6334	43.3000	France, Bouches-du-Rhone	A. Baumel (AIX)
Lavatera maritima	Grand Vallon	LAS	5.5710	44.3483	France, Bouches du Rhône	A. Baumel (AIX)
Lavatera maritima	Île de Ratonneau	FRI	5.3069	43.2881	France, Bouches du Rhône	A. Baumel (AIX)
Lavatera maritima	Col de l'Arma	ROY	7.5251	43.9010	France, Alpes maritimes	M. Pires (AIX)
Lavatera maritima	Roche amère	VIL	5.8449	43.8943	France, Alpes de Haute Provence	A. Baumel (AIX)
Lavatera maritima	Île de Gargalu	IGA	8.5534	42.3693	France, Corsica	F. Médail (AIX)
Lavatera maritima	Cala Gonone	CER	9.6173	40.2800	Italy, Sardinia	P. Escobar (MA-709504)
Lavatera maritima	Imouzzer des Ida-Outanane	IMO	-9.4820	30.6760	Morocco, Agadir	J. Fuertes & G. Nieto (MA-853378)
Lavatera maritima	Jbel Ansitten	ANS	-9.6333	31.1667	Morocco, Agadir	E. Jahandiez (MA-77086)
Lavatera maritima	Abdadgadel	ABD	-2.6082	34.9490	Morocco, Zaio	A. Gonzalez & I. Villa (MA-910940)
Lavatera maritima	Badés (Peñon de la Gomera)	BAD	-4.2938	35.1710	Morocco, Al-Hoceima	A. Gonzalez & I. Villa (IVM 24)
Lavatera maritima	Beni Snassen Monts	BSN	-2.1390	34.7761	Morocco, Berkane	A. Gonzalez & I. Villa (MA-910944)
Lavatera maritima	Embalse Mechrá-Homadi	MEC	-2.8069	34.7376	Morocco, Selouanne	A. Gonzalez & I. Villa (MA-910942)
Lavatera maritima	Gareb	GAR	-3.1389	34.9175	Morocco, Tiztoutine	A. Gonzalez & I. Villa (MA-910941)
Lavatera maritima	Gorges du Zegzel	ZEG	-2.3685	34.8350	Morocco, Berkane	A. Gonzalez & I. Villa (MA-910946)
Lavatera maritima	Jbel Guilliz	JBG	-3.3237	34.4942	Morocco, Guercif	A. Gonzalez & I. Villa (IVM 25)
Lavatera maritima	Plaine du Gareb	PLA	-3.1126	34.8839	Morocco, Tiztoutine	A. Gonzalez & I. Villa (MA-910947)
Lavatera maritima	Puerto de Alhucemas	ALH	-3.9389	35.2500	Morocco, Al-Hoceima	A. Gonzalez & I. Villa (MA-910939)
Lavatera maritima	Saidia	SAI	-2.2120	35.0593	Morocco, Saidia	A. Gonzalez & I. Villa (MA-910945)
Lavatera maritima	Tazaguine	TAZ	-2.3439	34.8857	Morocco, Berkane	A. Gonzalez & I. Villa (MA-910943)
Lavatera maritima	Zaio	ZAI	-2.6986	34.9500	Morocco, Zaio	A. Gonzalez & I. Villa (IVM 26)
Lavatera maritima	Aguadulce	AGU	-2.4986	36.8379	Spain, Andalucía, Almería	F. Durán, I. Villa & J.C. Zamora (MA-910953)
Lavatera maritima	Rambla la Moladera	MOL	-2.2470	36.8182	Spain, Andalucía, Almería	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	Faro Cabo de Gata	GATA	-2.1848	36.7292	Spain, Andalucía, Almería	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	Los Lobos	LOB	-1.7606	37.3058	Spain, Andalucía, Almería	F. Durán, I. Villa & J.C. Zamora (MA-910954)
Lavatera maritima	Cala San Pedro (Las negras)	PED	-1.9986	36.8910	Spain, Andalucía, Almería	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	Calahonda	CAL	-3.4167	36.7018	Spain, Andalucía, Granada	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	Lobres (río Guadalfeo)	FEO	-3.5451	36.7871	Spain, Andalucía, Granada	F. Durán, I. Villa & J.C. Zamora (MA-910951)
Lavatera maritima	Pizarra	PIZ	-4.7007	36.7610	Spain, Andalucía, Málaga	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	Llanos de Líbar (Montejaque)	LIB	-5.2622	36.7365	Spain, Andalucía, Málaga	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	Frigiliana	GIL	-3.8964	36.7940	Spain, Andalucía, Málaga	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	Granja Suárez	SUA	-4.4569	36.7402	Spain, Andalucía, Málaga	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	La Araña	ARA	-4.3341	36.7203	Spain, Andalucía, Málaga	F. Durán, I. Villa & J.C. Zamora
Lavatera maritima	Otiñar	OTI	-3.7629	37.6912	Spain, Andalucía, Jaén	F. Durán, I. Villa & J.C. Zamora (MA-910955)
Lavatera maritima	Cabo Cope (Águilas)	COP	-1.4858	37.4338	Spain, Región de Murcia, Murcia	S. Manzano & D. Orgaz
Lavatera maritima	Cabo Tiñoso (Cartagena)	TIN	-1.1589	37.5490	Spain, Región de Murcia, Murcia	S. Manzano & D. Orgaz
Lavatera maritima	Sierra Espuña	ESP	-1.5263	37.8814	Spain, Región de Murcia, Murcia	S. Manzano & D. Orgaz
Lavatera maritima	Montgo	MON	0.1252	38.8123	Spain, Com. Valenciana, Alicante	A. Gonzalez (MA)
Lavatera maritima	Lliber	LLI	0.0030	38.7493	Spain, Com. Valenciana, Alicante	A. Gonzalez (MA)
Lavatera maritima	L'ocaive	LOC	-0.0141	38.7790	Spain, Com. Valenciana, Alicante	A. Gonzalez (MA)
Lavatera maritima	Barranco del Averno	AVE	-0.6075	38.7971	Spain, Com. Valenciana, Alicante	A. Gonzalez (MA)
Lavatera maritima	Marxuquera	MAR	-0.2300	38.9867	Spain, Com. Valenciana, Valencia	A. Gonzalez (MA)
Lavatera maritima	Borriol	BOR	0.0416	40.0291	Spain, Com. Valenciana, Castellón	Ricarda Riina
Lavatera maritima	Racó del Frare	RAC	0.1781	40.4619	Spain, Com. Valenciana, Castellón	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Lavatera maritima	Alquezar	ALQ	0.0276	42.1681	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Lavatera maritima	Cap de Norfeu	NOR	3.2505	42.2536	Spain, Cataluña, Girona	A. Gonzalez (MA)
Lavatera maritima	Castillonroy	CAS	0.5790	41.8821	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Lavatera maritima	Embid de la Ribera	EMB	-1.5991	41.4168	Spain, Aragón, Zaragoza	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Lavatera maritima	Estopiñán del Castillo	EST	0.6058	41.9764	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Lavatera maritima	Mallos de Riglos	MAL	0.7272	42.3553	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Lavatera maritima	Ólvena	OLV	0.2455	42.0997	Spain, Aragón, Huesca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Lavatera maritima	Cala de Sant Llorenç	MEN	4.0876	39.8864	Spain, Islas Baleares, Menorca	J.C. Moreno Saiz, J. Fuertes, A. Gonzalez (MA)
Lavatera acerifolia	Morro del Halconcillo	HAL	-13.9265	28.3569	Spain, Canary Islands, Fuerteventura	S. Scholz, I. Villa
Lavatera acerifolia	Vecindad de Enfrente (Agaete)	AGA	-15.6711	28.0823	Spain, Canary Islands, Gran Canaria	G. Nieto, I. Villa
Lavatera acerifolia	Barranco Guiniguada	GUIN	-15.4628	28.0660	Spain, Canary Islands, Gran Canaria	G. Nieto, I. Villa
Lavatera acerifolia	Barranco de Guayadeque 1	GUA	-15.5114	27.9368	Spain, Canary Islands, Gran Canaria	G. Nieto, I. Villa
Lavatera acerifolia	Barranco de Guayadeque 2	GUA	-15.4990	27.9360	Spain, Canary Islands, Gran Canaria	G. Nieto, I. Villa
Lavatera acerifolia	Barranco de Guayadeque 3	GUA	-15.4761	27.9335	Spain, Canary Islands, Gran Canaria	G. Nieto, I. Villa
Lavatera acerifolia	Hoyo de Pineda (Barranco Anzo)	HP	-15.6395	28.1131	Spain, Canary Islands, Gran Canaria	G. Nieto, I. Villa
Lavatera acerifolia	Agulo	AGU	-17.1902	28.1848	Spain, Canary Islands, La Gomera	J. Fuertes, A. Gonzalez
Lavatera acerifolia	Barranco Jorado (Tijarafe)	TIJ	-17.9605	28.7034	Spain, Canary Islands, La Palma	J. Fuertes, A. Gonzalez
Lavatera acerifolia	Barranco Famara	FAM	-13.4783	29.2183	Spain, Canary Islands, Lanzarote	Jose D. Naranjo
Lavatera acerifolia	Barranco de los Infiernos	INF	-16.7116	28.1338	Spain, Canary Islands, Tenerife	J. Fuertes, A. Gonzalez
Lavatera acerifolia	Masca	MAS	-16.8413	28.2986	Spain, Canary Islands, Tenerife	J. Fuertes, A. Gonzalez
Lavatera acerifolia	Barranco Guaria (Acojeja)	ACO	-16.7558	28.1949	Spain, Canary Islands, Tenerife	J. Fuertes, A. Gonzalez
Lavatera acerifolia	Güimar	GI	-16.4058	28.2938	Spain, Canary Islands, Tenerife	J. Fuertes, A. Gonzalez
Lavatera acerifolia	Bajamar	BAJ	-16.3408	28.5521	Spain, Canary Islands, Tenerife	J. Fuertes, A. Gonzalez
Lavatera acerifolia	Chamorga	CHA	-16.1408	28.5783	Spain, Canary Islands, Tenerife	J. Fuertes, A. Gonzalez
Lavatera acerifolia	Punta de Teno (Buenavista)	TENO	-16.8949	28.3491	Spain, Canary Islands, Tenerife	J. Fuertes, A. Gonzalez

Genetic structure analyses

The bioinformatic analysis for discovering SNPs in each individual genome was performed using GBS-SNP-CROP v.2 (Melo, Bartaula & Hale, 2016) according to the authors' recommendations for hexaploid genomes and following our conclusions in Chapter 2

about relative performances of three bioinformatics pipelines. The mock reference was built using five samples of *L. acerifolia* (AGU, FTV, GUIN, TENO and TIJ; Table 1). The minimum average read depth was set to 20.

A Bayesian-likelihood approach was applied to the SNPs resulting matrix in BAPS v.6.0 (Corander *et al.*, 2013) for inferring the number of genetic clusters from all populations of both species, without prior information about the sampling location. For the mixture analysis, the maximum number of genetically diverse groups was set to 48. The resulting mixture clustering was used for an admixture analysis with a minimum population size of 3, 100 interactions, 200 reference individuals from each population and 20 iterations for the reference individuals. Moreover, genetic substructuring within each species was assessed through independent analyses of the split data matrices following the same procedures described above (the maximum number of groups was set to the number of analyzed populations: 33 for *L. maritima* and 15 for *L. acerifolia*).
In order to investigate the phylogenetic relationship among the genetic groups of each species, we employed the coalescent based SVDquartets method (Chifman & Kubatko, 2014) implemented in PAUP v4 (Swofford, 2003), using *Navaea phoeniceae* as outgroup.

Environmental data

To accurately quantify the environmental niche overlap between the two species, a set of 19 bioclimatic layers obtained from WorldClim (www.worldclim.org, Hijmans *et al.*, 2005)was used for extracting ecological values from all sampled locations. Furthermore, three non-climatic variables were considered in this study: mean annual values of solar radiation (kW/m2, CCAFS, 2104), subsoil pH (30-100 cm, FAO *et al.*, 2012) and slope extracted from the altitude layer of WorldClim. These non-climatic layers were added based on our knowledge of the specific ecological requirements of these species, which limit their presence to particular habitats (see, e.g., Chapter 2). All environmental layers were converted to the same resolution of 1 km^2.

Cross-similarity between species distributed modelling (SDM)

In order to assess the spatial distribution of the climate suitability of *L. maritima* and *L. acerifolia* under current time and one past scenario (Last Interglacial period, 130-

115 kyr, LIG), two analyses were carried out using Maxent v3.3 (Phillips, Anderson & Schapire, 2006). This software works better than other methods for studies with a small sample size and only with presence data (Elith *et al.*, 2006). The distribution of climate suitability for *L. maritima* along the western Mediterranean region, including the Canary Islands, was described in Chapter 1.

For *L. acerifolia*, analyses were performed using a set of 19 bioclimatic variables from WorldClim, which were clipped to consider the Canary Islands, North African continent and western Mediterranean basin as model calibration area. The occurrence dataset comprised 17 presences points of the species (Table 1). The species distribution model for *L. acerifolia* was projected to the distribution area of *L. maritima* for exploring suitable areas over the continent. The occurrence data were randomly split into training (75%) and test (25%) data for model evaluation. Ten subsample replicates were performed and the species distribution model was constructed with the average prediction from all these model replicates.

To represent a binary map of both species, it was necessary to select a probability threshold in order to classify each map pixel either as suitable or unsuitable. The threshold chosen was "equal test sensibility and specificity", in which the percent of true presences correctly classified as present (sensitivity) is equal to the percent of true absences correctly classified as absent by the model (specificity) (Bean, Stafford & Brashares, 2012; Liu *et al.*, 2005)

Niche characteristics: overlap and equivalence test

The occurrence of niche differentiation was tested by comparing models at two levels: inter- and intraspecific. For this purpose, niche overlap was calculated: 1) between the two species using 74 presence points from all sampled populations (Table 1), and 2) between pairs of genetic groups obtained from the BAPS analyses of SNP data. This environmental dataset consisted of the geographic coordinates of 100 individuals from 48 populations for which genetic data were available. The number of occurrences in one of the three genetic groups of *L. acerifolia* (K3, see Results), which consisted of only three true presence points, was increased by adding two pseudo-occurrences, in order to achieve the minimum of 5 points required by the analysis.

Niche overlap was measured in the two-dimensional environmental space applying the method of Broennimann *et al.* (2012). This method comprises three steps. First, the environmental space is built through a multivariate analysis. Because we are working with coastal species that show a nearly linear distribution range, the geographic area for each species was estimated by placing a buffer area of approximately 190 kilometers around each location. This method creates a polygon that encloses all known occurrences and assumes that it includes all suitable habitats for the species. Such buffer was built using the gBuffer function of the rgeos v. 0.3-2 package (Bivand & Rundel, 2013) in R (R Core Team, 2015). The environmental values from 19 bioclimatic plus 3 non-climatic variables were extracted for each sample record and for 10000 random background points that were selected inside of this area. The environmental space was built from all these points via principal component analysis (PCA), reducing the number of environmental variables to two axes that are not correlated. For the niche overlap between genetic groups, the slope variable was removed due to its low contribution to the first two axes in the PCA. Second, the kernel density of occurrences was calculated along environmental axes of a PCA. The first two axes of this PCA were considered on the total environmental space sampled over the ranges of the species pair. This method makes optimal use of both geographical and environmental spaces, avoiding incorrect inferences by using only geographic predictions from the species distribution model (SDM). Third, the analysis measures the niche overlap along gradients of this PCA using I and D metrics. For this study, the D index (Schoener, 1970), which compares the occupancy of the environment for pairs of species in a given environmental space, was selected. The value of D ranges from 0, when the two species have no overlap in the environmental space, to 1 when the two species share the same environmental space.

The degree of niche overlap between genetic groups was assessed pairwisely in a retrospective manner along the SVDquartets topology starting with the most closely related groups. Once the overlap between each of these pairs was estimated, the presence data of these two groups was aggregated and the niche overlap of this aggregated group was calculated with respect to their sister genetic group. This was repeated until all the groups were added.

Finally, a statistical test of equivalency was calculated (Warren *et al.*, 2008). The niche equivalence test was used to assess if the bioclimatic niches of these sister species are significantly different from each other and if the niche spaces of the first species predict those of the second. A null distribution was generated by using permutations of the environmental scores obtained for each species record at the PCA axes and calculating the D index for each permutation. This process was repeated 100 times. The equivalence of environmental niche was accepted if the observed value of D fell into the density of 95% of the simulated values.

Niche overlap analyses were implemented using the ecospat v.2 package (Broennimann *et al.*, 2014). In order to interpret niche overlap for each axis of the PCA, the function niceOverPlot (see code in supporting information) was implemented in R environment.

Environmental variables influencing the niche of species

To identify the factors behind the niche differences, analyses of variance (ANOVA) were conducted on the environmental variables extracted in the occurrence localities of each species. With these analyses, we tested whether the observed environmental conditions differ significantly between the two species or not. Only the most influential environmental variables for each axis of the PCA were analyzed. A probability density plot of these non-correlated variables was used for visualizing their behavior of each species in the niche.

RESULTS

Genotyping by sequencing

A total of 126 325 884 reads were identified from 28 individuals belonging to 15 populations of *L. acerifolia* and 70 individuals from 33 populations of *L. maritima*. The resulting matrix recovered 3629 SNPs. One individual from TIN was the sample of *L. maritima* with the highest number of SNPs (3624) whereas ZEG had the lowest (2405 SNPs). Among the populations of *L. acerifolia,* one individual from GUIN had the highest number of SNPs (3463) whereas GUA had the lowest (2497).

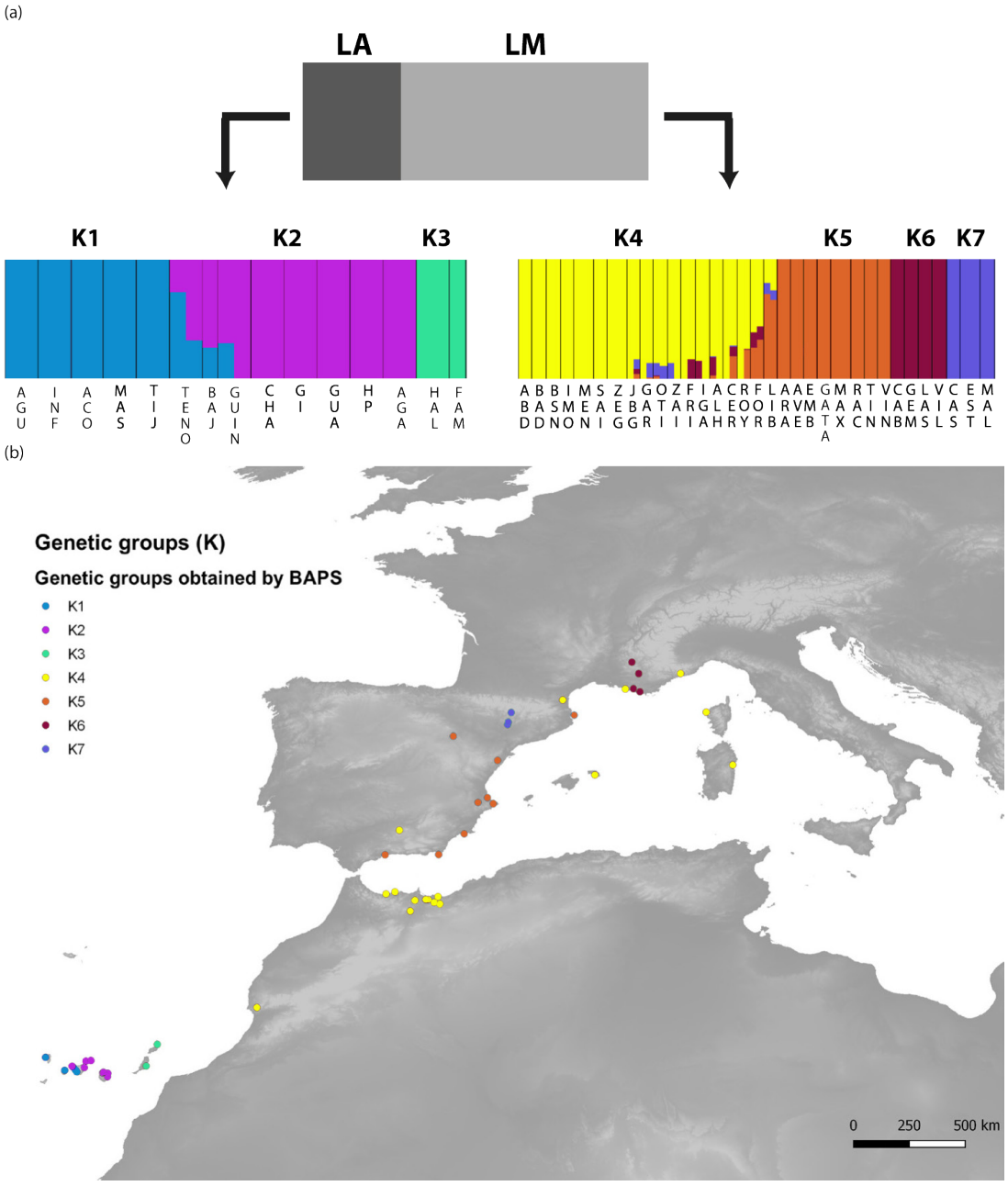


Figure 1. Bayesian clustering of genetic groups from 48 populations of *Lavatera maritima* and *Lavatera acerifolia* based on SNPs recovered from genotyping-by-sequencing (GBS) data. (a) Two genetic subgroups (LA and LM) are detected by BAPS from all populations. The optimal partition for each subgroup was K=3 and K=4 for LA and LM, respectively. Samples are represented by rectangles where the color indicates the probability of each sample to belong to a determined genetic group. (b) Distribution map of the seven genetic groups obtained by BAPS for both species.

Genetic structure of *Lavatera maritima-L. acerifolia*

The analysis of these 3629 SNPs from all populations in BAPS detected two genetic subgroups corresponding to the two species (K=2, LM and LA; Fig. 1): subgroup LA included 15 populations from *L. acerifolia* whilst subgroup LM was formed by 33 populations from

L. maritima. The genetic structure analysis of the subgroup LA identified three different groups within (K=3) whereas subgroup LM was divided into four genetic groups (K=4). The three genetic clusters within *L. acerifolia* subgroup (K1, K2, K3) were roughly distributed following an east-west pattern along the Canary archipelago. The results for *L. acerifolia* (Fig. 1) only showed admixture in three populations (Bajamar, Guiniguada and Teno, Table 1) involving the genetic clusters K1 and K2. K1 harboured several populations from the western Canary Islands: Agulo (GO), Barranco del Infierno (TF), Acojeja (TF), Masca (TF), Teno (TF, one individual) and Tijarafe (LP). K2 included populations from the central islands: Bajamar (TF), Chamorga (TF), Güimar (TF), Guayadeque (GC), Guiniguada (GC), Hoyo Pineda (GC), Teno (TF, one individual) and Agaete (GC). K3 comprised two populations from the eastern Islands: Halconcillo (FTV) and Famara (LZT) (Table 1).
The genetic structure analysis of the *L. maritima* subgroup (K4, K5, K6, K7; Fig. 1) identified three populations with admixture (Cap de Norfeu, Lliber and Otiñar, Table 1) and other populations in which one of their samples showed low admixture (Alhucema, Cerdeña, Ile de Ratonneau, Gareb, Ille de Gargalu, Jbel Guilliz, Col de l'Arma and Zaio, Table 1). K4 encompassed populations widely distributed along the western Mediterranean including all populations from Morocco (West and North), the Mediterranean continental islands, northern Spain and southern France; K5 comprised only Iberian populations; K6 comprised four French populations and K7 was formed by three populations from the Ebro valley.

The topology of the species tree obtained by the SVDquartets method (Fig. 2) showed two clearly supported clades (100% BS) each corresponding to one of the sister species. The *L. acerifolia* clade shows the three genetic groups (K1, K2 and K3), K2-K3 forming a subclade sister to K1. The *L. maritima* clade included four genetic groups, where K6 is sister to the remaining (K4, K5, K7), and K5-K7 form a supported subclade (98% BP).

Climate suitability modelling

The geographical distribution of the climate suitability of *L. acerifolia* along the Mediterranean region under current conditions showed an AUC value of 0.99 and encompassed its natural distribution in the Canary Islands, including all the sampled

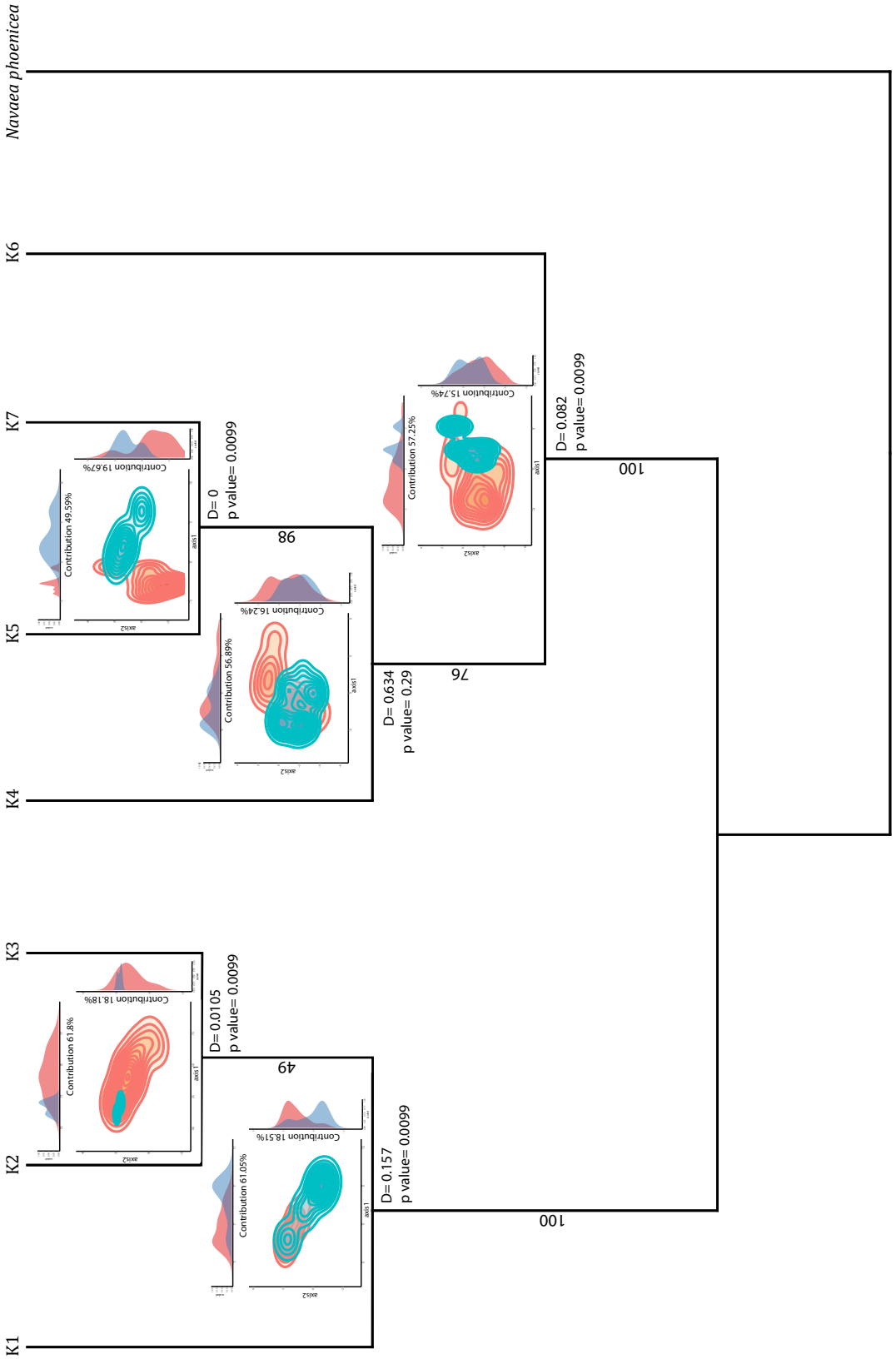


Figure 2. Species tree showing topological relationships among genetic groups found in 48 populations of *Lavatera maritima* and *Lavatera acerifolia* constructed with SDVquartets, using *Navaea phoenicea* as outgroup. Bootstrap support values are shown above the branches. Both niche overlap measure (D) among different genetic groups and the result from the equivalency test are shown for each node.

localities (Fig. 3). The predicted areas comprised the seven islands of the archipelago, including El Hierro, where *L. acerifolia* is absent. In addition, the analysis identified additional suitable areas such as the Madeira Islands as well as small continental regions that exhibited a high probability in the western coast of Morocco (Essaouira and Tarfaya) and coastal Portugal (Peniche).

The projected geographic distribution on the LIG was much more reduced compared with the current distribution. The analysis only predicted a high probability of climate suitability in the Madeira Islands and the same three reduced areas in the continent. The Canary Islands did not show an adequate probability of climate suitability.

The distribution map of the climate suitability for both species under both current and past conditions is showed in Fig. 3. In the present time, there are several regions where *L. maritima* and *L. acerifolia* share a high probability of climate suitability: the coast Portuguese area around Peniche where none of the two species currently occur, Essaouira (coastal Morocco) where *L. maritima* currently occurs albeit very scarcely possibly as a consequence of severe range reduction and genetic bottlenecks (see Chapter 1), and the Canary Islands, except for Lanzarote and Fuerteventura, where *L. acerifolia* appears. In addition, the Madeira Islands showed a high probability of climate suitability for both species. For the LIG scenario, only Essaouira and the Madeira Islands displayed a high probability of climate suitability for both species.

Environmental niche differentiation between species

The first two axes of the PCA analysis used for building the environment space together explained 69.83% of the total variance (Fig. 4a, c). The PC1 (54.36%) represented a gradient between wetter and drier climates. Variables such as annual total precipitation (bio12) and precipitation of the warmest and driest quarter (bio18 and bio17, respectively) contributed positively to this axis on the wetter end of the gradient whereas annual mean temperature (bio1) and mean temperature of the warmest quarter (bio10) contributed negatively to this axis on the drier opposite end (Fig. 4c). The PC2 (15.74%) seems to reflect seasonality and continentality since variables such as temperature seasonality (bio4) and temperature annual range values (bio7) contributed positively to this axis whereas minimum temperature of the coldest month (bio6) contributes negatively although not as clearly as the other two variables (Fig. 4c).

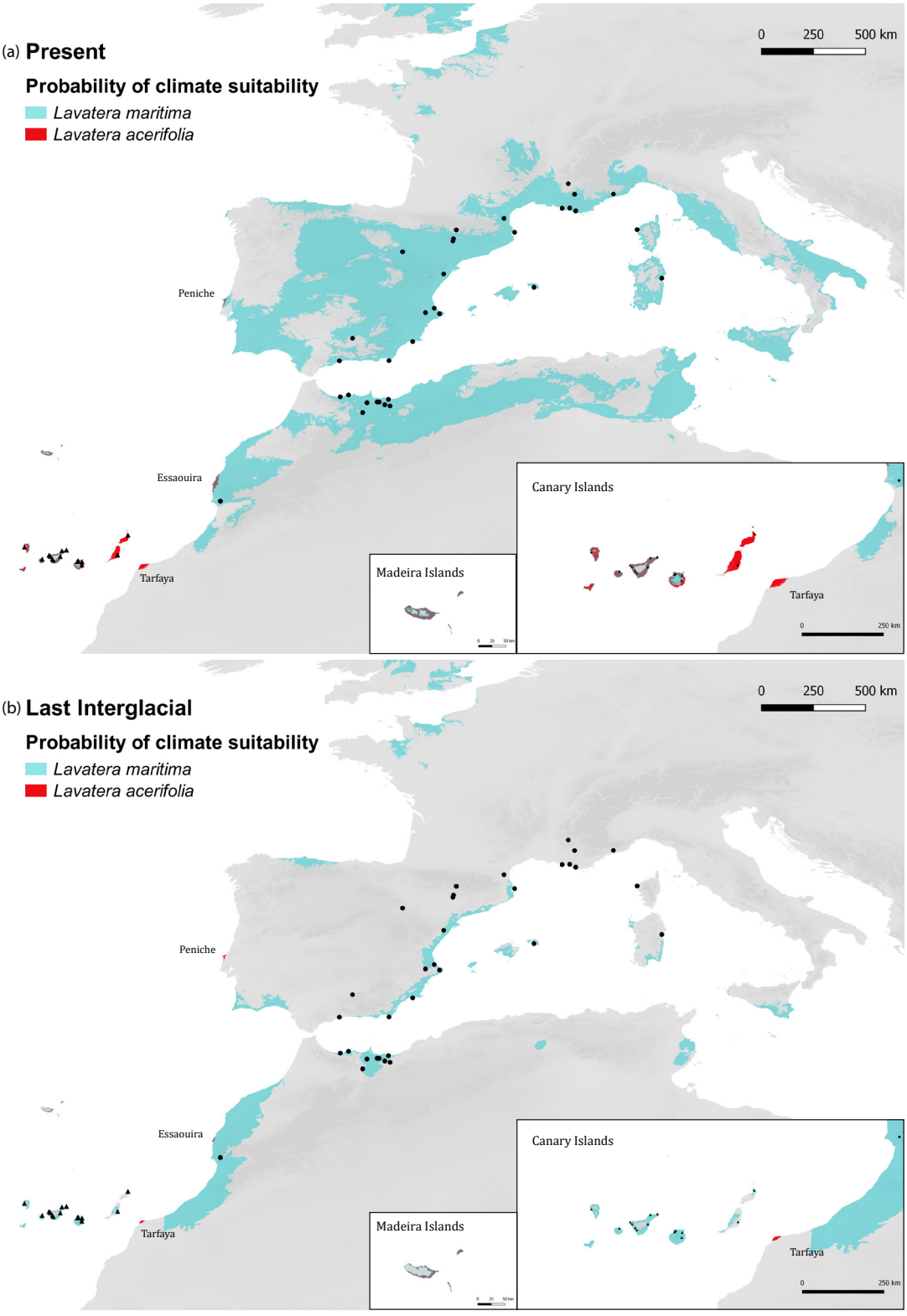


Figure 3. Climate suitability models calculated using Maxent for *Lavatera maritima* and *Lavatera acerifolia*, (a) under current and (b) past conditions (LIG, 130-115 ka). Dots indicate sampled populations of *L. maritima* and triangles represent those of *L. acerifola*. Darker areas indicate the overlap of environmental niches of both species.

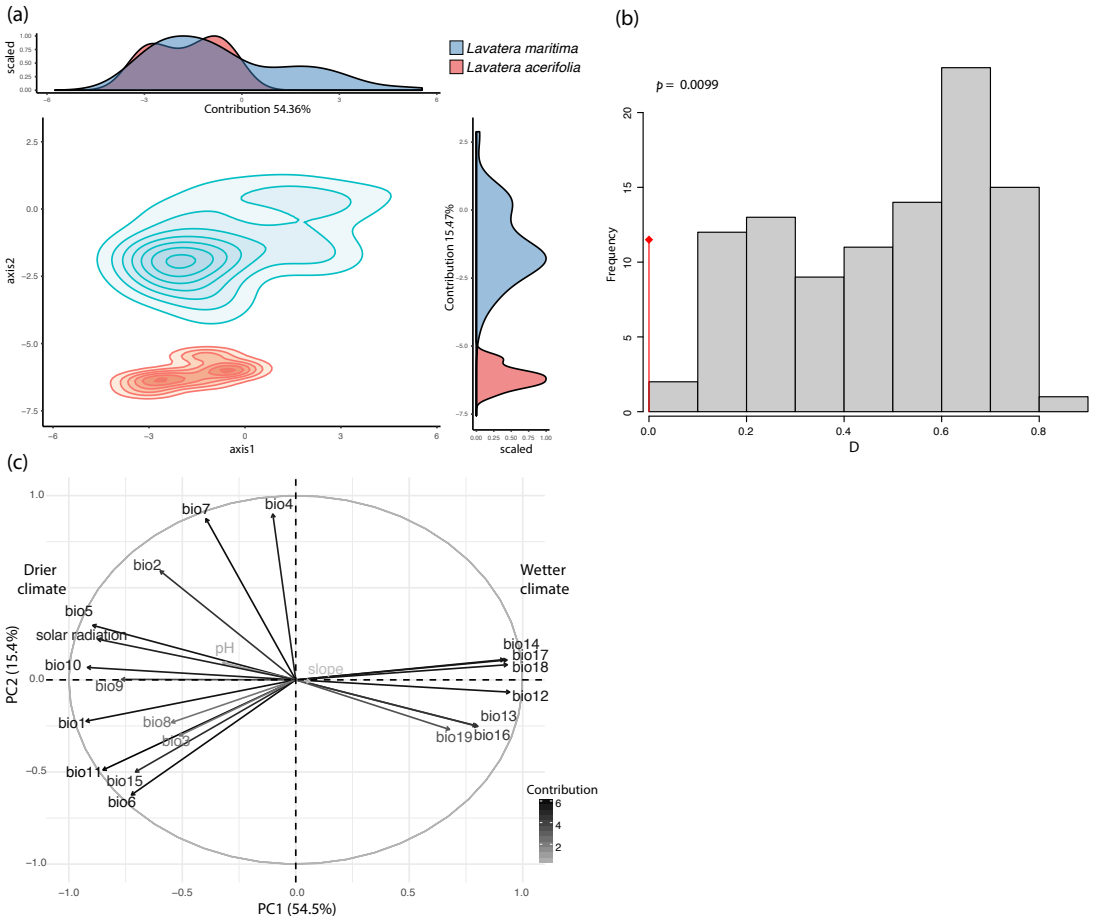


Figure 4. Environmental niche of *Lavatera maritima* and *Lavatera acerifolia* in the climatic space created through a principal component analysis (PCA). (a) Observed niche of both species along the two first axes of the PCA. Dark shading shows the density of the species occurrences by cell. The graphic on the top represents the resulting axis 1 from the PCA while the one on the right denotes the results of the PCA for the axis 2. (b) Equivalency test. Histogram showing the observed niche overlap D (red line) and simulated niche overlap (grey bars), obtained from tests of niche equivalency calculated with 100 iterations. (c) Contribution of the climatic variables on two axes of the PCA. The PC1 (axis x) illustrates a gradient between wetter and drier climates while the PC2 (axis y) describes a seasonality gradient.

The assessment of bioclimatic niche overlap using the first two axes of the environmental space revealed niche differentiation between these two sisters species (D=0, non-overlapping) (Fig. 4a). When environmental niche overlap is assessed separately for each axis, our analyses show that the two species share the same environmental space for axis 1 (54.36% of the variance), whereas for axis 2 (15.74% of the variance) there is no overlap between them in this bioclimatic scenario. The niche equivalence test was rejected (P-value=0.0099, Fig. 4b), supporting the conclusion that the bioclimatic niches of these species are not equivalent.

The most influential variables in the bioclimatic niche of both species were bio12 (annual precipitation) for axis 1 (overlapping values for the two species) and bio4 (temperature seasonality) for axis 2 (non-overlapping values for the two species) (Fig. 4c). The ANOVA showed significant differences between the values of these variables for each species (P<0.001) (Fig. 5a, b), indicating that overall each species has its own bioclimatic preferences. For bio4 (temperature seasonality) *L. maritima* presented higher values than *L. acerifolia* (Fig. 5a), whilst for bio12 (annual precipitation), the range of values were similar in *L. maritima* and *L. acerifolia* as indicated by their overlap along axis (Fig. 5b). Accordingly, *L. maritima* occurs in regions with high temperature seasonality and high annual precipitation, and *L. acerifolia* occupies areas with narrow range of temperatures, fitting with isothermal areas as oceanic islands.

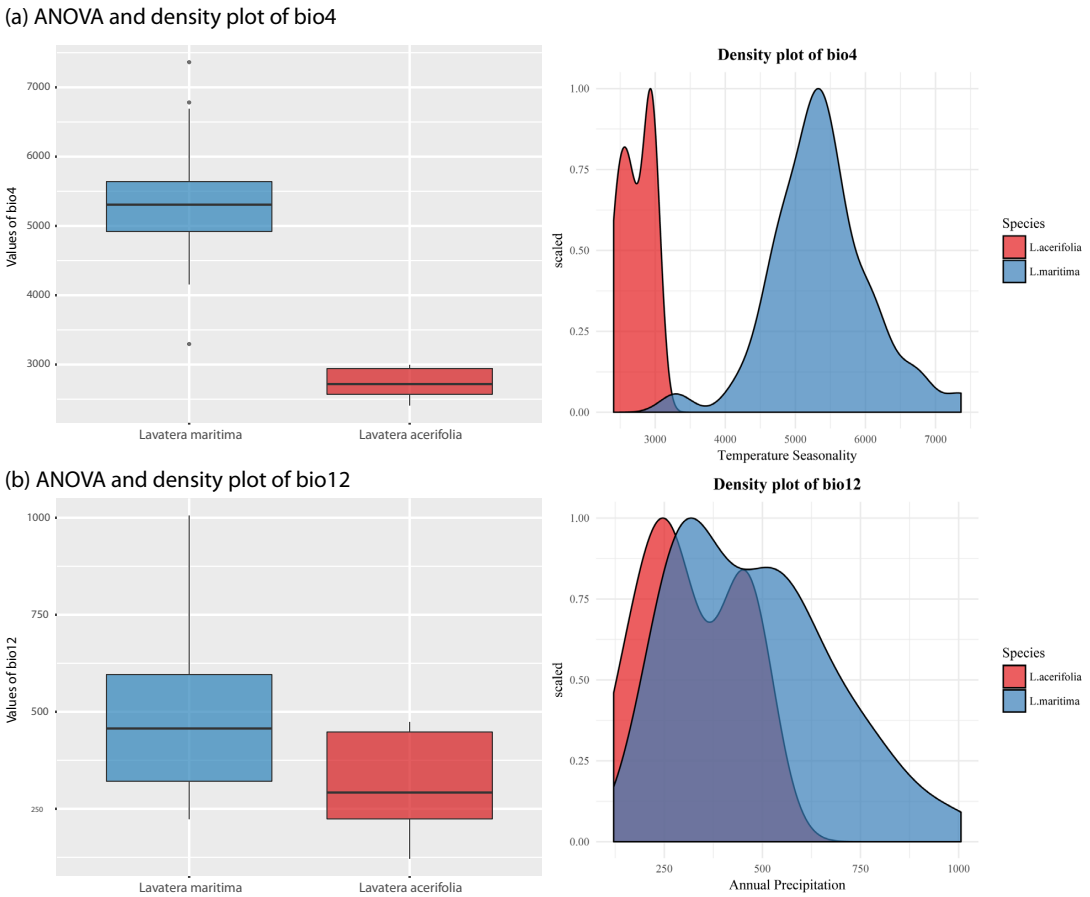


Figure 5. Exploratory analyses of the most influential variables on the niches of *Lavatera maritima* and *Lavatera acerifolia*. (a) ANOVA result and density plot of bio4 (temperature seasonality), and (b) ANOVA result and density plot of bio 12 (annual precipitation) for both species.

Environmental niche differentiation within species

Results concerning niche similarity between genetic groups were consistent with the lack of overlap between species. The results of pairwise comparisons among K1, K2 and K3 spanned the entire bioclimatic niche occupied by *L. acerifolia*, showing two different trends, one for western islands (K2, K3) and another for eastern islands (K1) (Fig 2). The result of pairwise K2-K3 analysis in which the equivalency test was rejected, was biased probably due to the sample size. The little K3 environmental niche space was totally included in the environmental space of K2 (Fig. 2).

For *L. maritima*, the K4 vs. (K5+K7) pairwise analysis that include the greatest number of populations, was the only case in which the niches were considered as equivalent (D=0.403, p-value=0.30, Fig. 2). The K5-K7 pair comparison did not reveal equivalent niches (Fig. 2). The bioclimatic niche occupied by K6 genetic group was significantly different from the rest of the genetic groups (K6 vs. (K4+K5+K6)), although this result might also be biased by the sample size.

DISCUSSION

A connection between Canary and Mediterranean lineages

Despite the occurrence of back-colonization from oceanic islands to the continent (Carine *et al.*, 2004; Caujapé-Castells, 2011) the onset of oceanic island lineages is always successful colonization events from continental species. There is abundant evidence from phylogenetic studies that Canarian lineages have their closest relatives in mainland Africa or the Mediterranean basin (Carvalho & Culham, 1998; Francisco-Ortega *et al.*, 1997; Helfgott *et al.*, 2000; Mairal *et al.*, 2015b; Rodríguez-Sánchez & Arroyo, 2008). But fine documentation of the closest relative of an island species in the continent is far less frequent (but see Talavera *et al.*, 2013), which makes the *L. maritima-L. acerifolia* species pair a most suitable case in which their genetic, biogeographic, and environmental relationships can help understand the connections between Mediterranean and Macaronesian floras.

The joint analysis of the SNP data for the two species reveals a clear genetic structure and is consistent with the independent analysis of *L. acerifolia* presented in

Chapter 2 (Fig. 2). Specifically, the eastern and western genetic groups of *L. acerifolia* are distinct and the populations from the central islands also form a group although some populations from Tenerife (Teno and Bajamar) and Gran Canaria (Guiniguada) showed admixture between groups K1 and K2 (Fig. 1). These three populations were recognized under one of the two bioinformatics pipelines as a fourth genetic group albeit a scarcely differentiated one (Chapter 2). The results concerning *L. maritima* reveal four genetic groups along the western Mediterranean basin (Fig. 1). Two of them (k4 and k5) are widely distributed, K4 being the most frequent while K5 only harbors populations from the Iberian Peninsula. Four populations from France (Cabasse, Gémenos, Grand Vallon, Roche amère) form one genetic group (K6), and three populations from the inland northern Spain (Castillonroy, Estopiñán del Castillo and Mallos de Riglos) encompass the fourth group (K7). The distribution of these genetic groups obtained by GBS data is congruent with the phylogeographic study of this species based on plastid DNA sequences (Chapter 1).

Feasibility of successful long-distance colonization depends on several factors including availability of dispersal agents sometimes, but not always, combined with specific adaptations in fruits or seeds (Arjona *et al.*, 2017) and niche availability or, more specifically, preadaptation of the dispersed genotypes allowing them to survive and become established in the environmental conditions of the colonized areas (Piñeiro *et al.*, 2007). In our case study there are several lines of evidence supporting the hypothesis that *L. acerifolia* originated from colonization of the eastern Canary Islands by *L. maritima*. First, *L. maritima* has been inferred to have successfully dispersed more than once from North Africa into Europe (Chapter 1), suggesting that even in the absence of known adaptations for dispersal in its mericarps, it has a capacity for long distance dispersal and establishment. Second, we have discovered a clear east-west pattern of colonization along the Canary Islands in *L. acerifolia* (Chapter 2). Third, SDM have found environmental suitable areas for *L. maritima* both in present time and in past projections (LGM and LIG) in coastal areas in western Morocco as well as in the Canary Islands. In fact, during the LIG the areas with the highest environmental suitability were this Moroccan area together with the Canaries (Chapter 1). Although our time-calibrated

phylogenetic analysis (Chapter 1) suggests that *L. acerifolia* diverged from *L. maritima* in the Plio-Pleistocene (2.77(0.43-5.75) Myr), the steady environmental suitability in those two areas across three different time scenarios suggest that suitable environment could have been projected deeper into older times. In fact, such a high suitability in past scenarios for *L. maritima* both in the Canaries and in western coastal Morocco holds for climatically very contrasting periods (LGM vs. LIG). Under this reasoning, we could infer that there were environmentally suitable and colonizable areas in the Canaries for *L. maritima*, at a relative geographical proximity, over different past time-frames. Fourth, the environmental niche of *L. acerifolia* is much narrower than that of *L. maritima* (see below; Fig. 4) and this is consistent with a founder event represented by plants dispersing from a geographically widespread and environmentally broader species such as *L. maritima*. In addition, we have detected hints of *L. maritima* capacity for enlarging its environmental niche towards inland areas in the Iberian Peninsula, albeit, compared to the Canary Islands climate, in a different direction of increased continentality (see Chapter 1).

Thus, it can be reasonably concluded that *L. acerifolia* originated from colonization of the eastern Canary Islands by *L. maritima*. However, there are aspects of the evolutionary relationships between these sister species that require further discussion.

Niche differentiation between sister species

Both the continental and the oceanic species have preferences for regions with Mediterranean climate, that is, warm and dry summers combined with cold and wet winters. However, *L. acerifolia* occurs in regions with lower annual precipitation than *L. maritima* and exhibited a narrow range for this value (Fig. 5). Also, the two species show differences concerning seasonality (PC2 axis; Fig. 4a, c). This factor represents the main difference in their environmental niche spaces. *Lavatera maritima* occupies regions with high temperature seasonality and even continentality, specifically, places with cold winters and warm summers whilst *L. acerifolia* occurs in areas where the temperature is constant throughout the year (Fig. 5). The woodiness and glabrous organs in *L. acerifolia* could be related to the more uniform and lower high temperatures throughout the year compared to *L. maritima* (Comes, 2004). Another result highlighting the contrast

between the climatic niches of the two species, and the narrowness of *L. acerifolia* niche is that climate suitability for this species is extremely restricted, with only minor suitable areas outside the Canaries (Fig. 3). In sharp contrast, the climate suitability distribution for *L. maritima* is not only much more extended, even reaching northern France and southern England, but it also includes the Canary Islands. In the context of *L. maritima* capacity for dispersal (Chapter 1), the absence of this species in the Canary archipelago might be due to competition with *L. acerifolia,* although we are not addressing this issue in the present study.

Still another element contributing to define a broader environmental niche for *L. maritima* is the finding of a visible expansion of the environmental niche in populations surrounding the Ebro Valley in Northern Spain. These represent exceptions to the subcoastal habitats where this species normally occurs and, based on the restricted distribution of these populations, it is likely that such an expansion occurred recently. These populations occupy a different bioclimatic more continental niche compared to the remaining populations of this species, which is determined by the precipitation of the driest month (bio14) that contributed positively to this axis 1and the maximum temperature of the warmest (bio5) that contributed negatively to this axis (Fig. S1). Therefore, *L. maritima* exhibits a broader climatic niche which is only partly overlapped with that of *L. acerifolia*, it shows differences across its range for some variable and it might have the capacity for adapting to substantially different climatic regimes represented by inland continental enclaves. Thus, the environmental niche analyses are also consistent with the hypothesis that *L. maritima* shifted its niche along with or after the colonization of the Canary Islands.

Implications for the speciation scenario of *Lavatera acerifolia*

The combination of the evidence from niche modelling, including niche overlap, and GBS genetic data suggest that both geographical and ecological factors played a role in the divergence of the two sister species and the speciation of *L. acerifolia*. We hypothesize that colonization of the Canary Islands took place by an ancestor of *L. acerifolia*, probably from a West African population of *L. maritima*, or its ancestors. A few individuals from such ancestral population might have reached the islands through long-distance

dispersal and differentiated from *L. maritima* via founder effects and adaptation to a new and narrower niche. The environmental niche differentiation between the two species here detected (Fig. 4) might be interpreted as a proxy for a divergence of *L. acerifolia* from *L. maritima* via ecological speciation. However, because there are minimal suitable areas for *L. acerifolia* outside the Canaries both in the present time and during the LIG (Fig. 3), we think that it is unlikely that *L. acerifolia* niche evolved during the course of ecological speciation and that occupying a different niche was the primary mechanism for reproductive isolation. The hypothesized scenario, including a founder event into an isolated –Island– area, involves abrupt geographical isolation that gradually allowed reproductive barriers to develop. This fits an allopatric speciation model and specifically, because of the small inferred size of the founder population, a peripatric model.

Notwithstanding, the occurrence of niche differentiation suggests that ecological factors were also important over the whole speciation process. We infer that niche evolution took place after the establishment of founder populations in the Canaries. Considering the variables that determine the niche differences between the two species, it is likely that niche shift involved less modifications of the biochemical and physiological machinery than the opposite shift, i.e., from a more uniform temperature regime to a seasonal and even continental one, as seen in the inland populations of *L. maritima* from the surroundings of the Ebro Valley (Chapter 1, Wiens & Donoghue, 2004). It is thus conceivable that adaptation of *L. maritima* inmigrants to the current *L. acerifolia* niche in the Canaries was an evolutionary feasible and relatively effortless process. Along this line, we could speculate that the polyploid nature of *L. maritima* and its ancestors could have facilitated adaptability to their genomes and ultimately climatic niche flexibility, which is not always found when studying niche overlap in closely related species (Chozas *et al.*, 2017).

In sum, we conclude that it is likely that *L. acerifolia* speciated in the Canary Islands following a peripatric model, after a successive colonization event and adapted to the new niche subsequently in isolation.

REFERENCES

Arjona Y, Nogales M, Heleno R, Vargas P. 2017. Long-distance dispersal syndromes matter: diaspore–trait effect on shaping plant distribution across the Canary Islands. *Ecography*.

Bean WT, Stafford R, Brashares JS. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography* 35: 250-258.

Bivand R, Rundel C. 2013. rgeos: interface to geometry engine-open source (GEOS). *R package version 0.3-2*.

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, Thuiller W, Fortin M-J, Randin C, Zimmermann NE, Graham CH, Guisan A. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* 21: 481-497.

Broennimann O, Petitpierre B, Randin C, Engler R, Breiner F, DAmen M, Pellissier L, Pottier J, Pio D, Mateo R. 2014. ecospat: Spatial ecology miscellaneous methods. *R package version* 1.

Burns JH, Strauss SY. 2011. More closely related species are more ecologically similar in an experimental test. *Proceedings of the National Academy of Sciences* 108: 5302-5307.

Butlin RK, Galindo J, Grahame JW. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363: 2997-3007.

Carine MA, Russell SJ, Santos-Guerra A, Francisco-Ortega J. 2004. Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in *Convolvulus* (Convolvulaceae). *American Journal of Botany* 91: 1070-1085.

Carvalho JA, Culham A. 1998. Conservation status and preliminar results on the phylogenetics of *Isoplexis* (Lindl.) Benth.(Scrophulariaceae) an endemic Macaronesian genus.

Caujapé-Castells J. 2011. Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. *The biology of island floras*: 284-324.

Caujapé-Castells J, García-Verdugo C, Marrero-Rodríguez Á, Fernández-Palacios JM, Crawford DJ, Mort ME. 2017. Island ontogenies, syngameons, and the origins and evolution of genetic diversity in the Canarian endemic flora. *Perspectives in Plant Ecology, Evolution and Systematics*.

CCAFS. 2104. Climate Change Agriculture and Food Security: http://www.ccafs-climate.org/data/ (accessed April 2017).

Comes HP. 2004. The Mediterranean region – a hotspot for plant biogeographic research. *New Phytologist* 164: 11-14.

Corander J, Cheng L, Marttinen P, Sirén J, Tang J. 2013. BAPS: Bayesian analysis of population structure V. 6.0. *Department of Mathematics and statistics. University of Helsinki, Finland*.

Coyne JA, Orr HA. 2004. Speciation. Sunderland, MA: Sinauer Associates, Inc.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317-3324.

Chozas S, Chefaoui RM, Correia O, Bonal R, Hortal J. 2017. Environmental niche divergence among three dune shrub sister species with parapatric distributions. *Annals of Bot*any, 119(7), 1157-1167

Díaz-Pérez A, Sequeira M, Santos-Guerra A, Catalán P, Mason-Gamer R. 2008. Multiple Colonizations, In Situ Speciation, and Volcanism-Associated Stepping-Stone Dispersals Shaped the Phylogeography of the Macaronesian Red Fescues (*Festuca* L., Gramineae). *Systematic Biology* 57: 732-749.

Elith J, H. Graham C, P. Anderson R, Dudík M, Ferrier S, Guisan A, J. Hijmans R, Huettmann F, R. Leathwick J, Lehmann A, Li J, G. Lohmann L, A. Loiselle B, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC. M. Overton J, Townsend Peterson A, J. Phillips S, Richardson K, Scachetti-Pereira R, E. Schapire R, Soberón J, Williams S, S. Wisz M, E. Zimmermann N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.

Escobar P, Schönswetter P, Fuertes Aguilar J, Nieto Feliner G, Schneeweiss GM. 2009. Five molecular markers reveal extensive morphological homoplasy and reticulate evolution in the *Malva* alliance (Malvaceae). *Molecular Phylogenetics and Evolution* 50: 226-239.

FAO, IIASA, ISRIC, ISSCAS, JRC. 2012. Harmonized World Soil Database v 1.2: http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/.

Fernández-Palacios JM, de Nascimento L, Otto R, Delgado JD, García-del-Rey E, Arévalo JR, Whittaker RJ. 2011. A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography* 38: 226-246.

Ferreira MZ, Zahradní, ek J, Kadlecová J, de Sequeira MM, Chrtek Jr J, ich, Fehrer J. 2015. Tracing the evolutionary history of the little-known Mediterranean-Macaronesian genus *Andryala* (Asteraceae) by multigene sequencing. *Taxon* 64: 535-551.

Francisco-Ortega J, Santos-Guerra A, Hines A, Jansen R. 1997. Molecular evidence for a Mediterranean origin of the Macaronesian endemic genus *Argyranthemum* (Asteraceae). *American Journal of Botany* 84: 1595-1595.

García-Aloy S, Vitales D, Roquet C, Sanmartín I, Vargas P, Molero J, Kamau P, Aldasoro JJ, Alarcón M. 2017. North-west Africa as a source and refuge area of plant biodiversity: a case study on *Campanula kremeri* and *Campanula occidentalis*. *Journal of Biogeography*. DOI: 10.1111/jbi.12997

Garnatje T, Garcia S, Canela MÁ. 2007. Genome size variation from a phylogenetic perspective in the genus *Cheirolophus* Cass. (Asteraceae): biogeographic implications. *Plant Systematics and Evolution* 264: 117-134.

Helfgott DM, Francisco-Ortega J, Santos-Guerra A, Jansen RK, Simpson BB. 2000. Biogeography and breeding system evolution of the woody *Bencomia* alliance (Rosaceae) in Macaronesia based on ITS sequence data. *Systematic Botany* 25: 82-97.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.

Liu C, Berry PM, Dawson TP, Pearson RG. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385-393.

Mairal M, Pokorny L, Aldasoro JJ, Alarcón M, Sanmartín I. 2015a. Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). *Molecular Ecology* 24: 1335-1354.

Mairal M, Sanmartin I, Aldasoro JJ, Culshaw V, Manolopoulou I, Alarcon M. 2015b. Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic *Canarina canariensis* (Campanulaceae). *Molecular Ecology* 24: 3944-3963.

Mayr E. 1942. *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Melo AT, Bartaula R, Hale I. 2016. GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC bioinformatics* 17: 29.

Nakazato T, Warren DL, Moyle LC. 2010. Ecological and geographic modes of species divergence in wild tomatoes. *Am J Bot* 97: 680-693.

Peterson A, Soberón J, Sánchez-Cordero V. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285: 1265-1267.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.

Piñeiro R, Aguilar JF, Munt DD, Feliner GN. 2007. Ecology matters: Atlantic–Mediterranean disjunction in the sand-dune shrub *Armeria pungens* (Plumbaginaceae). *Molecular Ecology* 16: 2155-2171.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rodríguez-Sánchez F, Arroyo J. 2008. Reconstructing the demise of Tethyan plants: climate-driven range dynamics of Laurus since the Pliocene. *Global Ecology and Biogeography* 17: 685-695.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor laboratory press New York.

Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323: 737-741.

Schoener TW. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology* 51: 408-418.

Soberón J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115-1123.

Swofford DL. 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4.

Talavera M, Navarro-Sampedro L, Ortiz PL, Arista M. 2013. Phylogeography and seed dispersal in islands: the case of *Rumex bucephalophorus* subsp. *canariensis* (Polygonaceae). *Annals of Botany* 111: 249-260.

Warren DL, Glor RE, Turelli M, Funk D. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62: 2868-2883.

Weingartner E, Wahlberg N, Nylin S. 2006. Speciation in Pararge (Satyrinae: Nymphalidae) butterflies – North Africa is the source of ancestral populations of all *Pararge* species. *Systematic Entomology* 31: 621-632.

Wiens JJ. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution* 58: 193-197.

Wiens JJ, Donoghue MJ. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* 19: 639-644.

Wiens JJ, Graham CH. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics* 36: 519-539.

SUPPORTING INFORMATION

Code. R script for NiceOverPlot function. Also, the code is available in the web page: http://allthiswasfield.blogspot.com.es/2017/05/niceoverplot-or-when-number-of.html

```
# niceOverPlot function is based on this two posts:
# http://stackoverflow.com/questions/20474465/using-different-scales-as-fill-based-on-factor
# http://rforpublichealth.blogspot.com.es/2014/02/ggplot2-cheatsheet-for-visualizing.html

# niceOverPlot function can be used in several ways. See example above to learn the basic use.
# Different approaches will be posted as soon as possible.

niceOverPlot<-function(sc1,sc2=NULL,n1=NULL,n2=NULL, plot.axis = TRUE, bw = NULL, b=NULL, a1cont=NULL, a2cont=NULL){

  # prepare the data, depending of the type of input ("pca"/"dudi" object or raw scores)
  if (is.null(sc2))
  {sc_1<-sc1
  sc_2<-sc1
  sc1<- sc_1$li[1:n1,]
  sc2<- sc_1$li[(n1+1):(n1+n2),]
  }

  if (class(sc1)==c("pca","dudi") && class(sc2)==c("pca","dudi"))
  {sc_1<-sc1
  sc_2<-sc1
  sc1<- sc1$li
  sc2<- sc2$li}

  # recognize both species
  scores<-rbind(sc1,sc2)
  g<-c(rep(0,nrow(sc1)),rep(1,nrow(sc2)))
  df<-data.frame(cbind(scores$Axis1,scores$Axis2,g))
  names(df)<-c("x","y","g")
  df$g<-as.factor(df$g)

  # establish an empty plot to be placed at top-right corner (X)
  empty <- ggplot()+geom_point(aes(1,1), colour="white") +
    theme(
      plot.background = element_blank(),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      panel.border = element_blank(),
      panel.background = element_blank(),
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      axis.text.x = element_blank(),
      axis.text.y = element_blank(),
      axis.ticks = element_blank()
    )
  # sp1
  p1 <- ggplot(data = df, aes(x, y,color = as.factor(g))) +
    stat_density2d(aes(fill = ..level..), alpha = 0.2, bins=b, geom = "polygon", h=c(bw,bw)) +
    scale_fill_continuous(low = "#fdae61", high = "#d7191c", space = "Lab", name = "sp1") +
    scale_colour_discrete(guide = FALSE) + scale_x_continuous(name = "axis1", limits= c(min(df$x)-100, max(df$x)+100))+
    scale_y_continuous(name = "axis2", limits= c(min(df$y)-100, max(df$y)+100))+
    theme(legend.position="none")
  # sp2
  p2 <- ggplot(data = df, aes(x, y, color = as.factor(g))) +
    stat_density2d(aes(fill = ..level..), alpha = 0.2, bins=b, geom = "polygon", h=c(bw,bw)) +
    scale_fill_continuous(low = "#abd9e9", high = "#2b83ba", space = "Lab", name = "sp2") +
    scale_colour_discrete(guide = FALSE) +  scale_x_continuous(name = "axis1", limits=c(min(df$x)-100, max(df$x)+100))+
    scale_y_continuous(name = "axis2", limits= c(min(df$y)-100, max(df$y)+100))+
    theme(legend.position="none")


  pp1 <- ggplot_build(p1)
  ppp1 <- ggplot_build(p1 + aes(alpha=0.15) + theme_classic() + theme(legend.position="none") +
     theme(text = element_text(size=15)) + xlab("axis1") + ylab("axis2") + xlim(c(min(pp1$data[[1]]$x)-
     0.5,max(pp1$data[[1]]$x)+0.5)) + ylim(c(min(pp1$data[[1]]$y)-0.5,max(pp1$data[[1]]$y)+0.5)))
  pp2 <- ggplot_build(p2 + aes(alpha=0.15) + theme_classic() + theme(legend.position="none")+ xlab("axis1") +
     ylab("axis2")  + xlim(c(min(pp1$data[[1]]$x)-0.5,max(pp1$data[[1]]$x)+0.5)) + ylim(c(min(pp1$data[[1]]$y)-
     0.5,max(pp1$data[[1]]$y)+0.5)))$data[[1]]

  ppp1$data[[1]]$fill[grep(pattern = "^2", pp2$group)] <- pp2$fill[grep(pattern = "^2", pp2$group)]

  grob1 <- ggplot_gtable(ppp1)
  grob2 <- ggplotGrob(p2)
  grid.newpage()
  grid.draw(grob1)

  #marginal density of x - plot on top

  if (class(sc_1)==c("pca","dudi") && class(sc_2)==c("pca","dudi"))
  {plot_top <- ggplot(df, aes(x, y=..scaled..,fill=g)) +
    geom_density(position="identity",alpha=.5) +
```

Code. (Contiuation)

```
    scale_x_continuous(name = paste("Contribution ",(round((sc_1$eig[1]*100)/sum(sc_1$eig),2)),"%",sep=""),
     limits=c(min(pp1$data[[1]]$x)-0.5,max(pp1$data[[1]]$x)+0.5))+
    scale_fill_brewer(palette = "Set1") +
    theme_classic() + theme(legend.position = "none")
  }

  else {

    if(is.null(a1cont)) plot_top <- ggplot(df, aes(x, y=..scaled..,fill=g)) +
        geom_density(position="identity",alpha=.5) +
        scale_x_continuous(name = "axis1", limits=c(min(pp1$data[[1]]$x)-0.5,max(pp1$data[[1]]$x)+0.5))+
        scale_fill_brewer(palette = "Set1") +
        theme_classic() + theme(legend.position = "none")

    else plot_top <- ggplot(df, aes(x, y=..scaled..,fill=g)) +
        geom_density(position="identity",alpha=.5) +
        scale_x_continuous(name = paste("Contribution ",a1cont,"%",sep=""), limits=c(min(pp1$data[[1]]$x)-
          0.5,max(pp1$data[[1]]$x)+0.5))+
        scale_fill_brewer(palette = "Set1") +
        theme_classic() + theme(legend.position = "none")

  }
  #marginal density of y - plot on the right

  if (class(sc_1)==c("pca","dudi") && class(sc_2)==c("pca","dudi"))
  {plot_right <- ggplot(df, aes(y, y=..scaled.., fill=g)) +
    geom_density(position="identity",alpha=.5) +
    scale_x_continuous(name = paste("Contribution ",(round((sc_1$eig[2]*100)/sum(sc_1$eig),2)),"%",sep=""), limits=
     c(min(pp1$data[[1]]$y)-0.5,max(pp1$data[[1]]$y)+0.5)) +
    coord_flip() +
    scale_fill_brewer(palette = "Set1") +
    theme_classic() + theme(legend.position = "none")
  }

  else {

    if(is.null(a2cont)) plot_right <- ggplot(df, aes(y, y=..scaled.., fill=g)) +
        geom_density(position="identity",alpha=.5) +
        scale_x_continuous(name = "axis2", limits= c(min(pp1$data[[1]]$y)-0.5,max(pp1$data[[1]]$y)+0.5)) +
        coord_flip() +
        scale_fill_brewer(palette = "Set1") +
        theme_classic() + theme(legend.position = "none")

    else plot_right <- ggplot(df, aes(y, y=..scaled.., fill=g)) +
        geom_density(position="identity",alpha=.5) +
        scale_x_continuous(name = paste("Contribution ",a2cont,"%",sep=""), limits= c(min(pp1$data[[1]]$y)-
          0.5,max(pp1$data[[1]]$y)+0.5)) +
        coord_flip() +
        scale_fill_brewer(palette = "Set1") +
        theme_classic() + theme(legend.position = "none")

  }

  if (plot.axis == TRUE) grid.arrange(plot_top, empty , grob1, plot_right, ncol=2, nrow=2, widths=c(4, 1),
    heights=c(1, 4))
  else grid.draw(grob1)

}
```

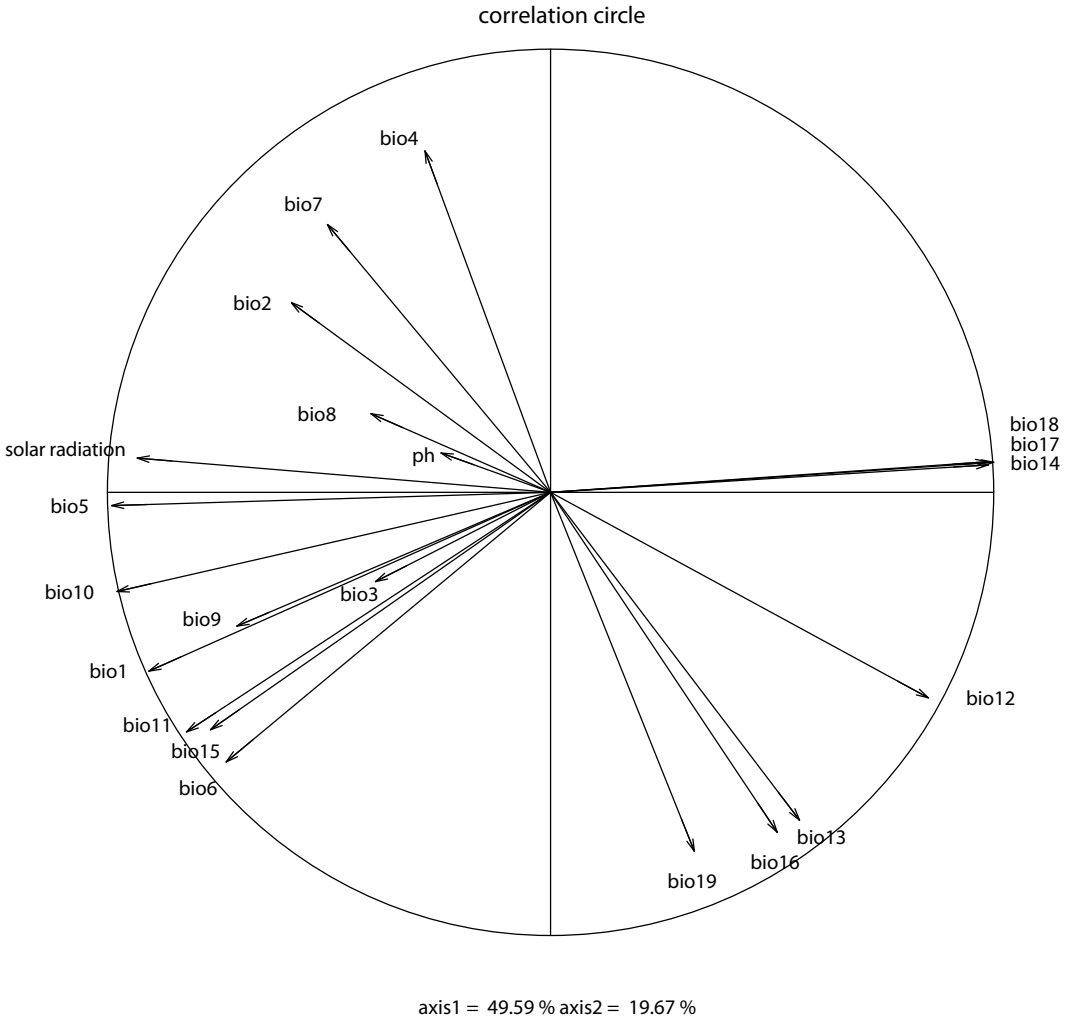


Fig S1. Contribution of the climatic variables on the two axes of the PCA for the pairwise K5-k7 analysis.

DISCUSIÓN

Esta memoria doctoral reconstruye la historia evolutiva de dos especies hermanas de angiospermas, *Lavatera maritima* y *L. acerifolia*, ambas con genomas poliploides, así como el proceso de divergencia que tuvo lugar desde su antepasado común. ¿Dónde surgieron?, ¿hacia dónde y cómo extendieron sus áreas? y ¿qué factores determinaron la divergencia de estas especies? son preguntas que se han abordado en este trabajo de investigación mediante el empleo de diferentes tipos de datos y enfoques metodológicos.

1. Historia evolutiva de *Lavatera maritima* y *L. acerifolia*

Gracias a que hemos podido documentar la relación genética, biogeográfica y ambiental entre *Lavatera maritima* y *L. acerifolia,* esta investigación puede ayudarnos a entender la conexión existente entre la flora Mediterránea y Macaronésica.

El tiempo estimado de divergencia entre ambas especies es c. 2,7 millones de años (Capítulo 1). *Lavatera maritima*, la especie mayoritariamente continental, pudo colonizar las regiones europeas desde el Norte de África –donde es muy probable que se originara— en dos eventos post-glaciares diferentes, coincidiendo con los dos linajes de ADN plastidial desvelados en el estudio filogeográfico (Capítulo 1). Los grupos genéticos identificados a partir de los SNPs, que idealmente muestran una representación amplia del genoma de cada individuo (Capítulo 3), son congruentes con los haplotipos detectados en el estudio del ADN plastidial (Capítulo 1). El primer linaje está ampliamente distribuido y abarca regiones que van desde las montañas del Rif hasta los Alpes Marítimos franceses, pasando por el este de la Península Ibérica. Además, presenta varios haplotipos locales, uno de ellos restringido en Aragón y otros en el Norte de África (Capítulo 1). La distribución de estos haplotipos muestra que la dispersión de este linaje pudo ser rápida, y siguiendo un modelo de colonización lineal a lo largo de la costa. La proyección al escenario pasado del último Máximo Glacial, donde el este de España, las montañas del Rif y el Norte de Argelia presentan alta idoneidad climática, sugiere un marco espacial y temporal adecuado para la colonización de la cuenca occidental mediterránea. Sin embargo, la proyección del modelo de distribución de especies al último Interglacial indica que las zonas con alta idoneidad quedan reducidas al SW de África e Islas Canarias, quedando prácticamente fuera el continente europeo

y las islas continentales mediterráneas. La reducción en la distribución de condiciones climáticas potencialmente idóneas durante el último Interglacial es, a primera vista, llamativa cuando estamos tratando con especies termófilas como son *L. maritima* y *L. acerifolia*. Explorando las razones, hemos comprobado que la extrema escasez de áreas con climas potencialmente idóneas para *L. maritima* en la costa oriental española, se debe a la continentalidad encontrada en el último Interglacial, que difiere sustancialmente de las condiciones encontradas durante el último Máximo Glacial y la actualidad en esa misma región. Las variables más influyentes en la proyección del modelo de distribución de especies al último Interglacial (temperatura estacional y precipitación de los cuatro meses más fríos) muestran este periodo como un escenario más húmedo y con una mayor oscilación de temperaturas máximas y mínimas entre los períodos fríos y cálidos (continentalidad). Este mismo patrón de reducción de climas potencialmente idóneos durante el último Interglacial ha sido encontrado en otras especies como *Ceratonia siliqua* L. (Viruel *et al.*, 2016) y *Linaria elegans* (Fernández-Mazuecos & Vargas, 2013).

El segundo linaje plastidial de *L. maritima* aparece en regiones del Norte de África, en localidades geográficamente distantes (Rif-SO Marruecos) y en poblaciones disyuntas (Montañas del Rif, Menorca-Córcega-Provenza y Pirineos orientales-Cerdeña) (Capítulo 1). El hecho de que el SW de Marruecos albergue haplotipos de cada uno de los dos linajes apoya la importancia de esta región en la historia evolutiva de estas especies (Capítulo 3), pudiendo haber actuado como un importante reservorio genético para *L. marítima* (Capítulo 1). Además, la diversidad haplotípica actual encontrada en las montañas del Rif y en el oeste de Argelia puede ser un reflejo de la diversidad genética que existió en el SW de Marruecos antes del declive de la especie en la zona debido probablemente a una combinación de factores climáticos (desertificación del Norte de África) y antropogénicos (pastoreo intensivo). El hecho de que los haplotipos que conforman este linaje no aparezcan en la Península Ibérica, puente de tierra entre las poblaciones africanas y francesas, es consistente con un escenario de dispersión a larga distancia, por el cual este linaje habría alcanzado el continente europeo y las islas del oeste del Mediterráneo. El factor más influyente para que una colonización tenga éxito después de un evento de dispersión a larga distancia es la idoneidad del hábitat de destino (Piñeiro

et al., 2007) y, según las proyecciones del modelo de distribución de especies, el sur de Francia y las islas del oeste del Mediterráneo aparecen como zonas climáticas idóneas desde el último Máximo Glacial, a excepción de Córcega que, curiosamente, alberga una población de *L. maritima* sobre un enclave muy especial y restringido de material volcánico que no aparece como idóneo en ninguna proyección de escenarios pasados. Este patrón filogeográfico que comprende poblaciones del Norte de África emparentadas con islas del oeste del Mediterráneo es compartido por otros grupos de plantas (García-Castaño *et al.*, 2014; González-Martínez *et al.*, 2010; Migliore *et al.*, 2012).

En cuanto a su especie hermana, *L. acerifolia*, los datos genéticos, específicamente SNPs obtenidos tras el filtrado de los datos GBS, revelan una estructura genética significativa. La combinación de marcadores moleculares y modelo de distribución de especies empleados en el Capítulo 2 sugiere que la especie siguió un claro patrón de colonización este-oeste. Varias fuentes de evidencia apoyan esta hipótesis: 1) la mayor diversidad genética y heterocigosidad encontrada en las poblaciones más cercanas al continente (Lanzarote y Fuerteventura), 2) la temprana separación de las poblaciones de estas dos islas (las más antiguas y más orientales) en el árbol filogenético, 3) el grupo genético bien diferenciado que forman las poblaciones de estas dos mismas islas, y 4) el origen africano de su especie hermana *L. maritima* (Capítulo 1). El estudio de modelo de distribución de especies aporta evidencia independiente que es consistente con la reconstrucción filogeográfica. El mantenimiento de las mismas zonas de idoneidad en escenarios pasados hace verosímil la persistencia de la distribución de *L. acerifolia* a lo largo del tiempo y minimiza la posibilidad de extinción en alguna isla y la recolonización desde islas cercanas. La distribución de idoneidad de nicho a lo largo del archipiélago y entre los diferentes sustratos geológicos pone de manifiesto que las islas orientales presentan menos zonas con idoneidad. Este resultado puede deberse a la importancia de las variables topográficas encontradas en el modelo de distribución de especies, y sugiere, además, que no sólo las erupciones volcánicas recientes influyen en la presencia de la especie sino la erosión gradual del modelado volcánico tendente a la suavización de la topografía. La idoneidad topoclimática tiende a aumentar hacia las islas jóvenes occidentales. Sin embargo, a pesar de las extensas áreas disponibles que podrían ser

colonizadas, estas islas solo albergan una población de *L. acerifolia*, y en el caso de El Hierro, ninguna. Este patrón puede ser reflejo de que el frente de colonización este-oeste es lento, tal vez por carecer de un mecanismo de dispersión evidente, pero está activo en el momento presente. Es decir, que aún no ha transcurrido tiempo suficiente para que la especie ultime una colonización más completa de esas dos islas occidentales. No puede descartarse la hipótesis alternativa de que el vulcanismo reciente en el suroeste de La Palma o el gran deslizamiento del Golfo en El Hierro hubieran podido extinguir poblaciones previamente establecidas. La estructura genética, el patrón de colonización inferido con los análisis moleculares y las evidencias aportadas por el modelo de distribución de especies confirman que la historia evolutiva de *L. acerifolia* encaja en la teoría clásica de biogeografía de islas, en la cual la colonización depende del tamaño (o áreas idóneas) y la edad de las islas (MacArthur & Wilson, 1963, 1967). Sin embargo, el patrón que detectamos en la isla de Tenerife –con la presencia de tres grupos genéticos, dos de los cuales se corresponden con la hipótesis de las paleo-islas que fueron el embrión de la isla– ilustra una peculiaridad de este archipiélago, que puede haberse sumado al efecto del mayor tamaño de esta isla.

2. Proceso de divergencia y especiación entre *Lavatera maritima* y *L. acerifolia*

¿Cómo de específico es el nicho ambiental de estas especies? ¿Están relacionadas las condiciones ambientales de ambas especies? Las dos especies muestran preferencia por el clima Mediterráneo, sin embargo *L. maritima* ocupa regiones con una mayor estacionalidad (e incluso continentalidad) en las temperaturas que *L. acerifolia,* la cual aparece en áreas donde la temperatura se mantiene constante a lo largo del año (Capítulo 3).

La versatilidad que presenta *L. maritima* para adaptarse a condiciones ambientales diferentes (Capítulo 1 y 3), su capacidad de dispersión a larga distancia y de establecimiento en diferentes áreas, la idoneidad climática encontrada en el SW de Marruecos en el tiempo presente y en proyecciones del modelo de distribución de especies al pasado (último Máximo Glacial y último Interglacial) (Capítulo 1), junto con

el patrón de colonización este-oeste de las Islas Canarias en *L. acerifolia* (Capítulo 2), y su nicho ambiental, mucho más reducido que el de *L. maritima* (Capítulo 3), son evidencias que apoyan el origen de *L. acerifolia* como el resultado de la colonización de las islas orientales por parte de *L. maritima*, o su ancestro. Esta colonización debió tener lugar a partir de poblaciones ancestrales de *L. maritima* situadas al oeste de África que consiguió alcanzar las islas y se diferenció de ésta gracias al aislamiento geográfico.

En resumen, es probable que *L. acerifolia* haya especiado en las Islas Canarias siguiendo un modelo peripátrico, debido a un evento de colonización y adaptación exitosos como consecuencia del aislamiento de los representantes ancestrales de *L. maritima*.

3. Aportaciones metodológicas exploradas durante la tesis doctoral

Durante el proceso de investigación, además de las técnicas expuestas en cada uno de los capítulos, se afrontaron una serie de desafíos metodológicos que han conducido a los resultados, pero que en el momento de presentación de esta memoria doctoral no han podido ser desarrollados con suficiente detalle e integrados en los tres capítulos que están ya elaborados. Se trata de varios proyectos aplicados a diferentes niveles de estudio, inter e intraespecífico, que requieren un tratamiento y análisis adicional de los datos antes de ser integrados en una publicación. En consecuencia, se da cuenta de ellos en este apartado, junto a la mención de filtrado bioinformático del GBS que sí se concluyó, así como en los apéndices de la memoria, pero serán completados en un futuro próximo.

Para conocer la historia evolutiva de un linaje es necesario indagar, no sólo en el ADN de los orgánulos (mitocondrial y plastidial), sino también en el ADN nuclear (Baldwin *et al.*, 1995; Feliner & Rosselló, 2007). Por este motivo, se llevó a cabo la amplificación y secuenciación de los espaciadores internos transcritos del ADN ribosómico nuclear, ITS (Apéndice 1). Se obtuvieron 225 secuencias ITS de 43 poblaciones de *L. maritima* y 27 secuencias ITS de 13 poblaciones de *L. acerifolia*. Posteriormente, los productos de PCR de ocho individuos de *L. maritima* y otros ocho de *L. acerifolia* fueron clonados, seleccionando diez colonias de cada producto de clonación. Por lo tanto, además de las

anteriores secuencias directas, se obtuvieron diez secuencias clonadas de cada fragmento de ITS amplificado en cada individuo. Sin embargo, debido a que estamos tratando con organismos hexaploides, por tanto con 6 juegos cromosómicos, la cantidad de posiciones polimórficas (Apéndice 1: Tabla 1) encontradas en cada una de las secuencias nos impidió obtener unos resultados concluyentes que pudiesen ser integrados en los estudios filogeográficos de los Capítulos 1 y 2 (Apéndice 1: Fig. 1, 2).

Además, se realizó un estudio de la variación del tamaño genómico de diferentes poblaciones de cada especie (Resultados en Apéndice 2). De forma general, se muestra que existen diferencias significativas entre el tamaño genómico de *L. maritima* y *L. acerifolia*, siendo ligeramente mayor el valor 2C en la especie *L. maritima*. Sin embargo, a nivel intraespecífico, las diferencias encontradas no fueron significativas.

La acumulación de sitios polimórficos intraindividulaes en las secuencias directas de los ITS se debe a la presencia de más de un fragmento de estas regiones dentro de los genomas de cada individuo, esto es, a que no ha sido completa la evolución concertada y por tanto los ITS no se han homogeneizado. Una de las causas de retardo en la homogenización de secuencias distintas de ITS conviviendo en un genoma es la presencia de diferentes loci ribosómicos que si están ubicados en distintas regiones del genoma son más difíciles de homogenizar por sobrecruzamientos desiguales. Para comprobar esta hipótesis, se realizó un estudio citogenético piloto mediante la técnica FISH con la que se puede identificar el número de loci de la región 45S del ADN ribosómico que presentan los genomas. Se pretendía comparar el número de loci entre las especies y entre las poblaciones que presentaron diferencias en el número de sitios polimórficos en el ITS así como entre aquellos que presentaron tamaño genómico distinto ya que el ADN ribosómico nuclear es parte del ADN repetitivo y diferencias en el número de estos genes multicopia pueden también motivar distintos tamaños genómicos (Resultados en Apéndice 3). Este estudio piloto se realizó en colaboración con Marcela Rosato (Jardín Botánico de la Universidad de Valencia).

4. Análisis bioinformáticos en genomas poliploides

Esta memoria doctoral ha abordado la vertiente bioinformática del filtrado, edición, análisis e interpretación de datos genómicos que, a diferencia de los aspectos metodológicos mencionados más arriba (secuencias ITS, tamaños genómicos, análisis FISH), sí han sido concluidos e integrados en los resultados (Capítulo 2 y 3). Mediante el desarrollo de varios flujos de trabajo, se ha extraído un conjunto de SNPs de genomas poliploides de organismos no modelo, como son las dos especies estudiadas, y sin genoma de referencia, empleando técnicas de secuenciación de nueva generación (Next Generation Sequencing, NGS) (Capítulo 2). Estos polimorfismos escaneados a lo largo del genoma del individuo ofrecen grandes ventajas para los estudios genómicos (Narum *et al.*, 2013). Sin embargo, el proceso de filtrado de datos, que es fundamental para maximizar la señal que podemos extraer de ellos –y minimizar el ruido—, es un arduo trabajo muy poco estudiado cuando se trata de genomas hexaploides. Tres flujos de trabajo diferentes fueron llevados a cabo: con genoma de referencia, creando una secuencia de referencia con datos propios, y ensamblando *de novo* (Capítulo 2). La conclusión en este apartado metodológico es que destacamos la necesidad de usar un genoma de referencia cercano a la especie de estudio. Sin embargo, esta carencia puede paliarse con la creación de una secuencia de referencia a partir de datos propios (Shafer *et al.*, 2016). Cuando comparamos las dos aproximaciones implementadas en el Capítulo 2 (*de novo* y con secuencia de referencia) comprobamos que la creación de una secuencia de referencia propia aporta mayor fiabilidad a los SNPs extraídos desde genomas poliploides sin genoma de referencia, que el realizar un ensamblado *de novo* (Capítulo 2).

REFERENCIAS

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*: 247-277.

Feliner GN, Rosselló JA. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44: 911-919.

Fernández-Mazuecos M, Vargas P. 2013. Congruence between distribution modelling and phylogeographical analyses reveals Quaternary survival of a toadflax species (*Linaria elegans*) in oceanic climate areas of a mountain ring range. *New Phytologist* 198: 1274-1289.

García-Castaño JL, Terrab A, Ortiz MÁ, Stuessy TF, Talavera S. 2014. Patterns of phylogeography and vicariance of *Chamaerops humilis* L. (Palmae). *Turkish Journal of Botany* 38: 1132-1146.

González-Martínez SC, Dubreuil M, Riba M, Vendramin G, Sebastiani F, Mayol M. 2010. Spatial genetic structure of *Taxus baccata* L. in the western Mediterranean Basin: past and present limits to gene movement over a broad geographic scale. *Molecular Phylogenetics and Evolution* 55: 805-815.

MacArthur RH, Wilson EO. 1963. An equilibrium theory of insular zoogeography. *Evolution*: 373-387.

MacArthur RH, Wilson EO. 1967. The theory of island biogeography. *Princeton, NJ*.

Migliore J, Baumel A, Juin M, Médail F. 2012. From Mediterranean shores to central Saharan mountains: key phylogeographical insights from the genus *Myrtus*. *Journal of Biogeography* 39: 942-956.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* 22: 2841-2847.

Piñeiro R, Aguilar JF, Munt DD, Feliner GN. 2007. Ecology matters: Atlantic–Mediterranean disjunction in the sand-dune shrub *Armeria pungens* (Plumbaginaceae). *Molecular Ecology* 16: 2155-2171.

Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW. 2016. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*: n/a-n/a.

Viruel J, Médail F, Juin M, Haguenauer A, Nieto Feliner G, Dagher-Kharrat MB, La Malfa S, Ouahmane L, Sanguin H, Baumel A. 2016. Mediterranean carob populations, native or naturalized? A continuing riddle. *International Conference of Ecological Sciences*. Marseille: doi: 10.13140/RG.2.2.33681.84328.

CONCLUSIONES

1. *L. maritima* y *L. acerifolia* son especies hermanas cuya edad de divergencia estimamos en c. 2,7 Ma (Plio-Pleistoceno).
2. El análisis de la distribución de haplotipos plastidiales en *L. maritima* indica que no muestra estructura filogeográfica clara.
3. En las poblaciones del Norte de África de *L. maritima* se concentra la mayor diversidad haplotípica y nucleotídica.
4. Según el estudio de modelización de distribución de especies, el SW de Marruecos aparece como un área potencialmente idónea para la presencia de *L. maritima* desde el último Interglaciar hasta la actualidad. Ello unido a la diversidad genética encontrada en la región para esta especie, apoya la idea del SW de Marruecos como reservorio genético así como la posibilidad de que sea también el origen de la especie.
5. La ausencia de idoneidad en las costas orientales españolas para *L. maritima* durante el último Interglaciar se explica por las condiciones de continentalidad existentes durante este período.
6. La historia evolutiva de *L. acerifolia* se ajusta bien al modelo de migración paso a paso recogido la Teoría clásica de Biogeografia Islas.
7. Las poblaciones de *L. acerifolia* presentan una estructura filogeográfica significativa basada en los datos del genoma.
8. *L. acerifolia* ha seguido una ruta de colonización este-oeste como muestran la distribución de la diversidad genética entre sus poblaciones y los valores de heterocigosidad, que disminuyen desde las poblaciones de las islas más antiguas (cercanas al continente) hacia las poblaciones de las islas jóvenes occidentales (más alejadas del continente).
9. La modelización de distribución de especies para *L. acerifolia* apoya un frente lento de colonización este-oeste en el que las islas occidentales, que presentan mayor idoneidad climática que las islas orientales tanto en la actualidad como en las proyecciones al pasado, apenas muestran poblaciones de esta especie.
10. Los nichos ambientales de *L. maritima* y *L. acerifolia* muestran una diferenciación, siendo el de *L. acerifolia* notablemente más estrecho. La temperatura estacional y

la precipitación anual son las variables más influyentes en el nicho ambiental de ambas especies.

11. Hipotetizamos que *L. acerifolia* divergió de *L. maritima* o un antepasado común y se estableció en las Islas Canarias siguiendo un modelo de especiación peripátrica, por tanto inicialmente promovido por aislamiento geográfico gracias a la capacidad de dispersión a larga distancia mostrada por *L. marítima.* La diferenciación de nicho habría ocurrido posteriormente mediante adaptación al nuevo hábitat.
12. El empleo de un flujo de trabajo bioinformático usando una secuencia de referencia creada con datos propios aporta mayor fiabilidad a los datos de SNPs generados con GBS cuando se estudia un organismo poliploide, no modelo, sin genoma de referencia próximo.

Apéndice 1

Tabla1. Tabla de polimorfismos obtenidos a partir de las secuencias ITS

POBLACIÓN	INDIVIDUO	97	123	149	237	242	256	257	402	455	473	510	517	535	552	560	621	632
TAZAGUINE	TAZ.2	C	A	C	G	G	C	C	G	A	T	G	C	A	T	C	G	G
	TAZ.4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	TAZ.5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	TAZ.1	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	TAZ.3	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ALQUEZAR	HU.ALQ.7	.	G	.	.	G	.	.	.	.	Y	.	.	.	.	.	K	.
	HU.ALQ.1	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	HU.ALQ.8	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	HU.ALQ.9	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	HU.ALQ.13	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
EMBID	ZGZ.EMB.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ZGZ.EMB.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ZGZ.EMB.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ZGZ.EMB.6	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ZGZ.EMB.10	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
MALLOS	HU.MAL.2	.	.	.	.	.	.	Y	.	R	Y	.	.	.	Y	.	.	.
	HU.MAL.14	.	.	.	.	.	.	.	.	R	Y	K	.	.	Y	.	.	.
	HU.MAL.15	.	.	.	.	.	.	.	.	R	Y	K	.	.	Y	.	.	.
	HU.MAL.16	.	.	Y	.	.	.	Y	.	G	Y	K	.	.	Y	.	.	.
ESTOPIÑAN B	HU.EST.B.1	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	HU.EST.B.2	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	HU.EST.B.3	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	HU.EST.B.4	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	HU.EST.B.5	.	.	.	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
CASTILLONROY	HU.CAS.1	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	HU.CAS.8	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	HU.CAS.14	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	HU.CAS.15	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	HU.CAS.20	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
OLVENA	HU.OLV.4	Y	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	HU.OLV.7	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	HU.OLV.9	Y	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	HU.OLV.11	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	HU.OLV.13	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
CREUS NORFEU	GE.FOR.1	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	GE.FOR.2	.	R	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
	GE.FOR.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	GE.FOR.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	GE.FOR.6	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
MENORCA	MEN.1	.	R	.	.	.	.	.	.	R	Y	.	.	.	.	.	.	.
	MEN.3	.	R	.	.	.	.	.	.	R	C	.	.	.	.	.	.	.
	MEN.7	.	R	.	.	.	.	.	.	R	Y	.	.	.	.	.	.	.
	MEN.8	.	.	.	.	.	.	Y	.	R	Y	K	.	.	.	.	.	.
	MEN.9	.	R	.	.	.	.	.	.	R	Y	.	.	.	.	.	.	.
IMOUZZER 1	IMO1.2	.	.	.	.	R	.	.	.	.	Y	.	.	.	.	.	.	.
	IMO1.8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	IMO1.9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	IMO1.10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	IMO1.11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
IMOUZZER 2	IMO2.1	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	IMO2.2	.	.	.	.	R	.	.	.	.	Y	.	.	.	.	.	.	.
	IMO2.5	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	IMO2.6	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	IMO2.7	.	.	.	.	R	.	.	.	.	Y	.	.	.	.	.	.	.
ESPUÑA	MU.ESP.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MU.ESP.2	.	G	.	.	.	.	.	S	.	Y	.	.	.	.	.	K	.
	MU.ESP.3	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MU.ESP.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MU.ESP.6	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
CERDEÑA	CER.2	.	R	Y	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	CER.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	CER.5	.	R	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	CER.8	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
RACO DEL FRARE	CAS.RAC.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	CAS.RAC.3	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	CAS.RAC.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	CAS.RAC.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	CAS.RAC.6	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	K	.
BORRIOL	CAS.BOR.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	CAS.BOR.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	CAS.BOR.3	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	CAS.BOR.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	CAS.BOR.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
LLIBER	ALI.LIB.3	.	R	.	.	.	.	.	.	R	Y	.	.	.	.	.	.	.
	ALI.LIB.5	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALI.LIB.6	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALI.LIB.10	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALI.LIB.12	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
MARXUQUERA	VA.MAX.1	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	VA.MAX.2	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	VA.MAX.3	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	VA.MAX.4	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	VA.MAX.5	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.

Tabla1. (continuación)

POBLACIÓN	INDIVIDUO	97	123	149	237	242	256	257	402	455	473	510	517	535	552	560	621	632
MONTGÓ	ALI.MON.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALI.MON.2	.	R	Y	K	.	.	.	.	R	Y	.	Y	.	Y	.	.	.
	ALI.MON.3	.	R	Y	K	.	.	.	.	R	C	.	Y	.	Y	Y	.	.
TIÑOSO	MU.TIÑ.3	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MU.TIÑ.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
COPE	MU.COP.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MU.COP.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MU.COP.3	.	G	.	.	.	.	.	S	.	Y	.	.	.	.	.	.	.
	MU.COP.4	.	G	.	.	.	.	.	S	.	Y	.	.	.	.	.	K	.
	MU.COP.5	.	G	.	.	.	.	.	S	.	Y	.	.	.	.	.	.	.
LOS LOBOS	ALM.LOB.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALM.LOB.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	ALM.LOB.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALM.LOB.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
SAN PEDRO	ALM.PED.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	ALM.PED.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	ALM.PED.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
GATA	ALM.GATA.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALM.GATA.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
OTIÑAR	J.OTI.1	.	R	.	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	J.OTI.2	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	J.OTI.3	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	J.OTI.4	.	R	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.
	J.OTI.5	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
AGUADULCE	ALM.AGU.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALM.AGU.2	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	ALM.AGU.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
FRIGILIANA	MLG.FRI.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MLG.FRI.2	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MLG.FRI.3	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MLG.FRI.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MLG.FRI.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
SUAREZ	MLG.SUA.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MLG.SUA.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MLG.SUA.5	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
LLANOS LIBAR	MLG.HAC.1	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MLG.HAC.2	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	MLG.HAC.3	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MLG.HAC.4	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MLG.HAC.5	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
CALAHONDA	GR.CAL.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	GR.CAL.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	GR.CAL.3	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	GR.CAL.4	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	GR.CAL.5	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
LOBRES	GR.FEO.1	Y	G	.	.	.	M	.	.	.	Y	.	.	.	.	.	.	.
	GR.FEO.2	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	GR.FEO.3	Y	G	.	.	.	M	.	.	.	Y	.	.	.	.	.	.	.
	GR.FEO.4	Y	G	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
	GR.FEO.5	Y	G	.	.	.	M	.	.	.	Y	.	.	.	.	.	.	.
PIZARRA	MLG.PIZ.1	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MLG.PIZ.3	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MLG.PIZ.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
LA ARAÑA	MLG.ARA.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	MLG.ARA.2	.	G	.	.	.	.	.	.	.	Y	.	.	M	.	.	.	R
	MLG.ARA.3	.	G	.	.	.	.	.	.	.	Y	.	.	M	.	.	.	R
	MLG.ARA.4		G	.	.	.	.	.	.	.	Y	.	.	M	.	.	.	R
	MLG.ARA.5		G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
MOLADERAS	ALM.AMO.1	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALM.AMO.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	K	.
	ALM.AMO.3	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALM.AMO.4	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALM.AMO.5	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
ZAIO	ZAI.1	.	R	Y	.	.	.	.	.	R	.	.	.	.	Y	.	.	.
	ZAI.2	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	ZAI.3	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	ZAI.4	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	ZAI.5	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
SAIDIA	SAD.1	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	SAD.2	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	SAD.3	.	R	Y	.	.	.	.	.	.	Y	.	.	.	Y	.	.	.
	SAD.4	.	R	.	.	.	.	.	.	.	Y	.	.	.	Y	.	.	.
	SAD.5	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
ZEGZEL	ZEG.2	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ZEG.3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	ZEG.4	.	R	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.
	ZEG.5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PLAINE	PLA.1	.	.	Y	.	.	.	Y	.	G	C	.	.	.	Y	.	.	.
	PLA.2	.	.	Y	.	.	.	.	.	G	C	.	.	.	C	.	.	.
	PLA.3	.	.	Y	.	.	.	.	.	G	C	K	.	.	Y	.	.	.
	PLA.4		.	Y	.	.	.	.	.	G	C	.	.	.	C	.	.	.
	PLA.5	.	.	Y	.	.	.	.	.	G	C	K	.	.	Y	.	.	.
GAREB	GAR.1	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	GAR.2	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	GAR.3	.	.	Y	.	.	.	.	.	G	C	K	.	.	Y	.	.	.
	GAR.4	.	.	Y	.	.	.	Y	.	G	C	K	.	.	Y	.	.	.
	GAR.5	.	.	Y	.	.	.	Y	.	G	C	K	.	.	Y	.	.	.

Tabla1. (continuación)

POBLACIÓN	INDIVIDUO	97	123	149	237	242	256	257	402	455	473	510	517	535	552	560	621	632
BADES	BAD.1	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	BAD.2	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	BAD.3	.	R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	BAD.4		R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	BAD.5		R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
MECHRA-HOMANI	MEC.1	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	MEC.2	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	MEC.3	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	MEC.5	.	.	T	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
AL-HOCEIMA	ALH.1	.	G	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
	ALH.2	.	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALH.3		G	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
	ALH.4		G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ALH.5		G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
JBEL GUILLIZ	JBG.1	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	JBG.2	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	JBG.3	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	JBG.4		R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	JBG.5		.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
BENI SNASSEN	BSN.1	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	BSN.2	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	BSN.3	.	.	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	BSN.4		.	Y	.	.	.	.	.	.	Y	.	.	.	Y	.	.	.
	BSN.12		R	Y	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
ABDADGADEL	ABD.1	.	R	Y	.	.	.	.	.	R	Y	.	.	.	.	.	.	.
	ABD.2	.	R	Y	.	.	.	.	.	R	Y	.	.	.	Y	.	.	.
	ABD.3	.	R	Y	.	.	.	.	.	R	Y	.	.	.	.	.	.	.
	ABD.4		R	Y	.	.	.	.	.	R	T	.	.	.	.	.	.	.
	ABD.5		R	Y	.	.	.	.	.	R	Y	.	.	.	.	.	.	.
EVENOS	EVE.1		G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	EVE.4		G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
VINGRAU	VIN.1		.	.	.	.	.	Y	.	R	C	.	.	.	Y	.	.	.
	VIN.2		.	.	.	.	.	Y	.	R	C	.	.	.	Y	.	.	.
	VIN.3		R	.	.	.	.	.	.	R	C	.	.	.	Y	.	.	.
	VIN.4		.	Y	.	.	.	Y	.	G	C	.	.	.	Y	.	.	.
	VIN.5		R	.	.	.	.	Y	.	R	C	.	.	.	Y	.	.	.
ILE DE RATONNEAU	FRI.1		R	Y	.	.	.	.	.	R	C	.	.	.	.	.	.	.
	FRI.2	Y	R	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
	FRI.3	Y	R	.	.	.	.	.	.	R	C	.	.	.	.	.	.	.
	FRI.4		R	Y	.	.	.	.	.	R	C	.	.	.	.	.	.	.
	FRI.5		R	Y	.	.	.	.	.	R	C	.	.	.	.	.	.	.
COL DE L'ARMA	ROY.1		R	Y	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ROY.2		R	Y	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ROY.4		R	Y	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	ROY.5		R	Y	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
GRAND VALLON	LAS.1		R	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	LAS.2	Y	G	Y	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	LAS.3		G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	LAS.4	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
	LAS.5	Y	G	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
ROCHE EMÉRE	VIL.1		G	.	.	.	.	.	.	R	C	.	.	.	.	.	.	.
	VIL.2		G	.	.	.	.	.	.	R	C	.	.	.	.	.	.	.
	VIL.3		G	.	.	.	.	.	.	R	C	.	.	.	.	.	.	.
	VIL.4		G	.	.	.	.	.	.	R	C	.	.	.	.	.	.	.
	VIL.5		G	.	.	.	.	.	.	R	C	.	.	.	.	.	.	.

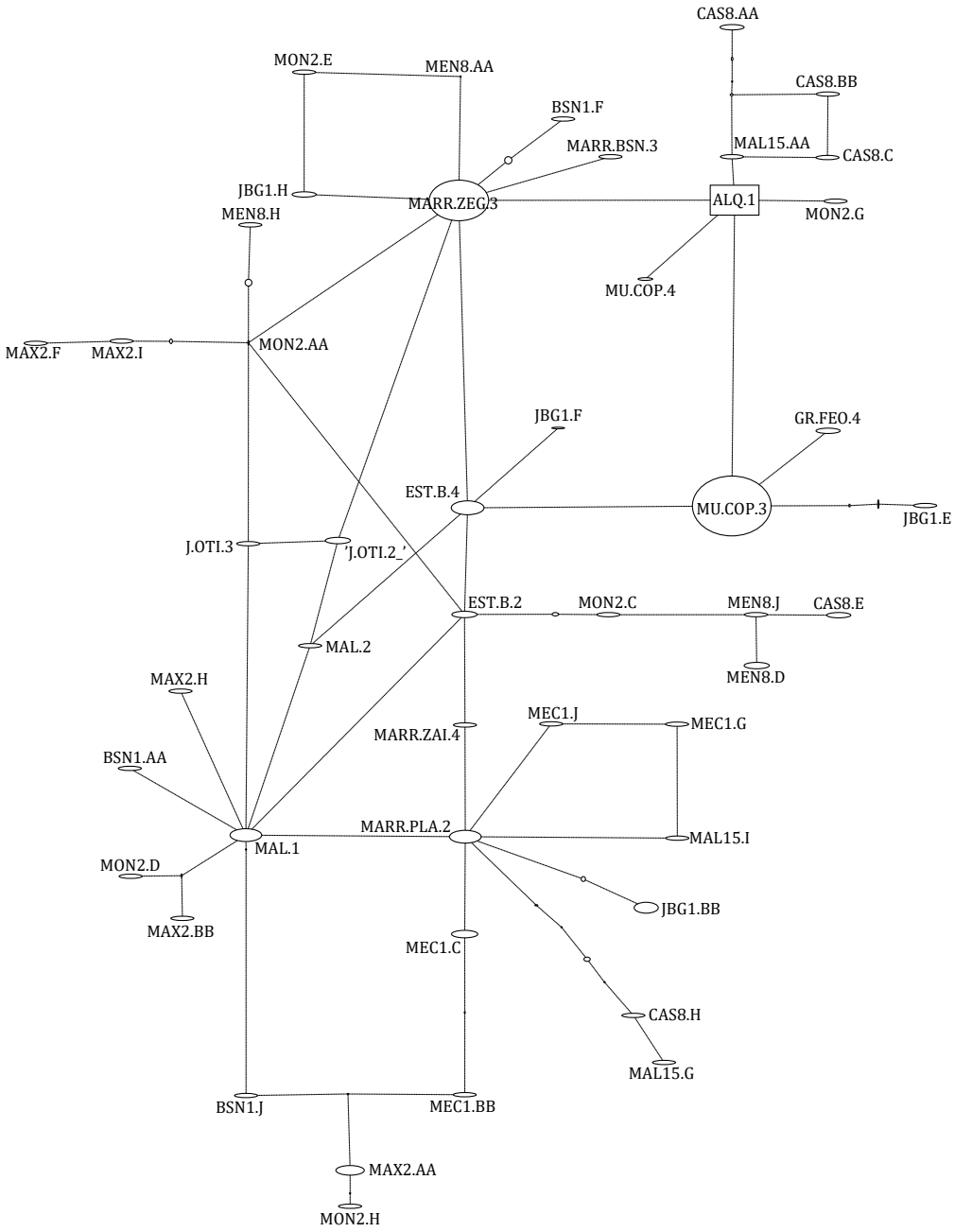


Figura1. Red de ribotipos obtenida con TCS v1.21 (Clement, Posada & Crandall, 2000) a partir de secuencias directas (122) y clonadas (80) de *L. maritima*.

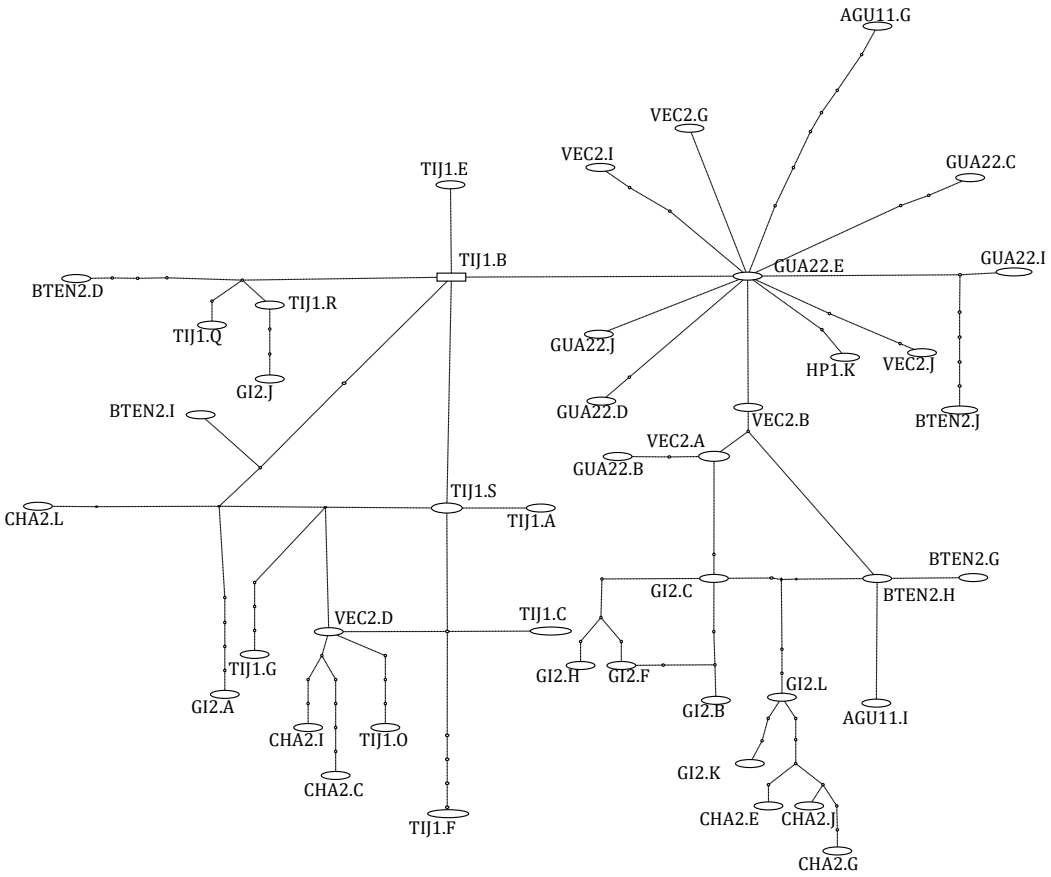


Figura2. Red de ribotipos obtenida con TCS v1.21 (Clement, Posada & Crandall, 2000) a partir de secuencias clonadas (46) de *L. acerifolia*.

Apéndice 2

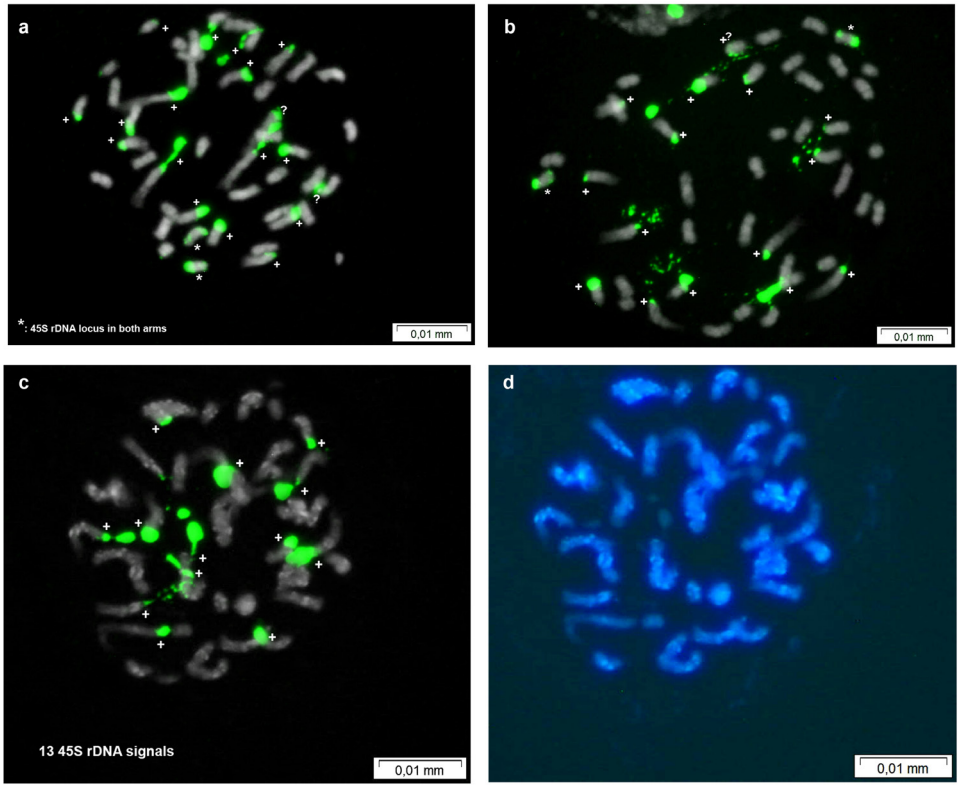
Tabla. Valores de tamaño genómico obtenidos de diez individuos de L*avatera maritima,* seis individuos de *L. acerifolia* y uno de *Navaea phoenicea*. Se realizaron 3 réplicas técnicas para cada medida individuo. Patrón: *Petunia hybrida* (2C=2.85 pg)

Especie	Individuo	Población	Provincia	2C (pg ± SD)
Lavatera maritima				
	ABD.6	Abdadgadel	Zaio	3,8 ± 0,06
	PLA	Plane du Gareb	Tiztoutine	3.76
	TAZ	Tazaguine	Berkane	3,75 ± 0,01
	ZEG.1	Gorges du Zegzel	Berkane	3,75 ± 0,01
	CAS,STANA	Castillonroy	Huesca	3,70 ± 0,06
	EST.B.10	Estopiñan	Huesca	3,74 ± 0,01
	SUA	Granja Suárez	Málaga	3,78 ± 0,06
	PED	San Pedro	Almeria	3.91
	GATA	Cabo de Gata	Almeria	3.74
	CAL	Calahonda	Granada	3.83
		PED-CAL-GATA		3,83 ± 0,09
Lavatera acerifolia				
	TENO1	Bco. Teno	Tenerife	3,43 ± 0,06
	GÜI.1	Güimar	Tenerife	3.47
	GUA2	Guayadeque	Gran Canarias	3,47 ± 0,05
	HP	Hoyo Pineda	Gran Canarias	3.5
	VEC	Agaete	Gran Canarias	3.53
	FUE	Morro del Halconcillo	Fuerteventura	3,37 ± 0,12
Navaea phoenicea			Tenerife	3,92 ± 0,02

Apéndice 3

Tabla. Resultado de la técnica citogénetica de marcaje de cromosomas FISH. Se encuentran diferentes números de loci de la región 45S del ADN ribosómico a nivel inter e intraespecífico, para el caso de *L. maritima*.

Specie	Locality	Code	2n	Ploidy level			Total No. 45S rDNA loci	DAPI bands
Lavatera maritima	Marroc, Gorges du Zegzel	ZEG	44	6x	8	1	10	-
	Spain, Málaga	SUA	44	6x	7	1	9	-
Lavatera acerifolia	Spain, Gran Canaria, Guayadeque	GUA1	44	6x	7	-	7	+
	Spain, Tenerife, Teno	TENO	44	6x	7	-	7	+



Physical mapping of 45S rDNA assesed by FISH. a. *L. maritima*, Marroc (ZEG); b: *L. maritima*, Spain (SUA); c,d: *L. acerifolia*, Spain (GUA1). d: DAPI+ bands of the same cell c.

Apéndice 4

Other article published during the PhD work

Heredity (2017) 118, 513–516

www.nature.com/hdy

NEWS AND COMMENTARY

Empiricist's view of homoploid hybrid speciation

Is homoploid hybrid speciation that rare? An empiricist's view

G Nieto Feliner, I Álvarez, J Fuertes-Aguilar, M Heuertz, I Marques, F Moharrek, R Piñeiro, R Riina, JA Rosselló, PS Soltis and I Villa-Machío

Heredity (2017) **118,** 513–516; doi:10.1038/hdy.2017.7; published online 15 March 2017

Natural hybridization and its role in evolution and specifically in generating new diversity is an old and yet endlessly revitalized topic (Lotsy, 1916; Anderson, 1949; Stebbins, 1959; Rieseberg *et al.*, 2003; Mallet, 2007; Soltis and Soltis, 2009; Larsen *et al.*, 2010; Arnold *et al.*, 2012b; Pereira *et al.*, 2014; Grant and Grant, 2015; Abbott *et al.*, 2016; Pennisi, 2016). Homoploid hybrid speciation (HHS) is the formation of a new-hybrid—species, independent from its parents, via hybridization with no whole-genome duplication and thus no increase in ploidy. Beyond this basic definition, complete agreement is lacking on key aspects of the process, such as the relative proportions of each parental genome present in a hybrid species, the mechanisms leading to reproductive isolation (RI), the degree of RI or the role played by hybridization in the process (Rieseberg, 1997; Abbott *et al.*, 2010). While our understanding of HHS has been improved by detailed evolutionary case studies documented by recent reviews (Abbott *et al.*, 2013; Yakimowski and Rieseberg, 2014; Payseur and Rieseberg, 2016) and empirical studies focused on mechanisms leading to HHS (Renaut *et al.*, 2014; Selz *et al.*, 2014; Lukhtanov *et al.*, 2015), there is controversy concerning the criteria to identify and demonstrate HHS, and even the range of situations that HHS might encompass.

For more than two decades, phylogenetic studies have reported the discovery of putative hybrid lineages at a continuous pace (for example, Rieseberg and Soltis, 1991; Rieseberg *et al.*, 1996; Soltis and Soltis, 2009; Blanco-Pastor *et al.*, 2012; Sousa *et al.*, 2016). In parallel, and partly to sort out the wealth of reported cases, attempts have been made to distinguish among those case studies that convincingly demonstrate HHS from those that correspond to other evolutionary contributions of hybridization or gene flow, for example, adaptive introgression (Rieseberg, 1997; Gross and Rieseberg, 2005). But HHS and such other evolutionary contributions of hybridization lie along a continuum, and in fact adaptive introgression may be involved in HHS (Seehausen, 2004,2013; Abbott *et al.*, 2013). Therefore, focusing our discussion just on HHS is a simplification if one is interested in understanding the role of hybridization (without polyploidy) in differentiation and speciation. However, the HHS concept is widely used, and we think that pointing out potential weaknesses in criteria that are too stringent is useful to avoid misconceptions and contribute to a solid and, at the same time, open conceptual framework (Wiens, 2004) for such a complex topic.

Yakimowski and Rieseberg (2014) list 19 putative cases of HHS among seed plants, two of them in genera in which more than one species is of hybrid origin. Previously, Gross and Rieseberg (2005) considered nine additional cases, including four invertebrates and one fish, and Abbott *et al.* (2013) recognized additional examples among fishes (Stemshorn *et al.*, 2011), sparrows (Elgvin *et al.*, 2011), and butterflies (Kunte *et al.*, 2011). Altogether, there are probably more than 30 cases that have received molecular support as homoploid hybrid species. In contrast, in a recently published paper, Schumer *et al.* (2014) suggested that a putative hybrid species should satisfy three criteria for confident consideration as such. These criteria are: (1) a strong RI mechanism between the putative parental and hybrid species; (2) genetic evidence of hybridization; and (3) isolating mechanisms derived from hybridization itself. They concluded that only four examples across the living world fulfil these three requisites and are thus considered as true homoploid hybrid species: the butterfly *Heliconius heurippa* (Salazar *et al.*, 2010) and the three hybrid sunflower species, *Helianthus anomalus, H. deserticola* and *H. paradoxus* (Rieseberg, 1991).

We think that the views in Schumer *et al.* (2014) illustrate a trend that narrows the concept of HHS, and we question, in this commentary, their concept by examining its pros and cons, for example, of concentrating the discussion of HHS primarily on RI, and discussing whether the importance and frequency of HHS can be assessed under such a position. We believe that the HHS concept remains operationally useful to account for the generation of stable novel diversity via hybridization without polyploidy, provided that it can fit a broader scope of scenarios than those depicted by the above-mentioned stringent criteria.

G Nieto Feliner, I Álvarez, J Fuertes-Aguilar, R Riina, I Villa-Machío are at Real Jardín Botánico, CSIC, Madrid, Spain; M Heuertz is at BioGeCo INRA, Université de Bordeaux, Cestas, France; I Marques is at Department of Agricultural and Environmental Sciences, High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain; and the UBC Botanical Garden & Centre for Plant Research, University of British Columbia, Vancouver, British Columbia, Canada; F Moharrek is at Department of Plant Biology, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran; R Piñeiro is at Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, UK; JA Rosselló is at Jardí Botànic, Universitat de Valencia, Valencia, Spain; PS Soltis is at Florida Museum of Natural History, University of Florida, Gainesville, FL, USA
Correspondence: Dr G Nieto Feliner, Department of Biodiversity and Conservation, Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain.
E-mail: nieto@rjb.csic.es
Received 13 May 2016; revised 19 January 2017; accepted 20 January 2017; published online 15 March 2017

HYBRIDIZATION-DERIVED ISOLATION

Schumer *et al.* (2014) consider that a case fulfils the requirements for being considered HHS if the hybridization event itself was the original trigger of RI. This criterion is based on the argument that the four most compelling cases of hybrid speciation combine genetic evidence of hybridization with evidence that hybridization led to the emergence of RI. We rather believe that the reason why those cases are convincing is that they have been more thoroughly studied in every aspect, not only the origin of RI but also the contribution of hybridization to ecological divergence (Rieseberg *et al.*, 2003) and how quickly hybrid genomes stabilize (Buerkle and Rieseberg, 2008), among other topics. Furthermore, we think that if there is evidence that a hybridization event has given rise to an established, persistent, morphologically and ecologically distinct hybrid lineage, the recognition of this fact should not be compromised by whether or not we can demonstrate that hybridization was directly the cause of RI. Hybridization can be causative of mechanisms that contribute to enhancing RI in hybrid lineages, for example, the sorting of chromosomal rearrangements along the recombinational speciation model (Lai *et al.*, 2005; Lukhtanov *et al.*, 2015) or the occurrence of new traits that change mating patterns (Vereecken *et al.*, 2010; Selz *et al.*, 2014; Marques *et al.*, 2016; Ma *et al.*, 2016). In addition, intrinsic changes in the hybrids not directly causing RI between hybrid lineages and their progenitors may ultimately lead to external RI by facilitating the colonization of new niches (Grant, 1981; Gross and Rieseberg, 2005). But the possibility that RI results from geographical and/or ecological barriers that are not traceable to the hybridization event cannot be excluded. *Senecio squalidus*, a hybrid species formed in Great Britain, acquired geographic isolation from its parents, both of which occur on Mount Etna, Sicily, when it was introduced into the UK (James and Abbott, 2005). We fail to see why this case in which hybridization itself is not the direct cause of RI should not be considered a homoploid hybrid species. Creating such eco-geographic barriers between hybrids and parental species need not rely on human-mediated dispersal. Mechanisms acting on small time scales and macrospatial scales, such as long distance dispersal (LDD) of hybrid lineages, may bring about a rapid isolation but gradual mechanisms probably more commonly lead to external RI. For instance, migration and recurrent bottlenecks seem to have isolated *Pinus densata* from its congeners (Wang *et al.*, 2011).

In addition, we argue that requiring hybridization to be the direct cause of RI may shift the focus of the research away from a crucial aspect of HHS: the production of raw genetic material for selection at higher rates than mutation (Grant and Grant, 1994; Arnold *et al.*, 2012a; Abbott *et al.*, 2013), which can be a source of evolutionary novelty (Soltis, 2013; also for allopolyploids, Soltis and Soltis, 2016). Furthermore, enforcing the hybridization-derived RI criterion might also imply uncritically assuming a role for RI in HHS that is pivotal under a specific model of speciation, which fits the biological species concept, but is not considered crucial under others (see below). From an epistemological point of view, establishing a stringent set of criteria for falsifying putative HHS hypotheses could be seen as an advantage, but this is at the cost of establishing an overly restrictive criterion.

The first criterion advocated by Schumer *et al.* (2014) for recognizing true cases of HHS, that of demonstrating strong RI, is not controversial in itself but altogether illustrates our insufficient understanding of the HHS process(es). RI is a *sine qua* non condition to initiate speciation (Coyne and Orr, 2004) and intrinsic reproductive isolating mechanisms, in particular, maintain integrity of species whenever they come into contact. There is much theoretical and empirical research on the components of RI (Lafon-Placette and Köhler, 2016; Pease *et al.*, 2016) and how to identify and measure them (Ramsey *et al.*, 2003; Martin and Willis, 2007; Sobel and Chen, 2014). However, there is also a growing concern about the actual role of intrinsic reproductive isolating mechanisms in the speciation process itself, particularly in allopatric speciation (Wiens, 2004) and specifically on whether they are drivers or merely by-products of divergent evolution (Sætre, 2013). Acknowledging this dilemma leads to rethinking whether RI should be considered the major factor for recognizing HHS. Furthermore, beyond the evidence that speciation can occur with considerable levels of gene flow (Mallet, 2005; Smadja and Butlin, 2011; Feder *et al.*, 2012) and that RI is frequently incomplete between well-established species (Grant and Grant, 2002), there is debate as to whether the (more or less episodic) interruptions of RI may stimulate speciation (Seehausen, 2004, 2013; Sætre, 2013; Lamichhaney *et al.*, 2015).

In sum, we think that an alternative view to the question of whether hybridization generates RI in HHS processes is to ask whether hybridization generates novel diversity which, by various means, becomes reproductively isolated and stabilized in a different niche, even if RI is not complete, as expected throughout most of the speciation process (Lowry and Gould, 2016).

AN EMPIRICIST'S APPROACH TO HHS

Are homoploid hybrid species as rare as the criteria of Schumer *et al.* (2014) imply? This question cannot be answered conclusively at this point, and we also ignore here the proportion of hybridization events that have led to speciation (Abbott *et al.*, 2013), but there are hints that HHS is not particularly rare, at least when putative cases of this process are considered with a less stringent view. In addition to the mentioned four paradigmatic cases recognized by Schumer *et al.* (2014), a number of examples of potential homoploid hybrid species have been confirmed: for example, the Oxford ragwort *Senecio squalidus* (James and Abbott, 2005; Brennan *et al.*, 2012), *Iris nelsonii* (Arnold, 1993; Taylor *et al.*, 2013), *Pinus densata* (Wang *et al.*, 2001; Gao *et al.*, 2012), *Penstemon clevelandii* (Wolfe *et al.*, 1998) and *Paeonia anomala* (Pan *et al.*, 2007). Significantly, there are many other potential examples of homoploid hybrid species detected in phylogenetic analyses, which have not been thoroughly studied but have been tested against incomplete lineage sorting and have some temporal trajectory and niche differentiation with respect to their progenitors. One can currently consider those cases as hybrid lineages, pending further study, but it is important to call attention to them because phylogenetic approaches offer powerful methods for discovering HHS and also provide complementary information for speciation studies, particularly when these follow stringent criteria such as those of Schumer *et al.* (2014).

As in any other scientific field, in speciation studies it is important that data are collected within a solid conceptual framework which, however, should remain open for debate (Wiens, 2004). This is especially so when disparate views exist on how theoretical studies about natural hybridization (Barton and Hewitt, 1985; Barton and Gale, 1993; Harrison, 1993) should affect empirical evidence (Butlin and Ritchie, 2013; Servedio *et al.*, 2013). Under this perspective, it would be more helpful to adopt broader conceptual frameworks for HHS than that of Schumer *et al.* (2014) such as those in Abbott *et al.* (2013) and Mallet (2007).

In addition, because all putative cases of HHS are detected and initially studied by empiricists, it would be impractical to rely on analysis of RI for recognition of homoploid

hybrid species and to apply restrictive criteria at this stage. We thus think that viewing empirical evidence more broadly will minimize false negatives and allow for other aspects that are as important as RI. In particular, a dimension that requires consideration equal to that of RI and the traits and genes responsible for it (barrier genes) is the ecological context of the HHS process ideally including the traits and genes related to the occupation of a new niche. Given that we cannot confidently expect general patterns in HHS and that the speciation process is a complex continuum (Lowry and Gould, 2016), we think it is preferable to encourage reporting rather than discouraging putative cases of HHS.

In summary, we agree that case studies should rigorously test the role of RI. However, we believe that the benefits that Schumer *et al.*'s restricted vision of HHS may have in terms of facilitating falsification of putative cases do not outweigh two questionable aspects: requiring that RI derives directly from hybridization, which we deem unnecessary, and focusing exclusively on RI, which may shift the interest away from other crucial elements in HHS, that is, the ecological dimensions of the process and the production of novel diversity.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Alex Twyford, and three anonymous reviewers for very helpful suggestions for the manuscript.

Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC (2010). Homoploid hybrid speciation in action. *Taxon* **59**: 1375–1386.

Abbott RJ, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N *et al.* (2013). Hybridization and speciation. *J Evol Biol* **26**: 229–246.

Abbott RJ, Barton NH, Good JM (2016). Genomics of hybridization and its evolutionary consequences. *Mol Ecol* **25**: 2325–2332.

Anderson E (1949). *Introgressive hybridization*. John Wiley: New York.

Arnold ML (1993). *Iris nelsonii*: origin and genetic composition of a homoploid hybrid species. *Am J Bot* **80**: 577–583.

Arnold ML, Ballerini ES, Brothers AN (2012a). Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. *Heredity* **108**: 159–166.

Arnold ML, Hamlin JA, Brothers AN, Ballerini ES (2012b). Natural hybridization as a catalyst of rapid evolutionary change. In: Singh RS, Xu J, Kulathinal RJ (eds). *Rapidly evolving genes and genetic systems*. Oxford Univ. Press: Oxford. pp 256–265.

Barton NH, Hewitt GM (1985). Analysis of hybrid zones. *Annu Rev Ecol Syst* **16**: 113–148.

Barton NH, Gale KS (1993). Genetic analysis of hybrid zones. In:Harrison RG (ed). *Hybrid zones and the evolutionary process*. Oxford Univ. Press: Oxford. pp 13–45.

Blanco-Pastor JL, Vargas P, Pfeil BE (2012). Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean. *Linaria*. *PLoS One* **7**: e39089.

Brennan AC, Barker D, Hiscock SJ, Abbott RJ (2012). Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of *Senecio squalidus* (Asteraceae). *Heredity* **108**: 87–95.

Buerkle CA, Rieseberg LA (2008). The rate of genome stabilization in homoploid hybrid species. *Evolution* **62**: 266–275.

Butlin RK, Ritchie MG (2013). Pulling together or pulling apart: hybridization in theory and practice. *J Evol Biol* **26**: 294–298.

Coyne JA, Orr HA (2004). *Speciation*. Sinauer Associates: Boston, Massachusetts.

Elgvin TO, Hermansen JS, Fijarczyk A, Bonnet T, Borge T, Sæther SA *et al.* (2011). Hybrid speciation in sparrows II: a role for sex chromosomes? *Mol Ecol* **20**: 3823–3837.

Feder JL, Egan SP, Nosil P (2012). The genomics of speciation-with-gene-flow. *Trends Genet.* **28**: 342–350.

Gao JIE, Wang B, Mao JF, Ingvarsson P, Zeng QY, Wang XR (2012). Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. *Mol Ecol* **21**: 4811–4827.

Grant V (1981). *Plant speciation* 2nd edn, Columbia Univ. Press: New York.

Grant PR, Grant BR (1994). Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution* **48**: 297–316.

Grant PR, Grant BR (2002). Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**: 707–711.

Grant PR, Grant BR (2015). Introgressive hybridization and natural selection in Darwin's finches. *Biol J Linn Soc* **117**: 812–822.

Gross BL, Rieseberg LH (2005). The ecological genetics of homoploid hybrid speciation. *J Hered* **96**: 241–252.

Harrison RG (ed) (1993). *Hybrid zones and the evolutionary process*. Oxford Univ. Press: New York.

James JK, Abbott RJ (2005). Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution* **59**: 2533–2547.

Kunte K, Shea C, Aardema ML, Scriber JM, Juenger TE, Gilbert LE *et al.* (2011). Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. *PLoS Genet* **7**: e1002274.

Lafon-Placette C, Köhler C (2016). Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol Ecol* **25**: 2620–2629.

Lai Z, Nakazato T, Salmaso M, Burke JM, Tang S, Knapp SJ *et al.* (2005). Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* **171**: 291–303.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A *et al.* (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**: 371–375.

Larsen PA, Marchán-Rivadeneira MR, Baker RJ (2010). Natural hybridization generates mammalian lineage with species characteristics. *Proc Natl Acad Sci USA* **107**: 11447–11452.

Lotsy JP (1916). *Evolution by Means of Hybridization*. M. Nijhoff: The Hague, The Netherlands.

Lowry DB, Gould BA. (2016). Speciation continuum. In: Kliman R (ed). *Encyclopedia of Evolutionary Biology*. Academic Press: Oxford, UK. pp 159–165.

Lukhtanov VA, Shapoval NA, Anokhin BA, Saifitdinova AF, Kuznetsova VG (2015). Homoploid hybrid speciation and genome evolution via chromosome sorting. *Proc Roy Soc B* **282**: 20150157.

Ma Y, Zhou R, Milne R (2016). Pollinator-mediated isolation may be an underestimated factor in promoting homoploid hybrid speciation. *Front Pl Sci* **7**: 1183.

Mallet J (2005). Hybridization as an invasion of the genome. *Trends Ecol Evol* **20**: 229–237.

Mallet J (2007). Hybrid speciation. *Nature* **446**: 279–283.

Marques I, Jürgens A, Fuertes Aguilar J, Nieto Feliner G (2016). Convergent recruitment of new pollinators is triggered by independent hybridization events in *Narcissus*. *New Phytol* **210**: 731–742.

Martin NH, Willis JH (2007). Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* **61**: 68–82.

Pan J, Zhang D, Sang T (2007). Molecular phylogenetic evidence for the origin of a diploid hybrid of *Paeonia* (Paeoniaceae). *Am J Bot* **94**: 400–408.

Payseur BA, Rieseberg LH (2016). A genomic perspective on hybridization and speciation. *Mol Ecol* **25**: 2337–2360.

Pease JB, Guerrero RF, Sherman NA, Hahn MW, Moyle LC (2016). Molecular mechanisms of postmating prezygotic reproductive isolation uncovered by transcriptome analysis. *Mol Ecol* **25**: 2592–2608.

Pennisi E (2016). A shortcut to a species. *Science* **354**: 817–821.

Pereira CSA, Aboim MA, Ráb P, Collares-Pereira MJ (2014). Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). *Heredity* **112**: 343–350.

Ramsey J, Bradshaw HD, Schemske DW (2003). Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**: 1520–1534.

Renaut S, Rowe HC, Ungerer MC, Rieseberg LH (2014). Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. *Phil Trans R Soc B* **369**: 20130345.

Rieseberg LH (1991). Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *Am J Bot* **78**: 1218–1237.

Rieseberg LH (1997). Hybrid origins of plant species. *Annu Rev Ecol Syst* **28**: 359–389.

Rieseberg LH, Soltis DE (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol Trends Pl* **5**: 65–84.

Rieseberg LH, Whitton J, Linder CR (1996). Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Bot Neerl* **45**: 243–262.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T *et al.* (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211–1216.

Sætre GP (2013). Hybridization is important in evolution, but is speciation? *J Evol Biol* **26**: 256–258.

Salazar C, Baxter SW, Pardo-Diaz C, Wu G, Surridge A, Linares M *et al.* (2010). Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLoS Genet* **6**: e1000930.

Schumer M, Rosenthal GG, Andolfatto P (2014). How common is homoploid hybrid speciation? *Evolution* **68**: 1553–1560.

Seehausen O (2004). Hybridization and adaptive radiation. *Trends Ecol Evol* **19**: 198–207.

Seehausen O (2013). Conditions when hybridization might predispose populations for adaptive radiation. *J Evol Biol* **26**: 279–281.

Selz OM, Thommen R, Maan ME, Seehausen O (2014). Behavioural isolation may facilitate homoploid hybrid speciation in cichlid fish. *J Evol Biol* **27**: 275–289.

Servedio MR, Hermisson J, Doorn GS (2013). Hybridization may rarely promote speciation. *J Evol Biol* **26**: 282–285.

Smadja CM, Butlin RK (2011). A framework for comparing processes of speciation in the presence of gene flow. *Mol Ecol* **20**: 5123–5140.

Sobel JM, Chen GF (2014). Unification of methods for estimating the strength of reproductive isolation. *Evolution* **68**: 1511–1522.

Soltis PS (2013). Hybridization, speciation and novelty. *J Evol Biol* **26**: 291–293.

Soltis PS, Soltis DE (2009). The role of hybridization in plant speciation. *Annu Rev Pl Biol* **60**: 561–588.

Soltis PS, Soltis DE (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Pl Biol* **30**: 159–165.

Sousa F, Bertrand YJ, Pfeil BE (2016). Patterns of phylogenetic incongruence in *Medicago* found among six loci. *Pl Syst Evol* **302**: 493–513.

Stebbins GL (1959). The role of hybridization in evolution. *Proc Am Philos Soc* **103**: 231–251.

Stemshorn KC, Reed FA, Nolte AW, Tautz D (2011). Rapid formation of distinct hybrid lineages after secondary contact of two fish species (*Cottus* spec.). *Mol Ecol* **20**: 1475–1491.

Taylor SJ, Rojas LD, Ho SW, Martin NH (2013). Genomic collinearity and the genetic architecture of floral differences between the homoploid hybrid species *Iris nelsonii* and one of its progenitors, *Iris hexagona*. *Heredity* **110**: 63–70.

Vereecken NJ, Cozzolino S, Schiestl FP (2010). Hybrid floral scent novelty drives pollinator shift in sexually deceptive orchids. *BMC Evol Biol* **10**: 103.

Wang XR, Szmidt AE, Savolainen O (2001). Genetic composition and diploid hybrid speciation of a high mountain pine, *Pinus densata*, native to the Tibetan Plateau. *Genetics* **159**: 337–346.

Wang B, Mao JF, Gao JIE, Zhao WEI, Wang XR (2011). Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Mol Ecol* **20**: 3796–3811.

Wiens JJ (2004). What is speciation and how should we study it? *Am Nat* **163**: 914–923.

Wolfe AD, Xiang QY, Kephart SK (1998). Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proc Natl Acad Sci USA* **95**: 5112–5115.

Yakimowski SB, Rieseberg LH (2014). The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *Am J Bot* **101**: 1247–1258.