UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Biológicas
Departamento de Genética

TESIS DOCTORAL

Genetic structure of forest trees in biodiversity hotspots at different spatial scales

Estructura genética de árboles forestales en regiones de alta biodiversidad a diferentes escalas espaciales

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Katharina Birgit Budde

Directores

Santiago C. González-Martínez
Myriam Heuertz

Madrid, 2014

GENETIC STRUCTURE OF FOREST TREES IN BIODIVERSITY HOTSPOTS AT DIFFERENT SPATIAL SCALES

(Estructura genética de árboles forestales en regiones de alta biodiversidad a diferentes escalas espaciales)

PhD thesis presented by Katharina Birgit Budde

2014

Centro de Investigación Forestal (CIFOR)
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)

Universidad Complutense de Madrid
Facultad de Ciencias Biológicas
Departamento de Genética

Universidad Complutense de Madrid
Facultad de Ciencias Biológicas
Departamento de Genética

Genetic structure of forest trees in biodiversity hotspots at different spatial scales

(Estructura genética de árboles forestales en regiones de alta biodiversidad a diferentes escalas espaciales)

Memoria presentada por Katharina Birgit Budde para optar al grado de Doctor Europeo en Ciencias Biológicas, dirigida por el Doctor Santiago C. González-Martínez y la Doctora Myriam Heuertz (Departamento de Genética y Ecología Forestal), del Centro de Investigación Forestal, INIA-CIFOR.

Madrid, 2014

La Doctoranda	V°B° del Director	V°B° de la Directora
Katharina B. Budde	*Dr Santiago C. González-Martínez*	*Dra Myriam Heuertz*

Prologue

This PhD thesis has been conducted at the Forest Research Centre INIA-CIFOR, supervised by Dr Santiago C. González-Martínez and Dr Myriam Heuertz in Madrid. The adviser at the University Complutense of Madrid was Dr María del Pilar de Arana Montes. The PhD scholarship (FPI, Formación de Personal Investigador) was funded by the Spanish Ministry of Science and Innovation (ref. BES-2009-015443). The research included in this thesis was financed by the "Plan Nacional de I+D" Spanish national projects VaMPiro CGL2008-05289-C02-02 and AFFLORA CGL2012-40129-C02-02 and the European project LinkTree, BiodivERsA ERAnet.

Through my studies of Biology at the University of Marburg (Germany) and the University of Costa Rica (Costa Rica) with emphasis on botany, nature conservation and ecology, I became interested in the population genetics of European and tropical tree species. Later on, my curiosity to understand evolutionary processes was intensified during my Master of Evolutionary Biology at the University Complutense of Madrid (Spain). Luckily, I could combine these interests during my PhD.

The present thesis comprises five chapters which have partly already been published in international journals listed in the Science Citation Index (SCI) or are expected to be published in the near future. The studies can be grouped in three sections encompassing issues related to I.) Phylogeography (Chapter 1), II.) Spatial genetic structure (Chapters 2-4) and III.) Molecular basis of adaptation in nature (Chapter 5). In these publications or manuscripts I have the leading role as first author. Furthermore, during my PhD, I contributed as co-author to two additional publications related to the thesis and that will be cited throughout it.

Hardy OJ, Born C, **Budde KB**, Daïnou K, Dauby G, Duminil J *et al.* (2013) Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geoscience*, **345**, 284–296.

Pinzauti F, Sebastiani F, **Budde KB**, Fady B, González-Martínez SC, Vendramin GG (2012) Nuclear microsatellites for *Pinus pinea* (Pinaceae), a genetically depauperate tree, and their transferability to *P. halepensis*. *American Journal of Botany*, **9**, E362-E365.

INDEX

Resumen

Resumen

Introducción

Los bosques son los "pulmones verdes" de nuestro planeta y proporcionan servicios ecosistémicos importantes, de gran valor para el mantenimiento de los ecosistemas terrestres. Sin embargo, la cubierta forestal está disminuyendo, principalmente debido al impacto humano. Con el objetivo de preservar la biodiversidad y los ecosistemas, incluyendo los bosques, se han desarrollado diferentes estrategias de conservación. En particular, se han identificado regiones de alta biodiversidad (que son también muy vulnerables) como prioridades de conservación, incluyendo la cuenca del Mediterráneo y los bosques tropicales de África y América. Hoy en día, una de las principales amenazas a la biodiversidad es el cambio climático. Se cree que las poblaciones (o especies) que sean incapaces de soportar las nuevas condiciones ambientales a través de respuestas plásticas, tendrán que adaptarse o migrar para no extinguirse. El mantenimiento de la diversidad genética en las poblaciones naturales es un factor clave para su adaptación al cambio climático. Para poder predecir su futura supervivencia o para poder implementar estrategias de conservación adecuadas, es de suma importancia entender correctamente los procesos y factores ecológicos históricos y contemporáneos que configuran la diversidad y la estructura genética espacial de las poblaciones y especies.

La filogeografía tiene como objetivo identificar linajes genealógicos en un marco histórico y a escalas espaciales amplias, con el fin de comprender los procesos que crearon la divergencia dentro y entre especies que están relacionadas estrechamente. Se espera que la divergencia genética entre las poblaciones de una especie surja, por ejemplo, cuando los cambios ambientales conducen a la fragmentación de la población. En condiciones de aislamiento, las poblaciones naturales evolucionan de forma independiente a través de procesos genéticos neutrales y/o adaptativos. La interrupción del flujo de genes puede dar lugar a la divergencia entre los linajes genéticos o poblaciones. Este proceso se ve reforzado por el aumento de la deriva genética en poblaciones pequeñas. Las huellas genéticas de aislamiento del pasado todavía se pueden detectar con marcadores moleculares. Un problema típico que se aborda en estudios filogeográficos es el estudio de las contracciones poblacionales y la fragmentación causada por las oscilaciones climáticas del pasado. Los datos paleoecológicos sugieren que los bosques tropicales de la región guineo-congoleña en África podrían haber persistido durante los períodos más fríos y secos del Pleistoceno en refugios forestales bien delimitados. Sin embargo, los estudios filogeográficos han comenzado solo recientemente a desvelar la historia de las especies arbóreas de la selva

tropical, mientras que la ubicación de las rutas de recolonización de refugios glaciales son más conocidas para especies de zonas templadas.

La naturaleza sésil de las plantas promueve típicamente la existencia de estructura genética espacial, es decir una distribución no aleatoria de los genotipos. Las plantas dependen del polen y las semillas para su dispersión, pero sus capacidades de dispersión son a menudo limitadas. El término “aislamiento por distancia” (“Isolation by distance”, IBD) se refiere a la mayor diferenciación genética entre poblaciones (o individuos) en función de la distancia causada por la reducción progresiva del flujo de genes. Las consecuencias evolutivas del IBD pueden ser múltiples, por ejemplo una mayor probabilidad de apareamiento entre vecinos emparentados puede aumentar la consanguinidad y disminuir la diversidad genética, influyendo así en la evolución de los sistemas de reproducción.

Sin embargo, la limitación de la dispersión no es el único factor que regula el IBD puesto que la selección o los procesos de colonización pueden causar patrones similares. Los estudios recientes de “genética del paisaje” (“landscape genetics”) han comenzado a identificar las características del paisaje, tales como la topografía de las montañas, la posición de los ríos y las carreteras, o la fragmentación antropogénica de hábitats, que afectan de manera significativa el flujo de genes. Así, en un contexto de paisaje, se ha reconocido el efecto de los factores y procesos ecológicos sobre la estructura genética espacial. A escala local, sin embargo, se espera que el flujo de genes entre las poblaciones supere el equilibrio entre selección y migración. En dicho caso, la capacidad de dispersión es entonces el principal determinante de la estructura genética espacial a escala fina (SGS) dentro de las poblaciones. Sin embargo, es frecuente considerar que la selección contribuya también a la formación de SGS.

En la presente Tesis Doctoral, se han examinado la diversidad y la estructura genética espacial en dos casos de estudio, a diferentes escalas espaciales. El desarrollo de análisis paralelos en una especie arbórea tropical, *Symphonia globulifera* L.f., y dos especies de pinos Mediterráneos*, Pinus pinaster* Aiton y *Pinus halepensis* Miller, me permitieron ampliar el conocimiento sobre los factores ambientales que determinan la estructura genética de los árboles forestales.

Symphonia globulifera es una especie arbórea de la selva tropical con gran amplitud ecológica y está presente en las regiones tropicales de África y América. Las flores, de color rojo, se polinizan mediante insectos y los frutos se dispersan por pequeños mamíferos, como los murciélagos, monos, rumiantes y roedores que acaparan semillas. Dado que *S. globulifera* es una especie ancestral, con amplia distribución en el África

Ecuatorial Atlántica anterior al Pleistoceno, esta especie representa un caso ideal para testar la hipótesis de los refugios forestales. Por otra parte, las diferencias en el tipo de polinizadores y dispersores de semillas en África y América pudieron haber producido patrones de SGS contrastadas, lo que constituye una hipótesis interesante para testar con marcadores moleculares.

Pinus pinaster y *P. halepensis* son especies típicas del bosque Mediterráneo que crecen bajo condiciones secas durante el verano. Las flores se polinizan por el viento y el viento también dispersa sus semillas aladas. Ambas especies muestran adaptaciones notables a los diferentes regímenes de incendios bajo los cuales habitan. La existencia de conos serótinos, que permanecen cerrados en los árboles y forman un banco de semillas aéreo, es de suma importancia. La apertura de los conos serótinos se induce por temperaturas altas (por ejemplo, durante los eventos de fuego). En el este de la Península Ibérica, las poblaciones de pino con baja recurrencia de incendios (LoFi) conviven con otras afectados por incendios frecuentes (HiFi). Esta situación ofrece una oportunidad extraordinaria para estudiar los efectos demográficos y selectivos inducidos por diferentes regímenes de fuego en las poblaciones naturales de los pinos Mediterráneos. Además, la alta variabilidad fenotípica para caracteres relacionados con la incidencia de incendios forestales en la región permitió el desarrollo de un estudio de asociación genética entre marcadores moleculares y serotinia.

Objetivos

El objetivo principal de la presente Tesis Doctoral es evaluar el efecto de los procesos ecológicos y evolutivos históricos y actuales sobre la diversidad y la estructura genética de especies forestales estudiadas a diferentes escalas espaciales. La tesis consiste en cinco estudios complementarios (presentados como capítulos) que se agrupan en tres secciones temáticas.

A escala espacial amplia, se llevó a cabo un estudio de filogeografía en *S. globulifera* procedente del África Ecuatorial Atlántica basado en microsatélites nucleares y secuencias del cloroplasto. El objetivo era testar la teoría de refugios forestales del Pleistoceno y el papel de las barreras geográficas en la estructura genética de las poblaciones (Capítulo 1).

Los efectos de los factores ecológicos sobre la diversidad y la estructura genética espacial a escala regional y local, se trataron en ambos casos de estudio, *S. globulifera* y los pinos Mediterráneos. Para investigar la influencia de distintos agentes de dispersión en los patrones de SGS, se compararon poblaciones de *S. globulifera* de África y América

(Capítulo 2). Además, se examinaron los efectos de regímenes de incendios contrastados sobre la diversidad genética, la historia demográfica y la formación de SGS en las poblaciones naturales de *P. pinaster* y *P. halepensis* (Capítulo 3). El fuego no parece afectar la SGS de las poblaciones de *P. pinaster*, pero se ha encontrado una señal de selección potencial que estaría correlacionada con otros factores ambientales. Por lo tanto, se desarrolló también un estudio sobre el efecto de algunos factores ecológicos relevantes en las poblaciones de estudio (altitud, pendiente, orientación, cubierta vegetal herbácea, y riqueza de especies herbáceas) sobre la diferenciación genética entre árboles dentro de poblaciones (Capítulo 4).

Finalmente, considerando la fuerte presión de selección que suponen los fuegos forestales y mis descubrimientos en los capítulos anteriores, se llevó a cabo un estudio de asociación genética para identificar loci relacionados con fenotipos de fuego, reflejados por un alto grado de serotinia (Capítulo 5).

Resultados y discusión

En general, a lo largo de los diferentes capítulos de esta Tesis Doctoral, describo el fuerte efecto que los factores y procesos ambientales tienen sobre la variabilidad genética y la estructura genética espacial en especies arbóreas. El estudio de filogeografía en *S. globulifera* reveló una alta diversidad genética, tanto con los microsatélites nucleares como con las secuencias del cloroplasto (*psbA-trnH*). Los valores de diversidad obtenidos fueron similares a los encontrados por otros autores en poblaciones americanas de esta especie. Los datos de microsatélites nucleares sugieren, además, la existencia de cuatro grupos genéticos (GPs) en el África Ecuatorial Atlántica: el GP1 consiste en las poblaciones costeras de Benin, el oeste de Camerún representa el GP2, el GP3 abarca el sur de Camerún y Gabón, y el GP4 está presente en São Tomé. El patrón de divergencia entre los GPs sugiere un papel importante de las barreras geográficas (tales como la línea volcánica de Camerún, el corredor Togo-Dahomey y el Golfo de Biafra) en la conformación de la estructura genética de *S. globulifera*. Por último, se detectaron señales de cuellos de botella demográficos en todos los grupos genéticos, lo que probablemente esté relacionado con la existencia de contracciones poblacionales durante el último máximo glacial (LGM), al menos en el caso de Benin (GP1), el oeste de Camerún (GP2) y São Tomé (GP4). Nuestros resultados proporcionan un apoyo parcial a la hipótesis de los refugios forestales y a algunos de los lugares de refugio propuestos por Maley (1996). Sin embargo, se detectó una señal demográfica de cuello de botella más antigua en la región de Camerún y Gabón (GP3), lo que sugiere que el LGM no causó cambios de tamaño de población tan fuertes en esta región. El elevado número de

haplotipos endémicos con rangos de distribución estrechos indica también la persistencia de *S. globulifera* tanto dentro como fuera de lugares postulados como refugios forestales.

Los patrones filogeográficos de *S. globulifera* contrastan con los de otras especies arbóreas de la selva tropical africana. Aunque es difícil generalizar, una amplia gama de estudios indican respuestas específicas de las especies a los cambios ambientales pasados, probablemente relacionadas con rasgos de historia vital distintos. En cualquier caso, la divergencia genética entre la Alta y Baja Guinea es congruente entre las especies estudiadas hasta la fecha. Con respecto a la hipótesis de las contracciones de los bosques y la supervivencia en los refugios forestales durante las oscilaciones climáticas del Pleistoceno, la evidencia es también contradictoria, con algunas especies apoyando dicha teoría pero otras no.

La capacidad limitada de dispersión es el factor principal que influye la SGS. En las especies polinizadas y dispersadas por animales, la presencia y abundancia de vectores son factores clave. Generalmente, la SGS detectada en poblaciones de *S. globulifera* fue similar a la de otras especies con modos de dispersión similares, densidades bajas y apareamiento cruzado (“outcrossing”). Sin embargo, la SGS en las poblaciones neotropicales de *S. globulifera* es menos pronunciada y la dispersión de genes más leptocúrtica que en África. La dispersión de semillas a largas distancias por murciélagos en las poblaciones neotropicales podría ser una de las explicaciones posibles. Por el contrario, la dispersión de polen es aparentemente similar en las poblaciones de ambos continentes. De hecho, los principales polinizadores en ambas regiones, los colibríes en América y el género Nectarina en África, pertenecen al mismo tipo funcional de polinizador.

Aparte de los procesos de dispersión, existen eventos ecológicos (tales como la fragmentación del hábitat) que pueden afectar también a la SGS de las poblaciones arbóreas. La comparación de poblaciones naturales de *P. pinaster* y *P. halepensis* afectadas por regímenes de fuego contrastados (HiFi vs. LoFi, ver arriba) reveló que los incendios forestales cuando son frecuentes aumentan la SGS en poblaciones de *P. halepensis* en el este de la Península Ibérica. Además, existe una autocorrelación espacial significativa de los fenotipos de serotinia, lo que junto con el aumento de la SGS detectado con marcadores potencialmente adaptativos (SNPs) en las poblaciones de HiFi, parecía reflejar una similitud funcional de árboles vecinos. Estos patrones pueden deberse a la selección microambiental y/o a capacidades de dispersión alteradas inducidas por la frecuencia de los incendios forestales. Curiosamente, no se detectó ningún efecto relacionado con el fuego en la SGS de las poblaciones de *P. pinaster*. A pesar del efecto

claro del fuego sobre la dinámica poblacional de *P. halepensis*, la diversidad genética poblacional e historia demográfica son similares cuando se comparan poblaciones de HiFi y LoFi de ambas especies. Este hecho apunta a la fuerte resiliencia de las poblaciones de pinos Mediterráneos, una característica facilitada por la fuerte adaptación de estas especies a los incendios, en particular, gracias al enorme banco de semillas aéreo resultante de la existencia de conos serótinos y a una edad temprana de la primera floración.

Sorprendentemente, la SGS en *P. pinaster* no estaba afectada por los diferentes regímenes de fuegos estudiados. Sin embargo, la fuerza de la SGS difería entre las tres poblaciones analizadas. Dado que dichas poblaciones son ecológicamente heterogéneas, este resultado sugiere un papel de los factores ambientales no relacionados con los incendios forestales en la conformación de la SGS. A pesar de la creciente evidencia sobre la contribución de la heterogeneidad microambiental a la SGS, todavía no se conocen los factores ambientales subyacentes. Al correlacionar las distancias genéticas con distancias ambientales (controlando por los efectos de autocorrelación espacial) en mi estudio, encontré una diferenciación genética significativa a lo largo de un gradiente altitudinal de 300 m en la población de Eslida (este de la Península Ibérica). Las diferencias genéticas encontradas podrían tener origen, por ejemplo, en una respuesta plástica de la fenología de la floración debida a las diferentes temperaturas que los árboles soportan a lo largo de gradientes altitudinales acusados, que a su vez evitan la sincronía de la floración entre la parte alta y baja de la montaña. Sin embargo, algunos efectos selectivos, como la existencia de diferencias en la germinación y reclutamiento de semillas inmigrantes de alturas superiores o inferiores, podrían ser también una explicación. Los estudios futuros en esta zona deberían tratar de identificar mejor los factores ambientales subyacentes que cambian a lo largo de esta clina altitudinal, a la vez que validar mis resultados en otros gradientes altitudinales independientes.

Teniendo en cuenta que el fuego juega un papel importante en la conformación de los rasgos adaptativos de las plantas, llevé a cabo también un estudio de asociación genética para fenotipos relacionados con el fuego en poblaciones naturales. La alta variabilidad fenotípica y la ausencia de estructura genética de poblaciones en el este de la Península Ibérica contribuyeron al éxito de este estudio. Diecisiete SNPs se asociaron a los fenotipos de serotinia en *P. pinaster*. Sobre la base de estos 17 SNPs, se creó un modelo que predice el 29% de la variabilidad fenotípica. Estos valores son similares en estudios de asociación de otros caracteres adaptativos de especies de coníferas. La precisión del modelo predictivo fue mayor para las poblaciones dentro del mismo linaje materno (tal como las poblaciones en el este de la Península Ibérica), pero disminuyó en otros más

lejanos. Los estudios futuros deberían tratar de identificar la función de estos genes potencialmente asociados con rasgos adaptativos de respuesta a los incendios forestales.

Summary

Summary

Introduction

Forests are the “green lungs” of our planet; they provide important ecosystem services and are of utmost importance for the maintenance of the Earth’s terrestrial ecosystems. Nevertheless, forest cover is decreasing mainly because of human impacts. Different conservation strategies target the preservation of biodiversity and ecosystems, including forests. Biodiversity hotspots, which are highly jeopardized and have outstanding species richness, have been identified as conservation priorities, including the Mediterranean Basin and the tropical forests of Africa and America. Nowadays, one of the major threats to biodiversity is climate change, which is expected to force plant populations to adapt, migrate or go extinct, unless they are able to keep up with altered climatic conditions through plastic responses. The maintenance of genetic diversity in natural populations is a key factor for adaptation to climate change. The understanding of how historic and contemporary ecological processes and factors shape the genetic diversity and spatial genetic structure of populations and species is therefore of extreme importance to predict their future survival or to implement suitable conservation strategies.

Phylogeography aims to identify patterns of genealogical lineages in a historical framework and at wide spatial scales, with the purpose of understanding the processes that created divergence within and among closely related species. Genetic divergence among populations of a species is expected to arise, for instance, when environmental changes lead to population fragmentation. In isolation, populations evolve independently through neutral and/or adaptive genetic processes. The interruption of gene flow can lead to divergence among genetic lineages or populations, a process that is enhanced by increased genetic drift in small populations. The genetic footprints of past isolation can still be detected with molecular markers. A typical issue addressed in phylogeographic studies is that of species range contractions and fragmentation caused by past climatic oscillations. Palaeoecological data suggested that tropical rainforests of the Guineo-Congolian region in Africa possibly persisted in a few major refugia, during colder and drier periods of the Pleistocene. Only recently, phylogeographic studies have started to unveil the history of tropical rain forest tree species, while the location of and recolonization routes from glacial refugia are better known for temperate species.

The sessile nature of plants typically promotes spatial genetic structure, a non-random distribution of genotypes. Plants rely on pollen and seeds to disperse, but their dispersal abilities are often limited. Isolation by distance (IBD) refers to the typically higher genetic

differentiation between geographically distant than nearby populations (or individuals) produced by reduced gene flow. The evolutionary consequences of IBD can be manifold; for example a higher mating probability between related neighbors can increase bi-parental inbreeding and decrease genetic diversity, thus influencing the evolution of mating systems. However, limited dispersal is not the only factor shaping IBD, as selection or serial colonization processes can cause similar patterns. Recently, landscape genetic studies started to identify landscape features, such as the topography of mountains, rivers, roads or anthropogenic habitat fragmentation, that significantly affect gene flow. Thus, in a landscape context, the effect of ecological factors and processes on the spatial genetic structure has been recognized. At local scales, however, gene flow within populations is expected to exceed the selection-migration equilibrium. Dispersal abilities are then the strongest determinants of fine-scale spatial genetic structure (SGS) within populations. However, selection has often been suspected to contribute to SGS. Identifying the underlying factors and processes (apart from limited dispersal) causing SGS within populations remains a challenge.

In the present thesis, genetic diversity and spatial genetic structure have been examined in two case studies at different spatial scales. Parallel analyses of a tropical tree species, *Symphonia globulifera* L.f., and two Mediterranean pine species, *Pinus pinaster* Aiton and *Pinus halepensis* Miller, allowed me to produce more comprehensive views on relevant biotic and abiotic environmental factors shaping genetic structure in forest trees.

Symphonia globulifera is a rainforest tree species with wide ecological amplitude, present in tropical regions of Africa and America. The scarlet red flowers are insect- and bird-pollinated and fruits are dispersed by small mammals, such as bats, monkeys, ruminants or scatter hoarding rodents. As an ancient species with a wide distribution throughout Atlantic Equatorial Africa anterior to the Pleistocene, *S. globulifera* represents an ideal study case to test the forest refuge hypothesis. Furthermore, differences in the assembly of pollinators and seed dispersers in Africa and America may have produced contrasted SGS patterns.

Pinus pinaster and *P. halepensis* are typical Mediterranean tree species that grow under summer dry conditions. The flowers are wind-pollinated and the winged seeds are wind-dispersed. Both species show adaptations to different fire regimes, most importantly serotinous cones, which remain closed on the trees and form a canopy seed bank until high temperatures (e.g. during fire events) cause cone opening and seed release. In the east of the Iberian Peninsula, pine populations with low recurrence of fires (LoFi) cohabit with others under frequent fires (HiFi). This situation offered a remarkable opportunity to

study the demographic and selective effects of fire regimes on natural Mediterranean pine populations. In addition, high phenotypic variability for fire phenotypes in the region allowed the study of the molecular basis of fire adaptation.

Objectives

The main aim of the present thesis is to evaluate the effects of historic and ongoing ecological and evolutionary processes on genetic diversity and structure of tree species at different spatial scales. The thesis consists in five different studies (i.e., chapters) that are grouped in three sections.

At a wide spatial scale, a phylogeographic study was conducted for *S. globulifera* in Atlantic Equatorial Africa based on nuclear microsatellites and chloroplast sequences, in order to test the Pleistocene forest refuge theory and the role of geographical barriers in population genetic structure of tropical species (Chapter 1).

The effects of ecological factors on the genetic diversity and spatial genetic structure at regional and local scale were addressed in both *S. globulifera* and Mediterranean pines. To investigate the influence of distinct dispersal agents on SGS patterns, *S. globulifera* populations from Africa and America were compared (Chapter 2). Furthermore, the impact of distinct fire regimes on the genetic diversity, demographic history and SGS in natural populations of *P. pinaster* and *P. halepensis* were examined (Chapter 3). No fire-related effect was reflected in SGS patterns from *P. pinaster* populations, but I found potential evidence of selection due to micro-environmental factors. Thus, the influence of ecological factors (altitude, slope, aspect, herbaceous plant cover, and herbaceous species richness) on genetic differentiation among trees within-populations was tested (Chapter 4).

Finally, based on previous knowledge of strong selection pressure of wildfire and my discoveries in previous Chapters, a genetic association study was conducted to identify loci related to fire phenotypes, as gauged by serotiny (Chapter 5).

Results and discussion

Overall, throughout the different studies that constitute this PhD thesis, I have found a strong effect of environmental factors and processes on forest tree genetic variability and spatial genetic structure. The phylogeographic study *of Symphonia globulifera* revealed high genetic diversity for nuclear microsatellites and *psbA-trnH* chloroplast sequences. Values were similar to the ones reported from American populations of this species. Nuclear microsatellites data exposed four gene pools (GPs) in Atlantic Equatorial Africa:

GP1 consisting of coastal populations in Benin, GP2 representing West Cameroon, GP3 encompassing South Cameroon and Gabon and GP4 present in São Tomé. The divergence pattern of GPs suggested geographical features, such as the Cameroon volcanic line, the Dahomey gap and the Bight of Bonny to play a role in shaping the genetic structure. Bottleneck signals were detected in all GPs, most likely related to population contractions during the LGM in the case of Benin (GP1), West Cameroon (GP2) and São Tomé (GP4). Our results provided partial support for the forest refuge hypothesis and some of the refuge locations proposed by Maley (1996). However, an older bottleneck signal was detected in the Cameroon and Gabon region (GP3), suggesting that LGM climatic oscillations did not cause as strong population size changes in this region as compared to other GPs and/or that GP3 populations recovered better. The high number of endemic haplotypes with narrow ranges indicated long-term persistence of *S. globulifera* inside and outside of postulated forest refuge locations.

Phylogeographic patterns in *S. globulifera* contrasted with those found in other African rain forest tree species. Although generalizations are difficult to draw, a wide range of studies indicated species-specific responses to past environmental changes that are probably related to distinct life-history traits. Congruent among species, however, is the genetic divergence between Upper and Lower Guinea. Furthermore, the hypothesis of forest contractions and survival in refugia during Pleistocene climatic oscillations is supported by some species but not by others.

Limited dispersal is the main factor for the built-up of SGS. In animal pollinated and dispersed species, the presence and abundance of vectors are key determinants. Generally, SGS detected in populations of *S. globulifera* was similar to other species with similar dispersal modes, low individual densities and predominantly outcrossing mating systems. However, the SGS in Neotropical populations of *S. globulifera* was slightly less pronounced and gene dispersal was more leptokurtic than in Africa. Long-distance seed dispersal by bats in Neotropical populations could be a plausible explanation. In contrast, pollen dispersal was possibly similar in populations from both continents. In fact, the main pollinators, hummingbirds in America and sunbirds in Africa, belong to the same functional pollinator type.

In addition to dispersal processes, ecological events such as habitat fragmentation can affect the SGS of tree populations. The comparison of natural *P. pinaster* and *P. halepensis* stands under contrasted fire regimes (HiFi vs. LoFi, see above) revealed that wildfire is a driver of SGS in *P. halepensis* HiFi populations from the east of the Iberian Peninsula. In HiFi populations, significant spatial autocorrelation of serotiny phenotypes

and increased SGS at single nucleotide polymorphism (SNP) markers, seemed to reflect a functional similarity of neighboring trees. These patterns can be due to micro-environmental selection and/or altered dispersal capacities induced by frequent stand-replacing fires. Interestingly, no fire-related effect was found in the SGS of *P. pinaster* stands. Despite the clear effect of fire on the population dynamics in *P. halepensis*, the genetic diversity and demographic history was similar in HiFi and LoFi stands of both species, respectively. This fact points to strong stand resilience facilitated by fire adaptive traits, in particular the large canopy seed bank resulting from serotinous cones and the early age of first flowering.

Remarkably, SGS in *P. pinaster* was not affected by distinct fire regimes. However, the strength of SGS differed among the three ecologically diverse stands studied, suggesting a role of environmental factors unrelated to wildfire. Although micro-environmental heterogeneity has frequently been suspected to contribute to SGS, the underlying environmental factors remain obscure. By correlating genetic distances with environmental distances, while controlling for spatial autocorrelation effects, I found significant genetic differentiation along a 300 m altitudinal gradient in the population from Eslida. Genetic differences could be caused, for instance, by a plastic response of the flowering phenology to different temperatures along altitudinal gradients, in turn leading to non-overlapping flowering in upper and lower parts. However, selective effects, such as limited germination and establishment success of immigrating seeds from upper or lower altitude, could also be a plausible explanation. Future studies should aim at further identifying the underlying environmental factors that change along this altitudinal cline and to validate results in other independent gradients.

Bearing in mind that fire plays an important role in shaping adaptive plant traits, I conducted a genetic association study for fire phenotypes in natural conditions. The high phenotypic variability and absence of population genetic structure in the east of the Iberian Peninsula facilitated this study. Seventeen SNPs were potentially associated to fire phenotypes, as gauged by serotiny, in *P. pinaster*. Based on these 17 SNPs it was possible to predict 29% of the phenotypic variability. Similar values have been obtained for association studies of other adaptive traits in conifer species. The accuracy of the predictive model was best for populations of the same maternal lineage as the focal populations, in the east of the Iberian Peninsula, but decreased in other maternal lineages. Future studies should try to disentangle the function of genes potentially associated to fire adaptive traits.

General part

General Introduction

Forests are of utmost importance for the functioning of the Earth's ecosystems. They cover 31% of the total land area (FAO 2010) and make up around 90% of the Earth's biomass (Whittaker 1975). Forests provide important ecosystem services such as carbon stocks, purification of air, maintenance of biodiversity, generation and renewal of soil and soil fertility, mitigation of droughts and floods, avoidance of erosion, etc. (Daily 1997). However, forest cover is decreasing: 13 million hectares of forests per year were destroyed by human use or natural causes during the last decade (FAO 2010).

Different conservation strategies have been defined to efficiently preserve biodiversity during the last decades. To protect the most species with the least costs, biodiversity hotspots have been identified (Myers 1988, 1990; Mittermeier & Myers 1998), including the Mediterranean Basin and West African Forests (Myers 1990; Médail & Quézel 1999). These hotspots cover 44% of all vascular plant species while they only encompass 1.4 % of the Earth's landmass and they are threatened by exceptional habitat loss (Figure 1, Myer et al. 2000). In the particular case of Mediterranean and tropical forests, the extinction of one species can have severe effects on associated species, such as pollinators, pathogens, herbivores, seed predators and dispersers. It has been estimated that the disappearance of one plant species can lead to between 30 to 150 extinctions of other taxa (Raven 1976). However, the loss of a rare herb species most likely affects less species than the extinction of a widespread forest tree (Myers 1990).

Climate change is one of the major threats to biodiversity and also to natural forest ecosystems (Thomas *et al.* 2004). Trees, which are keystone elements in forests and provide habitat for many species, are long-lived, sessile organisms that have to cope with continuous changes in their surroundings. If current tree populations are not able to tolerate new climatic conditions, they will have to 1) adapt, 2) migrate or 3) go extinct (Aitken *et al.* 2008). Long generation times lead to a slow accumulation of mutations, but many tree populations maintain high genetic diversity and thus are expected to adapt rapidly (Petit & Hampe 2006, Fallour-Rubio et al. 2009). However, whether adaptation and migration will be fast enough to keep up with climate change is still an open question. The understanding of the relevant evolutionary and ecological processes affecting the genetic structure and adaptive capacities of tree species are therefore crucial. Furthermore, establishing experimental and selection studies in species with long life-spans is difficult, therefore trees are typically under-studied (Linhart 2000).

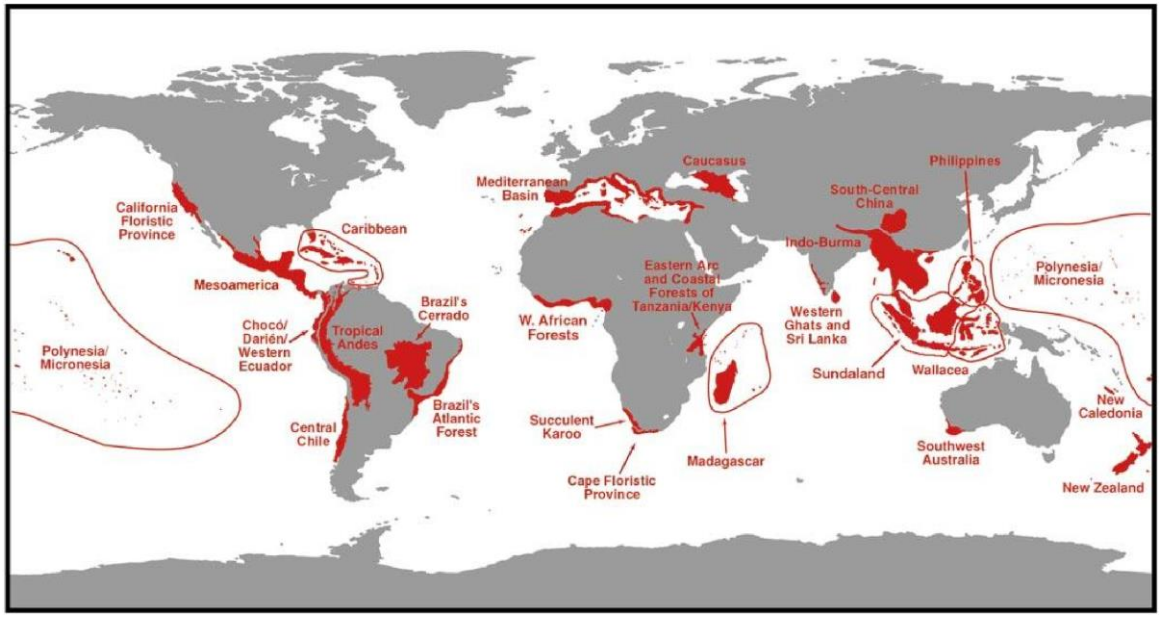

Figure 1 Red areas indicate the location of the 25 biodiversity hotspots identified by Myers *et al.* (2000).

To address multifaceted questions at different spatial and temporal scales, genetic markers provide powerful tools and may enable research sometimes without the need of field experiments. During this thesis, the genetic structure of tree species at different spatial scales was investigated in order to understand the effects of past and on-going ecological impacts on forest ecosystems from biodiversity hotspots. Based on previous knowledge and availability of genetic markers, two different study systems, *Symphonia globulifera* (Clusiaceae) in Africa and two Mediterranean pine species, *Pinus pinaster* and *Pinus halepensis* (Pinaceae), were chosen and specific questions were addressed at the interface of different biological disciplines, as described below.

Phylogeography

Phylogeographical approaches are helpful to investigate the distribution of genetic diversity and to understand patterns of genetic structure at a broad spatial scale. This discipline investigates the geographic structure of genealogical lineages in a historical framework in order to understand the processes that shaped divergence within and among closely related species. Phylogeography is a discipline that emerged from the conjunction of phylogenetics, population genetics and biogeography. Phylogeographic studies are typically performed on a wide geographic scale covering the distribution of a species or closely related species. Similarly to biogeography, phylogeography is based on the principle of vicariance, i.e., populations or species were separated by barriers, such as oceans, mountains or rivers inhibiting gene flow. During isolation, populations or species

evolved independently and distinct genetic lineages arose through neutral and/or adaptive genetic processes. Patterns of the present day genetic variation still reflect these historical processes (Avise 2000). Phylogeographic studies give insights on the evolutionary processes that shaped the distribution of genetic variation and contribute important knowledge for the development of appropriate conservation strategies of species (Moritz 2002).

Investigating how forests responded to past climatic changes is crucial with regard to present-day climate change (Petit *et al.* 2008). An important issue addressed in phylogeographic studies is the location of Pleistocene forest refugia, i.e. places with a stable climate (not exposed to climatic extremes) which provided shelter for forest species during glacial periods. Isolation of populations in distinct refugia commonly causes intraspecific divergence. Numerous studies addressed this topic for tree species from the Northern Hemisphere and identified refuge locations and postglacial migration routes, especially in Europe and North America (Beheregaray 2008), including for the Mediterranean conifers studied in this thesis (Burban & Petit 2003; Gómez *et al.* 2005; Bucci *et al.* 2007). In Europe many species survived adverse climatic conditions in the Balkan, the Iberian, and the Italian Peninsulas, and subsequently recolonized northern territories (e.g. Taberlet *et al.* 1998; Heuertz *et al.* 2004; Petit *et al.* 2005b). Highest genetic divergence is typically found among populations located in former forest refugia, while divergence is lower in recolonized areas due to the admixture of distinct lineages (Petit *et al.* 2003). Apart from major glacial refugia in southern Europe, cryptic refugia further north have been identified for several species (Provan & Bennett 2008). In tropical regions, phylogeographic studies have been conducted more recently. Although conditions were colder and drier during glacial periods (Hamilton & Taylor 1991), there are no clear *a priori* expectations on genetic diversity and structure, as no obvious temperature gradient is assumed. Based on paleoecological data, Maley (1996) and Bonnefille (2007) assumed contractions of forest ecosystems and subsequent expansions into grasslands due to Pleistocene climatic oscillations in all three Guineo-Congolian sub-centers of endemism: Upper Guinea, Lower Guinea and Congolia, as defined by White (1979, 1983; Figure 2). This finding supported the idea of forest refugia also in tropical Africa. Tropical forest refugia are assumed to have accumulated high species richness and promoted the development of endemic species due to long-term climatic stability (Mayr & O'Hara 1986). Several locations have been proposed to have served as such, based on species richness, species endemism and paleoecological data (reviewed by Maley 1996). However, phylogeographic studies of keystone African rain forest species are essential to

complement the knowledge on possible refuge locations and to enable comparative analyses with their Mediterranean and temperate counterparts.

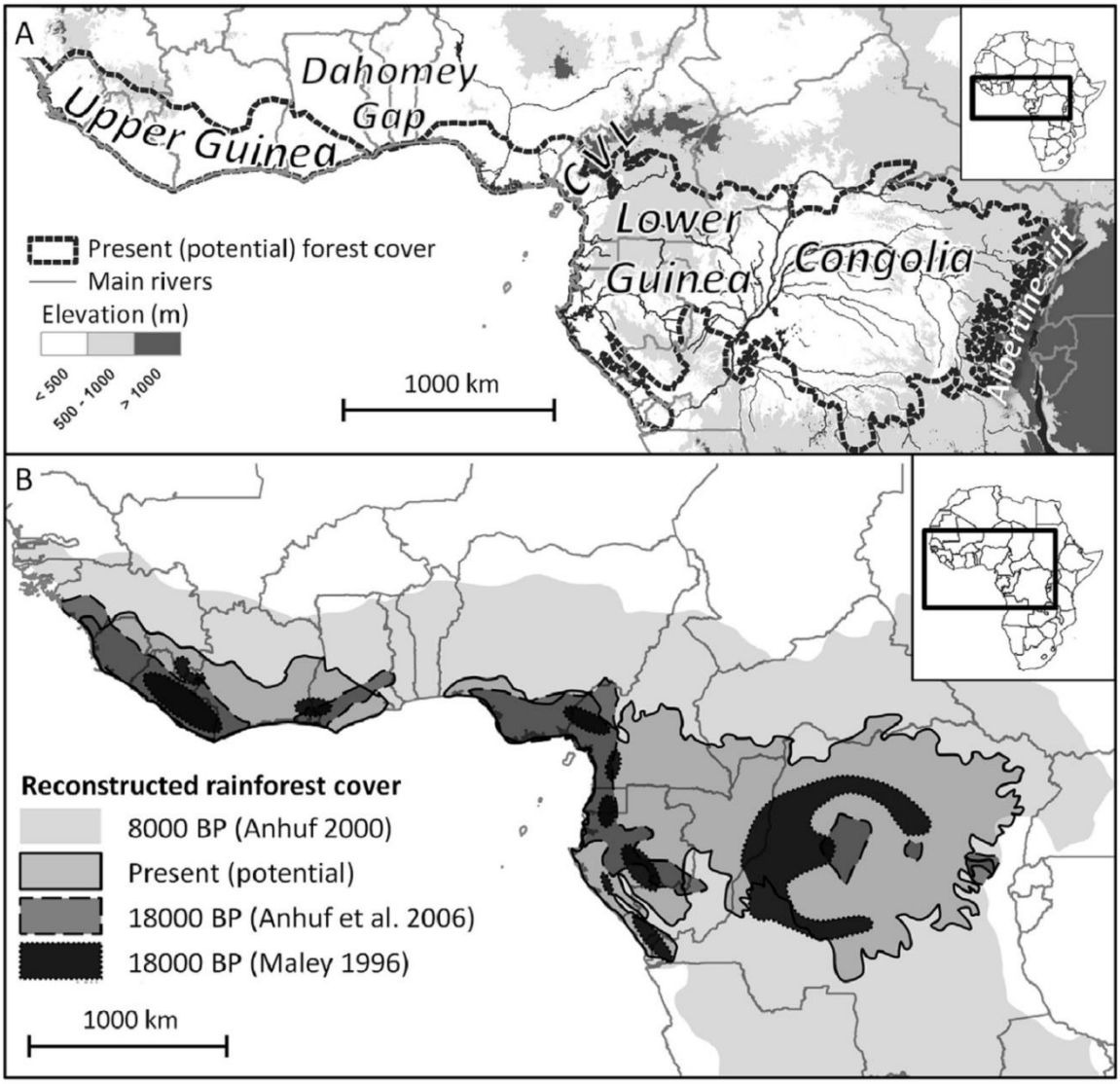

Figure 2 Guineo-Congolian forest delimitation after Hardy *et al.* (2013), subdivision and topography (A), and hypothetical range shifts in the Late Quaternary (B). CVL: Cameroonian Volcanic Line.

Historic processes also shaped the genetic constitution of populations at regional and local scales. Historical ecological impacts or anthropogenic habitat fragmentation can severely reduce effective population sizes and genetic diversity, potentially compromising the survival of populations in changing environments. A high heterozygosity can positively affect individual fitness and high genetic diversity provides better chances for populations to respond to selection pressures (Reed & Frankham 2003). Therefore, the genetic diversity of a population may reflect its evolutionary potential (Frankham *et al.* 2002). Habitat fragmentation has been shown to induce genetic bottlenecks, increase genetic drift

and reduce gene flow, which leads to lower genetic diversity and increased inbreeding (reviewed in Young *et al.* 1996; Aguilar *et al.* 2008). However, the reduction in gene flow between population fragments is not always drastic (de-Lucas *et al.* 2009). Also, natural disasters such as climate change (e.g. glacial periods), floods, volcanic eruptions, pest outbreaks, droughts or wildfires can reduce the effective population sizes drastically. However, natural disturbances have a long history and the dimension of the damage to species and ecosystems depends on species-specific life-history traits and adaptations, and the resilience of ecosystems (Holling 1973). Genetic markers provide important tools to investigate the distribution of genetic diversity and the demographic history of populations in order to understand these past and ongoing threats.

Maternally inherited genetic markers, such as chloroplast DNA sequences (in the case of most Angiosperms) are often used in phylogeographic studies. These markers allow to study the distribution of distinct maternal lineages and show sharper patterns than paternally inherited markers or nuclear DNA (Petit *et al.* 2005a). Chloroplast DNA evolves more slowly than nuclear DNA, does not experience recombination (Palmer 1990), and is usually dispersed by seeds in Angiosperms (due to maternal inheritance). Nevertheless, despite phylogeographic structure is less pronounced in nuclear DNA, the joint analysis of several diploid and independent nuclear genetic markers allows to identify more robust patterns and often gives more detailed views than organelle markers (Hare 2001). Furthermore, statistical approaches have been developed which can deal with different genetic markers and, most importantly, allow testing scenarios based on coalescent theory. Likelihood-based or Bayesian methods enable researchers to calculate the probabilities of genetic data with respect to distinct evolutionary scenarios underlying observed genetic patterns (reviewed in Knowles 2009).

Spatial genetic structure

Spatial genetic structure, the non-random distribution of alleles or genotypes in space is driven by mutation, migration, selection and drift (Wright 1943, 1951). The establishment of spatial genetic structure is influenced by neutral and selective genetic processes, such as dispersal, colonization events, bottlenecks, bi-parental inbreeding or (micro-) environmental adaptation. Genetic patterns are the products of complex space-time processes and are apparent on continental, regional and local scales (Epperson 2003). Similarly to phylogeographic studies, the knowledge of the distribution of present-day genetic variation and the involved processes at different spatial scales are important for species conservation management.

Plants are sessile organisms and due to limited dispersal abilities, isolation by distance (IBD) patterns are characteristic for many continuously distributed plant species and populations. The concept of IBD was introduced by Wright (1943). Under continuous distribution, populations (or individuals) that are spatially closer are also genetically more similar than distant ones. This can be caused by dispersal limitations in conjunction with genetic drift. This theory has been extensively developed, e.g. by Malécot (1968), Kimura & Weiss (1964) or Maruyama (1971). Many empirical studies described IBD patterns based on neutral genetic markers, both among (e.g. Slatkin 1993) and within (e.g. Hardy *et al.* 2006; de-Lucas *et al.* 2009) populations. Patterns of IBD can have notable evolutionary consequences for plants. Because of the higher probability of interaction of related neighbors, they may enhance biparental inbreeding, which might influence the evolution of mating systems and reduce genetic diversity (reviewed in Heywood 1991). However, IBD patterns are not exclusively generated by limited dispersal but can also be caused by other factors, e.g. by selection or serial colonization processes (reviewed in Orsini *et al.* 2013).

The relative contribution of random drift and selective processes to create genetic patterns has already been a subject of debate to S. Wright, R. Fisher and J.B.S. Haldane and remains a challenging topic nowadays (Etheridge 2011). During the last decade, a new discipline, termed landscape genetics, arose. Landscape genetics relates population genetic structure to ecological factors (Manel *et al.* 2003; Holderegger *et al.* 2010). Several recent studies identified genetic discontinuities that coincided with landscape features, such as rivers, roads, mountains, or that could be attributed to anthropogenic habitat fragmentation (reviewed in Storfer *et al.* 2010). Association studies revealed correlations between adaptive genotypes/phenotypes or allele frequencies for particular loci and environmental clines, such as temperature or drought gradients (e.g. Bergmann 1978; Eckert *et al.* 2010; Holliday *et al.* 2010; Samis *et al.* 2012) and outlier-detection approaches revealed loci involved in local adaptation (e.g. Beaumont & Nichols 1996).

Local adaptation to distinct habitats can shape genetic structure both at specific loci and across the genome. Divergence is strongest at loci under positive selection and weaker at neutral loci or loci under balancing selection. However, general barriers to gene flow caused by adaptive divergence and enhanced by genetic drift, can promote genome-wide differentiation and lead to isolation by adaptation (IBA, Nosil *et al.* 2007; Andrew *et al.* 2012). This can cause ecotype divergence with phenotypic and neutral genetic differentiation, a pattern described for walking-sticks (*Timema cristinae*) showing preference for distinct host plants (Nosil *et al.* 2007) and for sunflowers (*Helianthus petiolaris*) growing in sand dune versus non-dune habitat (Andrew *et al.* 2012). Even in

cases where no geographical barriers inhibit gene flow, reduced establishment success of immigrant genotypes can reinforce IBA. However, as patterns of IBD and IBA can be confounded, it is crucial to identify the underlying evolutionary drivers and processes.

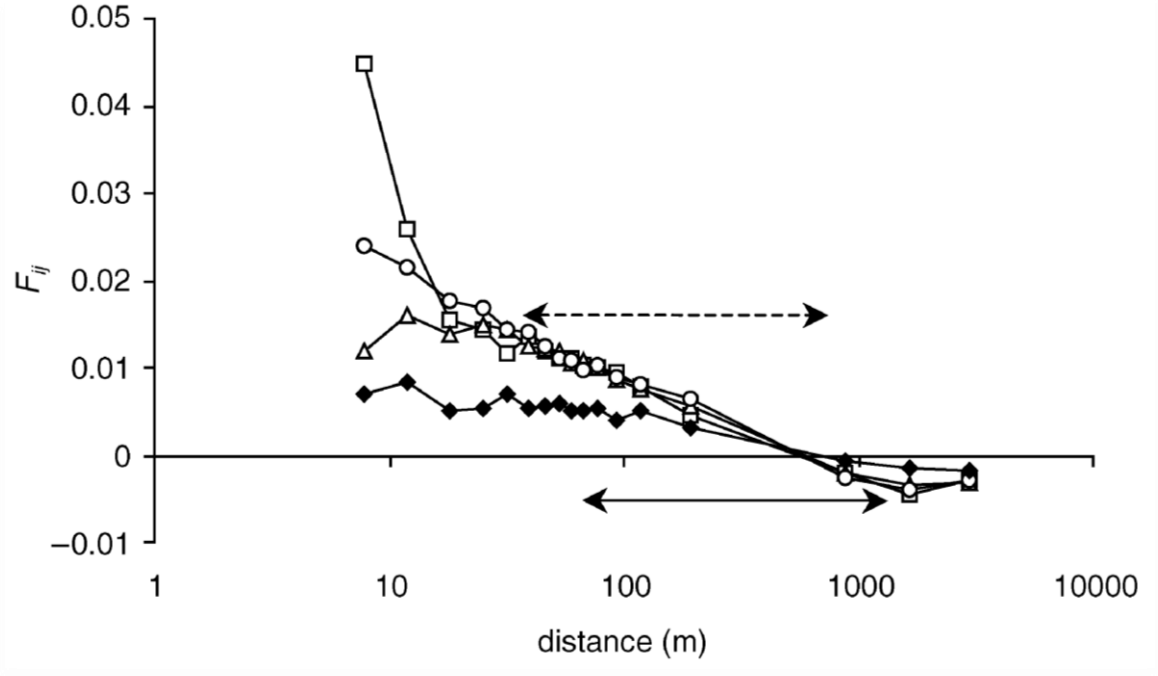

Figure 3 Graphical display of the decay of inter-individual relatedness with spatial distance after Heuertz *et al.* (2003). Average kinship coefficients F_{ij} over 100 independent replicates of four simulated parameter combinations, plotted against the logarithm of distance. Parameters defined for the simulations were σ_s, seed dispersal; σ_p, pollen dispersal and σ_g, gene dispersal. For squares, σ_s=5 m, σ_p=49 m; for circles, σ_s=14 m, σ_p=49 m; for triangles, σ_s=28 m, σ_p=35 m; for diamonds, σ_s=49 m, σ_p=49 m. Open symbols stand for parameter combinations with σ_g≈35–37 m, whereas filled symbols feature σ_g≈60 m. Arrows represent the distance range between σ_g and 20 σ_g, where a linear decrease of F_{ij} is expected, the dashed arrow corresponding to σ_g=37 m, the solid arrow to σ_g=60 m.

At a very local scale, within populations, genotypes are also seldom randomly distributed. Fine-scale spatial genetic structure (SGS) refers to the decay of inter-individual kinship with spatial distance within populations (Figure 3) and can be quantified using the *Sp* statistic (Vekemans & Hardy 2004). This statistic was designed to make SGS patterns comparable between populations, species and different genetic markers (e.g. Hardy *et al.* 2006; Dick *et al.* 2008). Comparative and simulation studies revealed that pollen and seed dispersal abilities are most important in determining the strength of SGS (Figure 3, Heuertz *et al.* 2003). Wind-pollinated species show less SGS than animal-pollinated species (Dick *et al.* 2008), selfing species have higher *Sp* values and therefore stronger SGS than

outcrossing species, and herbs typically show stronger SGS than trees (Vekemans & Hardy 2004). A higher effective population density decreases SGS (Vekemans & Hardy 2004) and population fragmentation increases SGS within fragments (de-Lucas *et al.* 2009). Additionally, populations at range margins normally show stronger SGS than populations at the core of the distribution range (Gapare & Aitken 2005; Pandey & Rajora 2012).

SGS studies based on nuclear microsatellites (nuSSRs) have been conducted for many plant species (e.g. Degen *et al.* 2001; Heuertz *et al.* 2003; Hardy *et al.* 2006; de-Lucas *et al.* 2009; Piotti *et al.* 2013). However, the effects of local adaptation at neutral and adaptive loci have rarely been taken into account (but see van Heerwaarden *et al.* 2010; Audigeos *et al.* 2013). Recent studies started to identify environmental factors shaping genetic structure in a landscape context. However, the role of environmental heterogeneity within populations has often been neglected. While local adaptation was often assumed to affect SGS patterns (see e.g. Troupin *et al.* 2006; Volis *et al.* 2010), the driving ecological factors at short spatial scales (i.e. at distances from a few to several hundred meters) have rarely been specifically addressed.

Molecular basis of adaptation in nature

The identification of the genetic basis of local adaptations is one of the main goals in evolutionary biology (Orr & Coyne 1992). Ongoing climate change makes the consolidation of this knowledge increasingly urgent (Savolainen *et al.* 2013). Despite ubiquitous evidence for local adaptation in plants (Leimu & Fischer 2008), its molecular basis remains relatively unknown (Savolainen *et al.* 2013).

The strict definition of local adaptation is that a population must have the highest fitness at its home site compared to other populations introduced to this site (Kawecki & Ebert 2004). However, also the distribution of genetic and phenotypic variation along environmental clines or divergence between populations from contrasting habitats might indicate local adaptation (Conover *et al.* 2009). Local adaptation typically occurs when a phenotype conferring high fitness in one environment is costly in another environment due to recent or ongoing spatially varying selection (Kawecki & Ebert 2004). Other factors such as temporally varying selection, frequent extinctions and recolonizations or adaptive plasticity might hamper local adaptation (reviewed in Savolainen *et al.* 2013).

Selection acts on phenotypes and many adaptive plant traits are polygenic, with each gene having a relatively small effect. Association genetic approaches (i.e. based on correlations

between genotypes and phenotypes) are increasingly used to identify genes underlying adaptive traits. However, apart from the complex inheritance modes of polygenic quantitative traits, several challenges hamper genetic association studies in natural populations. Patterns of genetic variation resulting from selection can be confounded with population genetic structure (Freedman *et al.* 2004) and plasticity can blur genetic and environmental effects. Recently, statistical approaches for association genetics have been developed that take population genetic structure and inter-individual relatedness into account (e.g. Yu *et al.* 2006; Eckert *et al.* 2010). The use of experimental approaches and common garden experiments that rely on growing plants in controlled environments can avoid phenotypic differences due to genotype x environment interactions. However, understanding the genetic basis of local adaptations in natural populations, where phenotypes confer their adaptive value, is of utmost importance.

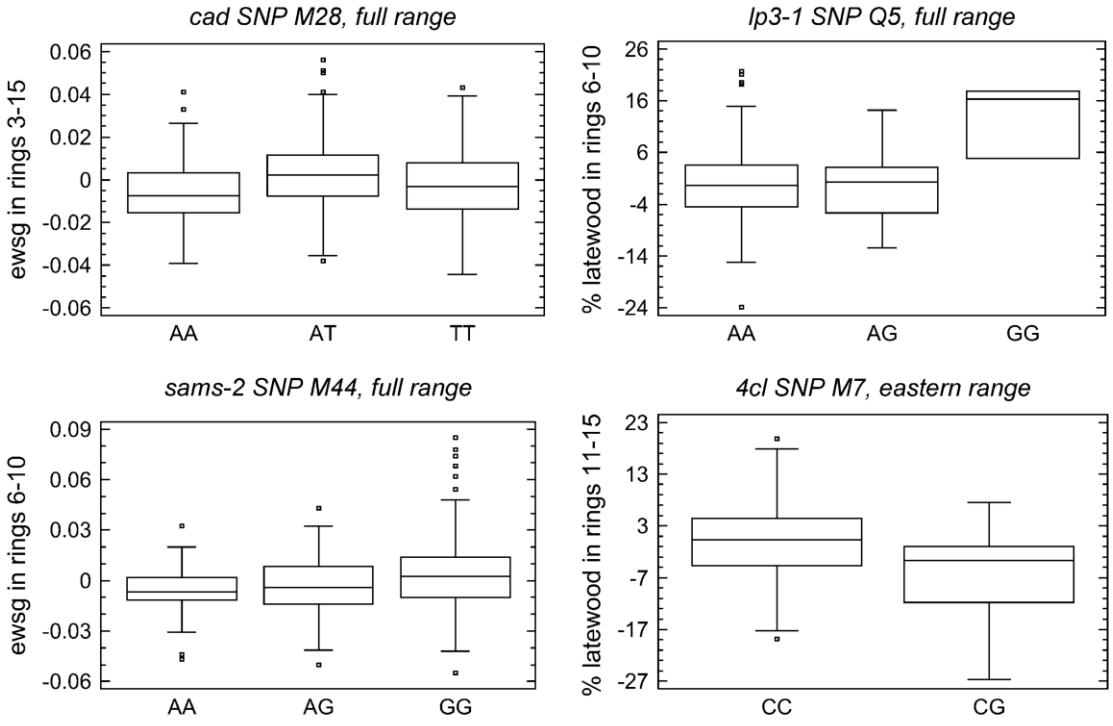

Figure 4 Genotypic effects (box plots) of SNPs that showed significant genetic association (after correction for multiple testing) with earlywood specific gravity (cad SNP M28 andsams-2 SNP M44) and percentage of latewood (lp3-1 SNP Q5 and4clSNP M7 in the east of the Mississippi Valley range) in *Pinus taeda*, after González-Martínez *et al.* (2007).

Most association studies have been conducted in model species or commercial crops with well-known genomes and by taking advantage of inbred lines, e.g. *Arabidopsis thaliana* or *Zea mays*. In Arabidopsis, genome wide association studies (GWAS) identified loci underlying 107 distinct phenotypes related for instance to flowering or plant resistance

(Atwell *et al.* 2010), and they revealed geographic and climatic signatures of local adaptation (Fournier-Level *et al.* 2011). In maize, GWAS identified genes associated, e.g., to plant height (Weng *et al.* 2011), oil biosynthesis (Li *et al.* 2013) or leaf architecture (Tian *et al.* 2011). However, in non-model species, genomic resources are generally limited. As an alternative to GWAS, candidate gene approaches that target loci with possible relevance for the phenotypes under study, are often applied (Figure 4, González-Martínez *et al.* 2007; see also Eckert *et al.* 2009). Conifers have large genomes and a fast decay of linkage disequilibrium (LD), which has been suggested to make genetic association analyses difficult (Neale & Savolainen 2004). However, recent publications showed the feasibility of association studies in this plant group (e.g. González-Martínez *et al.* 2008; Eckert *et al.* 2010a; Holliday *et al.* 2010; Parchman *et al.* 2012). To avoid the confounding effects of environmental variation typically found in natural populations, most of these studies were conducted in field trials or under experimental conditions. However, in some cases, for example when selection pressure is strong, promising candidate genes are known and/or highly heritable plant traits are targeted, association studies can also be conducted successfully in natural populations of conifers (Chapter 5).

Case studies

Symphonia globulifera (Clusiaceae)

Symphonia globulifera L.f. is an ancient and widespread tropical tree species that grows at low to medium-high altitudes (from sea level to 2600 m a.s.l.) in Africa and America (Figure 5). Other species of the genus *Symphonia* are endemic to Madagascar, where *S. globulifera* does not occur (Stevens 2007). *S. globulifera* is a medium-sized late-succession tree (25-40 m height) typical of humid tropical rain forests and common also along rivers and close to mangroves. The species is hermaphroditic and the bright red flowers (Figure 6) are insect- and bird-pollinated, while fruits are dispersed by birds and different mammals, such as bats, ruminants, scatter hoarding rodents and monkeys. Seeds require shade for germination (Oyen 2005). The density of *Symphonia* varies strongly among populations. In the Neotropics, population density of trees with a diameter at breast height (d.b.h.) > 10 cm ranged from 6.12 to 122 trees/ha (Degen *et al.* 2001; van Andel 2003; Hardy *et al.* 2006).

The first pollen fossil records of *Symphonia* (fossil taxon *Pachydermites diederexi*) date back to 45 MA and were found in Nigeria (Jan du Chêne & Salami 1978). There is broad evidence that *S. globulifera* was already widely distributed in Africa before the Pleistocene

(reviewed in Dick *et al.* 2003). The fossil record of this species from the Neotropics is younger (ca. 15 MA, Fournier 1982) and Dick *et al.* (2003) assumed a migration from Africa to the Neotropics via sweepstakes dispersal during the Miocene. A phylogeographic study of *S. globulifera* revealed contrasting demographic histories in lower Mesoamerica and Amazonia (Dick & Heuertz 2008) while its phylogeographic history in Africa remains unknown.

Figure 5 Distribution range of *Symphonia globulifera*, including data from GBIF (circles), Tropicos (triangles) and the collection of CIFOR-INIA/ULB (squares).

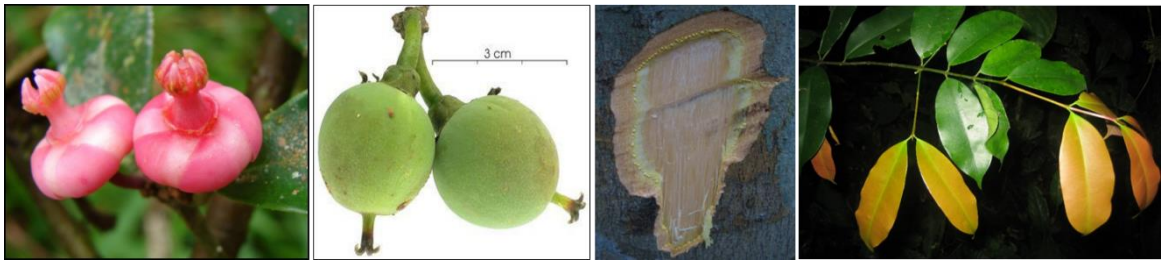

Figure 6 Conspicuous red flowers (left, ©Tobias Sandner, University of Marburg), immature fruits (center left, © Smithsonian Tropical Research Institute), a stem cut revealing yellow latex (center right, © Myriam Heuertz), and opposite leaves with fine parallel nerves and young leaves in yellow brownish color (right, © Myriam Heuertz) of *Symphonia globulifera*.

The species is mainly outcrossing but bi-parental inbreeding (Degen *et al.* 2004) and increased selfing rates in fragmented forest patches have been reported (Aldrich *et al.* 1998). Significant fine-scale spatial genetic structure (SGS) was detected in populations from French Guiana using RAPD and microsatellite markers. The mean historical gene

dispersal estimated based on SGS patterns was 141 m, which was the lowest value in a comparative SGS study of 10 Neotropical tree species (Hardy *et al.* 2006).

Several characteristics of *S. globulifera* made it especially attractive for this PhD thesis. As ancient species with widespread distribution in Atlantic Equatorial Africa, it was particularly suitable for studying Pleistocene forest contractions and the existence of tropical forest refugia. Furthermore, the gradual variation in morphological traits pointed to a single species (Abdul-Salim 2002; Oyen 2005) which avoided studying a cryptic species complex. As *S. globulifera* flowers are animal pollinated and seeds are animal dispersed it is an ecological key stone species providing resources for associated species. The disjunct distribution of *S. globulifera* in the tropical parts of Africa and America allows studying the role of distinct vector species. In Africa sunbirds have been described as pollinators and monkeys, hornbills and ruminants as seed dispersers, while hummingbirds and perching birds are the main pollinators and bats the most important seed dispersers in America. Finally, microsatellite markers developed for *S. globulifera* trees in Neotropical populations (Aldrich *et al.* 1998; Degen *et al.* 2004; Vinson *et al.* 2005) can successfully be transferred to African samples, providing adequate tools for our study.

Pinus halepensis and *Pinus pinaster* (Pinaceae)

Pinus halepensis Mill. (Aleppo pine) and *P. pinaster* Ait. (cluster or maritime pine) are typical Mediterranean conifer tree species that grow under summer dry conditions. These long-lived trees are monoecious, with female and male cones flowering in spring. Cones with seeds reach maturity in autumn two years after pollination (Figure 7). Pollen and seeds are wind-dispersed. Some cones remain closed on the trees and build a canopy seed bank. These serotinous cones are assumed to be an adaptation to wildfire, as high temperatures in the environment trigger cone opening and seed release (Lamont *et al.* 1991).

Both pine species have wide ecological amplitudes. *Pinus halepenis* is distributed throughout the Mediterranean Basin, from easternmost populations in Greece, Turkey and Israel to westernmost ones in Spain and Morocco (Figure 8). *Pinus pinaster* is limited to the western part of the Mediterranean Basin (Figure 9), but grows from sea level to over 2000 m of altitude and in both calcareous and siliceous soils. Genetic diversity is generally higher in *P. pinaster* than in *P. halepensis,* especially in the Iberian Peninsula. This is often explained by the long-term persistence of *P. pinaster* in this region (Salvador *et al.* 2000; Bucci *et al.* 2007), while a more recent recolonization from the Balkan Peninsula to the

west is assumed for *P. halepensis* (Morgante *et al.* 1997; Bucci *et al.* 1998; Gómez *et al.* 2005).

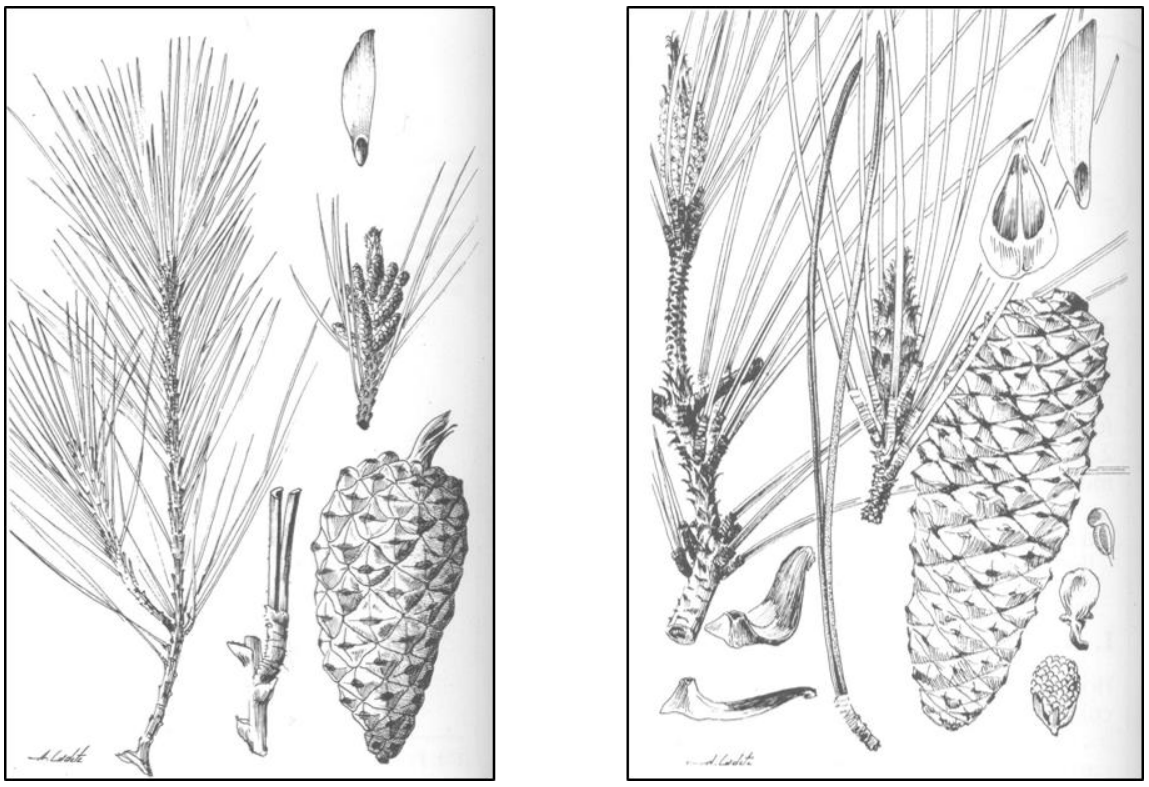

Figure 7 Appearance of needles, cones, and seeds in *Pinus halepensis* (left) and *Pinus pinaster* (right) (© Valdés B., Talavera. S., & Fernández-Galiano E. Flora Vascular de Andalucía Occidental, Ed. Ketres).

Wildfire shapes adaptive traits in Mediterranean pines. For example, some pyrophyte pines show an absence of self-pruning of old cones and dead branches as well as high resin contents, which both enhance flammability (Schwilk & Ackerly 2001). Early flowering (with female function at first flowering) and serotinous cones are also typical fire adaptations in pines (Schwilk & Ackerly 2001; Ne'eman *et al.* 2004). *Pinus halepensis* and *P. pinaster* are highly variable in the expression of fire-adapted traits. *Pinus halepensis* is considered a typical fire evader as adult trees are normally killed by fire events and seeds are released from the numerous serotinous cones (Ne'eman *et al.* 2004; Tapias *et al.* 2004; Fernandes *et al.* 2008). Large seed crops and long-distance dispersal make this species a good colonizer, in particular of abandoned agricultural land. In *P. pinaster*, the fire-response strategy is less clear. Thick barks enable individual trees to survive ground and low intensity fires, a typical feature of fire-resistant species. However, many cluster pine provenances also exhibit serotinous cones which enable regeneration after stand-replacing fires (Tapias *et al.* 2004; Fernandes & Rigolot 2007).

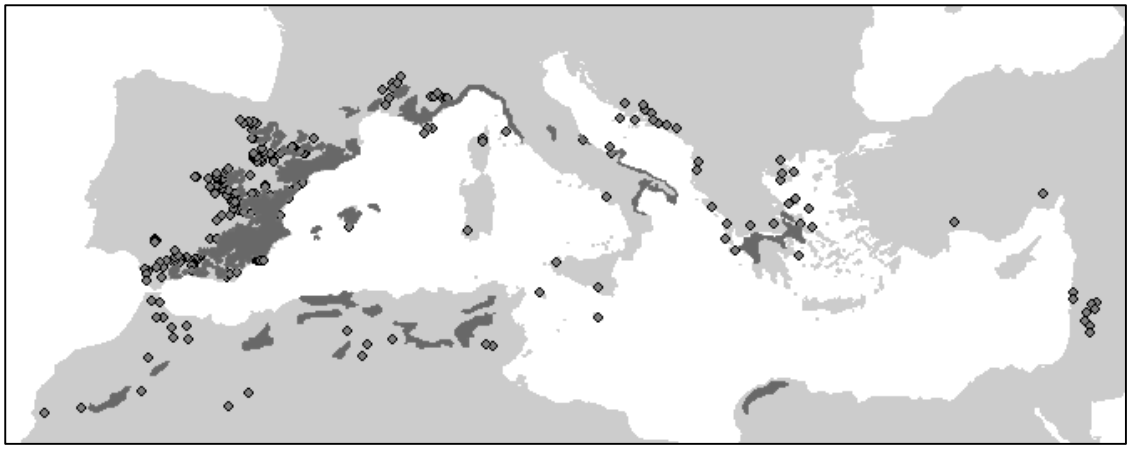

Figure 8 Distribution of *Pinus halepensis*. Grey areas indicate continuous populations while circles mark small fragmented populations (data source: EUFORGEN, European Forest Genetic Resources Programme 2009).

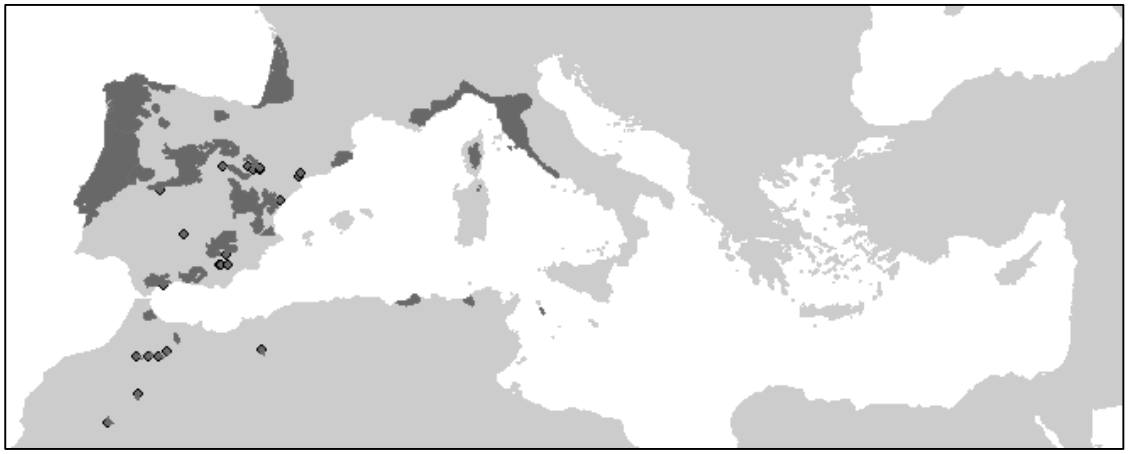

Figure 9 Distribution of *Pinus pinaster*. Grey areas indicate continuous populations while circles mark small fragmented populations (data source: EUFORGEN, European Forest Genetic Resources Programme 2009).

Forests of *Pinus halepensis* and *P. pinaster* from the East of the Iberian Peninsula provide an interesting case study. In this region, Verdú & Pausas (2007) have described areas with contrasting fire regimes (high recurrence vs. low recurrence of fire). Both pine species harbour a single gene pool in the region (Burban & Petit 2003; Gómez *et al.* 2005; Bucci *et al.* 2007). This study setting allowed investigating the effects of recurrent fire events on genetic diversity, demographic history and spatial genetic structure in two species with similar life-history traits while avoiding confounding effects due to population genetic structure. Furthermore, as fire is a strong selection pressure and serotiny is a heritable trait, the study system provided excellent requirements for an association study in natural populations. Finally, previous knowledge on SGS in natural pine populations and the high

environmental heterogeneity in the study populations allowed targeting the effects of environmental factors on genetic structure at fine spatial scale.

Aims of the thesis

The general objective of the present thesis is to improve the understanding of genetic processes shaping the genetic variation in tree species at different spatial scales. With this aim in mind, two contrasted study systems were chosen, (1) *Symphonia globulifera* (Clusiaceae) in Africa and (2) two Mediterranean pine species in eastern Spain, *Pinus pinaster* and *Pinus halepensis* (Pinaceae). All species are representatives of ecosystems in biodiversity hotspots with outstanding conservation value. Based on previous knowledge on the study species, their natural ecosystems and the availability of genetic markers, specific questions of ecological and evolutionary relevance were addressed. For *Symphonia globulifera*, I unveiled its phylogeographic history in Atlantic Equatorial Africa and studied the effects of distinct vector species on the fine scale spatial genetic structure within populations. For the pine species, I studied effects related to fire and other ecological factors on the genetic diversity and fine scale spatial genetic structure at regional and local scale. An overview of the specific objectives of each chapter included in this thesis is given below (see also Table 1):

Section I: Phylogeography

Chapter 1 The ancient tropical rainforest tree *Symphonia globulifera* L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa

Chapter 1 presents a phylogeographic study of *Symphonia globulifera* in Atlantic Equatorial Africa. Based on species endemism and diversity, Maley (1996) proposed a number of glacial refuge areas for tropical rain forest taxa. We used widespread samples of *S. globulifera* to test the hypothesis of persistence in these proposed glacial refuge areas and to describe the species' population genetic structure. The following aims were targeted:

1) To identify the structure of genetic diversity and the geographical features that shaped them. By comparing the patterns of nuclear gene pools and maternal lineages the effect of current geographical features (mountain or ocean barriers and the forest-savannah zone of the Dahomey gap) or historical processes coherent with forest contractions in glacial refugia were investigated.
2) To detect signals of demographic population size changes in current regional gene pools and interpret these signals with respect to forest fragmentation during the LGM.

3) To examine the pollen-to-seed dispersal distance ratio (σ_p/σ_s) using markers with contrasting inheritance, in order to better explain the distribution of nuclear gene pools and maternal lineages.

Section II: Spatial genetic structure

<u>Chapter 2</u> Fine-scale spatial genetic structure within *Symphonia globulifera* populations from different continents as related to dispersal vector species

Symphonia globulifera is an ancient and widespread tropical tree species. In Chapter 2, the fine-scale spatial genetic structure (SGS) within populations from different locations in Africa and South America was compared. One of the main determinants of the strength of SGS is the dispersal ability of plants which is, in this case, related to the presence of distinct vector species acting as pollinators and seed dispersers. The objective of Chapter 2 was:

1) To identify the strength and shape of SGS within populations from Africa and the Neotropics and to estimate the relative contribution of pollen and seed dispersal to overall gene dispersal in different sites and with respect to the different pollen and seed disperser agents.

<u>Chapter 3</u> Effects of fire regime on the population genetics of natural pine stands

In the East of the Iberian Peninsula, previous studies have identified a region with high frequency of stand replacing fires (hereafter HiFi) and another region with more rare fire events (LoFi). *Pinus pinaster* and *P. halepensis* belonging to the same historical gene pool grow naturally in both regions. Stands from the HiFi and LoFi regions were used to study the effect of distinct fire regimes on the population genetics of natural pine stands. Chapter 3 had the following objectives:

1) To test whether HiFi populations show stronger signs of demographic bottlenecks due to the negative impact of frequent fires on seed bank diversity.
2) To investigate if HiFi populations show stronger SGS, in agreement with spatial autocorrelation for serotiny in *P. halepensis* (Hernández-Serrano *et al.* 2013).
3) To test if single nucleotide polymorphism (SNP) markers display a stronger SGS than microsatellites, which would suggest evidence for fire-driven spatially explicit selection.

4) To evaluate if fire effects are stronger in *P. halepensis,* as more fire-dependent species, compared to *P. pinaster*.

<u>Chapter 4</u> Local scale genetics: genetic and phenotypic structure in natural *Pinus pinaster* populations at short spatial scales

Recent studies have started to reveal the environmental and physiographical factors that shape plant genetic structure in a landscape context. However, the effect of micro-environmental variation on the spatial genetic structure within populations has rarely been taken into account so far. In Chapter 4, the correlations between environmental variables (such as elevation, slope, aspect, herbaceous plant cover and herbaceous species richness) and genetic structure were studied in three natural *Pinus pinaster* stands, with the following aims:

1) To disentangle patterns of isolation by distance (IBD) and isolation by adaptation (IBA) at short spatial scales.
2) To identify environmental factors shaping spatial genetic structure within Mediterranean pine populations.
3) To test whether the same environmental factors also determine the distribution of adaptive phenotypes.

Section III: Molecular basis of local adaptation in nature

<u>Chapter 5</u> *In situ* genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*)

Fire is a strong selection pressure that shapes adaptive plant traits. The number of serotinous cones is a heritable plant trait correlated with other fire adaptations. An association study using single nucleotide polymorphisms (SNPs) derived from candidate genes and serotiny phenotypes in natural *Pinus pinaster* stands was conducted. The specific targets of Chapter 5 were:

1) To identify loci potentially associated (or linked to loci associated) to serotiny, a trait that gauges fire phenotypes.
2) To construct a predictive model for serotiny phenotypes based on SNP genotypes potentially associated with serotiny.
3) To evaluate the predictive value of the model at wide spatial scales.

Table 1 Overview of the spatial scales, objectives, material and methods and results in form of publications and manuscripts included in the present thesis

Section	Chapter	Scale	Objectives	Material and methods				Results
				Study sites	Species	Genetic markers	Data analyses	
I. Phylogeography	1	conti-nental	- Genetic patterns and their interpretation in a historic framework and Maley's (1996) forest refuge hypothesis - Demographic history	Atlantic Equatorial Africa	*Symphonia globulifera*	*psbA-trnH* sequences + 5 nSSRs	• Estimation of genetic diversity and population divergence • Identification of gene pools • Haplotype network • Bottleneck analyses • ABC to test divergence and bottleneck scenarios	Budde *et al.* (2013) Heredity
II. Spatial genetic structure	2	local	- Strength and shape of SGS in the presence of distinct pollinator and seed disperser species	Sites in Cameroon, Brazil, French Guiana and Panama	*Symphonia globulifera*	3- 5 nSSRs	• Evaluation of strength and shape of SGS • Estimation of contribution of selfing to overall inbreeding	Budde *et al.* in preparation
	3	regional + local	- Effects of distinct fire regimes on the population genetics of natural pine stands	East of the Iberian Peninsula	*Pinus pinaster* + *Pinus halepensis*	9 and 11 nSSRs + 2 x 251 SNPs	• Estimation of genetic diversity and SGS • SGS simulations • Bottleneck analyses • ABC and skyline plots	Budde *et al.* sub. to Molecular Ecology
	4	local	- Identification of environmental factors shaping the genetic structure and phenotypes within natural populations	East of the Iberian Peninsula	*Pinus pinaster*	251 SNPs	• Mantel test (partial) • Multiple regression of distance matrices • Redundancy analysis • Linear regressions	Budde *et al.* in preparation
III. Molecular basis of local adaptation in nature	5	regional + conti-nental	- Genetic association of serotiny, a fire adapted trait with SNPs in natural stands - Validation of potential asociations at wide spatial scale	East of the Iberian Peninsula Western Mediterranean	*Pinus pinaster*	251 SNPs	• Mixed linear models (single- and multivariate) • Bayesian estimation of allelic effects • Standard correlation Ridge regression (based on BLUPs) to build predictive model	Budde *et al.* (2014) New Phytologist

General discussion

Throughout the chapters of this thesis, patterns of genetic structure were identified at different spatial scales based on distinct types of genetic markers. These patterns in combination with signals of recent or historic population size changes were interpreted with respect to evolutionary processes in interplay with species specific life-history traits and ecology.

Both the phylogeographic study of *S. globulifera* in Atlantic Equatorial Africa (Chapter 1), in comparison with similar studies of tropical tree species in Africa (reviewed in Hardy *et al.* 2013), and the study of fire regime effects on the population genetics of *P. pinaster* and *P. halepensis* (Chapter 3) revealed species specific patterns. For this reason, the response to historic or contemporary ecological impacts can seldom be generalized across co-occurring taxa (see also Kettle *et al.* 2011; Heuertz *et al.* 2014). My research emphasizes the importance of species specific studies, at all spatial scales: continental, regional and local. Additionally, bottleneck signals were detected in populations of all three species, *S. globulifera*, *P. halepensis* and *P. pinaster,* but might be difficult to interpret, as they can reflect processes at other time scales than the one under study (Chapter 1 & 3).

Life-history traits, especially those related to dispersal ability, determine the strength of SGS, as shown by comparing SGS patterns from different plants (Vekemans & Hardy 2004). In Chapter 2, SGS patterns of different populations of *S. globulifera* differed depending on their location in Africa or America. These differences can be interpreted in the light of the presence and efficiency of distinct vector species (see also Hardy *et al.* 2006; Dick *et al.* 2008). Furthermore, SGS patterns in *P. pinaster* and *P. halepensis* indicated distinct responses to fire regimes, although limited SSR marker power made it difficult to tease apart the contributions of neutral processes (such as dispersal and drift) and adaptive processes on SGS (Chapter 3). By correlating environmental factors with genetic distance it was shown that, apart from dispersal processes leading to IBD, also environmental gradients can have a detectable effect on the genetic structure within populations, even in a species with wide-ranging gene flow (Chapter 4, see also van Heerwaarden *et al.* 2010; Audigeos *et al.* 2013).

Chapter 3 pointed to a possible selective effect of fire in natural pine stands, especially in *P. halepensis*, highlighting the role of fire as selective pressure, in agreement with other studies (Pausas *et al.* 2004; Keeley *et al.* 2011; Moreira *et al.* 2014). A high variability of fire adaptive traits, including serotiny, has been described in Eastern Spain with respect to different fire regimes (Hernández-Serrano *et al.* 2013). The combination of high

heritability and phenotypic variability of serotiny enabled us to perform a successful association study in natural pine stands. The correlation of serotiny with other fire adaptive traits (i.e. fire syndrome) possibly facilitated the detection of loci potentially underlying fire phenotypes, pointing to a strong fire adaptive signal throughout the genome (Chapter 5, see also Parchman *et al.* 2012).

Phylogeography

The phylogeographic study of *S. globulifera* presented in Chapter 1 revealed complex phylogeographic patterns based on microsatellites and chloroplast haplotypes, in comparison with similar studies of other African rain forest species (e.g. Dauby *et al.* 2010; Daïnou *et al.* 2010; Duminil *et al.* 2010; Koffi *et al.* 2011; Heuertz *et al.* 2014). Widespread species inhabiting more than one sub-center of endemism of the Guineo-Congolian region (Upper Guinea, Lower Guinea and Congolia) showed genetic divergence between sub-centers, especially between Upper and Lower Guinea, indicating glacial persistence of distinct lineages within each sub-center (reviewed in Hardy *et al.* 2013). The patterns of genetic divergence between gene pools or maternal lineages of *S. globulifera, Milicia excelsa* and *Santiria trimera* were affected by the Dahomey gap and in the case of *Symphonia globulifera* and *Santiria trimera* also by the Bight of Bonny (Chapter 1; Daïnou *et al.* 2010; Koffi *et al.* 2011). With respect to the proposed refuge locations *sensu* Maley (1996), endemic haplotypes and alleles coincided with some of them in e.g. *Santiria trimera* (Koffi *et al.* 2011), *Irvingia gabonensis* (Lowe *et al.* 2010) and *Aucomea klaineana* (Born *et al.* 2011), while in other species, such as in *S. globulifera* (Chapter 1), they were found in many places, inside and outside of proposed refuge areas (reviewed in Hardy *et al.* 2013). These differences can possibly be explained by species specific responses due to different life-history traits (Hardy *et al.* 2013). *Symphonia globulifera*, for example, is a late successional species and occurs frequently along rivers and mangroves (Chapter 1). These habitats may have provided shelter not restricted to forest refuge locations during glaciations (see also Leal 2004).

Within the Lower Guinean sub-center of endemism, a north-south divergence between gene pools became apparent in several species, which coincided approximately with the climate hinge at the meteorological equator (Hardy *et al.* 2013). At the climate hinge located between 0 and 3ºN, forests are exposed to two dry and two rainy seasons of about equal length. North of this latitude the longer dry season is in the boreal summer whereas south of the climate hinge, the longer dry season is in the austral summer. A recent comparative phylogeographic study addressed specific evolutionary scenarios, based on physio-climatic factors, testing whether a north-south or an east-west pattern of

differentiation could be observed across 14 rain forest tree species in Lower Guinea, but detected no congruent general pattern among all species (Heuertz *et al.* 2014). The most frequent scenario, also found in *S. globulifera* analyzing the *trnC-ycf6* chloroplast region, was the north-south divide along the climate hinge (Heuertz *et al.* 2014). However, in *S. globulifera* this result was not supported by the more diverse *psba-trnH* region (Chapter 1). A more in depth study with nuclear genetic markers of the South Cameroon and Gabon gene pool might reveal further genetic structure.

A challenge in phylogeography of tropical species is the timing of demographic or divergence events. Despite this having been addressed in some studies (e.g. Chapter 1; Daïnou *et al.* 2010), methods and accuracy have to be improved in the future (Hardy *et al.* 2013). Although recent advances pointed to the need of more species specific studies, a problem in tropical regions remains the poor taxonomic resolution, causing misleading phylogeographic patterns from species data including several cryptic species (Cavers & Dick 2013; Heuertz *et al.* 2014).

While the phylogeographic history of northern European trees indicated glacial refugia in southern Europe and subsequent recolonization routes to northern regions (e.g. Petit *et al.* 2002, 2003; Heuertz *et al.* 2006), the patterns are more complex in Mediterranean species (Nieto Feliner 2011). Also African tropical tree species show very diverse patterns and generalizations are difficult to draw (Hardy *et al.* 2013; Heuertz *et al.* 2014). Interestingly however, a preliminary study addressing the congruence in cpDNA structure and diversity across 23 woody species in Europe (dataset from Petit *et al.* 2003) and eight tree species in Lower Guinea found higher congruence among species in the African dataset (Dauby *et al.* 2014). This illustrates that identifying a common pattern is not an easy task in tropical regions, where there is no north-south temperature gradient or reliable palaeoecological information (Bonnefille 2007) facilitating the construction of testable hypotheses for vegetation history. In Neotropical tree species, numerous phylogeographic patterns were identified similarly as in Africa and related to evolutionary processes at distinct spatial and temporal scales (e.g. Cavers *et al.* 2003; Dick *et al.* 2013; Cavers & Dick 2013). In fact, Amazonian forest cover is assumed to have been less affected by glacial periods than the forest cover in Atlantic Equatorial Africa (Anhuf *et al.* 2006). In general, these findings reflected the stronger impact of glacial periods in Northern Hemisphere ecosystems. The adverse climatic conditions, with ice sheets covering parts of nothern Eurasia and America (Webb & Bartlein 1992), strongly limited the location of glacial forest refugia in this region. However, in tropical regions, although the climate became also colder and drier (Hamilton & Taylor 1991) and forests most probably became

fragmented, the genetic structure of many species was possibly conditioned more by their life-history traits than by overall climatic patterns.

Spatial genetic structure

The dispersal abilities of plants are among the strongest determinants of their fine-scale spatial genetic structure (Vekemans & Hardy 2004). Consequently, pollen and seed dispersal modes influence the strength of SGS (Dick *et al.* 2008). In *S. globulifera*, an animal pollinated and animal dispersed tree, the composition and behavior of distinct disperser guilds in Africa and America might cause differences in SGS. Several studies described SGS patterns of different tree species in relation to their dispersal agents (e.g. (Degen *et al.* 2001; Hardy *et al.* 2006). However, to my knowledge, the effect of distinct dispersal agents on SGS within different populations of the same plant species have not been addressed. The *Sp* values in *S. globulifera* populations ranged from 0.0086 in Ituberá, Brazil, up to 0.0266 in Mbikiliki, Cameroon. These values are typical for animal pollinated species with gravity or animal dispersed seeds (Vekemans & Hardy 2004), but might indicate slightly stronger SGS in African than in American populations. Gene dispersal was also less leptokurtic in Africa. However, our results contrasted with *a priori* expectations of similar pollen dispersal on both continents, mediated by sunbirds in Africa, and hummingbirds and perching birds in the Neotropics. In fact, both sunbirds and hummingbirds belong to the same functional pollinator group (Armbruster 2006). In contrast, seed dispersal was possibly restricted in Africa, where assumedly far ranging seed dispersal by bats (Dick *et al.* 2008), had never been described for *S. globulifera* (Chapter 2).

Restricted gene flow is doubtlessly important for the build-up of SGS. However, other processes, such as selection or micro-environmental adaptation, can also be relevant. In pine species, pollen and seed flow via wind are relatively far-ranging and previous studies indicated low or non-significant SGS in natural pine populations (Vekemans & Hardy 2004; Troupin *et al.* 2006; de-Lucas *et al.* 2009). In Chapter 3, numerical simulations indicated that the number and characteristics of genetic markers are critical to detect a significant SGS pattern. A higher number of markers is more likely to detect significant SGS due to lower stochasticity (see also Jump & Peñuelas 2007). In our study, a stronger SGS was detected in frequently burnt (i.e. HiFi) vs. rarely burnt *P. halepensis* populations with both microsatellite and SNP markers, while no fire-related differences were found for *P. pinaster*. The differential response of two species with similar life-history traits to this recurrent disturbance probably indicates fine differences in their adaptive traits. Furthermore, our results possibly reflected an additional SGS signal in *P. halepensis*

apparent only in SNPs. This signal might indicate selective processes. In the same *P. halepensis* populations where significant SGS was detected, also a significant autocorrelation of fire-adaptive traits has been detected (Hernández-Serrano *et al.* 2013). These concomitant patterns suggested either functional similarity of neighboring trees in HiFi populations due to micro-environmental selection and/or altered dispersal capacities induced by frequent stand-replacing fires.

Recent studies have pointed out that environmental selection can modify patterns of isolation by distance (IBD) at a regional scale (leading to isolation by adaptation, IBA; Wang 2013; Orsini *et al.* 2013; Sexton *et al.* 2014). At local spatial scales, this same process may also affect SGS within populations (e.g. Garroway *et al.* 2013). No general fire-related effect on SGS was found in *P. pinaster* populations, but some populations showed significant SGS while others did not (Chapter 3). The three studied populations differed substantially in their environmental heterogeneity, suggesting that micro-environmental factors unrelated to fire could be relevant in shaping SGS in *P. pinaster*. Therefore, in Chapter 4, patterns of IBA were studied for this species. For that, we examined the influence of environmental factors on the genetic distance between trees, while controlling for geographic distance (to correct for autocorrelation due to limited gene dispersal). Only the strongest environmental gradient, a 300 m altitudinal cline in Eslida was identified to be significantly related to genetic distance by all methods assayed. On the one hand, selection along this altitudinal cline, leading to divergence between trees from upper and lower altitudes, could explain our results. Possibly, germination and establishment of seeds from upper parts might be reduced in lower altitudes and vice versa due to the exposure to distinct selection pressures. Several ecological factors such as temperature, soil depth, water availability, UV or wind exposure etc. might vary along altitudinal gradients (Körner 2007). Thus, further research to identify the underlying forces of selection would be needed to confirm this explanation. On the other hand, abiotic factors, such as temperature change, can also cause divergence due to a plastic response leading to non-overlapping flowering phenology in upper and lower parts (Gauzere *et al.* 2013). Despite difficulties to identify the underlying ecological factors, Chapter 4 proved that the detection of genetic structure at small spatial scales not simply reflecting IBD is possible, even in a species with wide-ranging gene flow. This result agrees with other recent studies. For instance, Audigeos *et al.* (2013) identified divergence within populations of the tropical tree *Eperua falcata* in relation to seasonally flooded and *terra firme* soils in French Guiana. Furthermore, van Heerwaarden *et al.* (2010) assumed within population differentiation in teosinte to be related to topographical features. These results

indicated that IBD and IBA can, but need not, co-occur at short spatial scales, pointing out the importance of disentangling the underlying processes of SGS patterns.

Molecular basis of adaptation in nature

Wildfire is a strong selection pressure that shapes adaptive plant traits (Keeley *et al.* 2011). Depending on the species, a recurrent fire regime provokes distinct responses, reflected in the distribution of adaptive phenotypes (Hernández-Serrano *et al.* 2013) and in the population genetics of natural pine populations (Chapter 3). Based on this knowledge, in Chapter 5 it was shown that genetic association in natural populations can successfully be performed given some conditions. A high heritability and variability of the trait under study in a region lacking population genetic structure and the feasibility of accurate phenotyping in many individuals are of utmost importance. Trees are most suitable for this kind of studies, as their large and genetically diverse populations often show no genetic structure in large areas (Neale & Savolainen 2004; Petit & Hampe 2006). In Chapter 5, serotiny, a fire adaptive trait that is correlated to other fire adaptive characters (Schwilk & Ackerly 2001), was used as indicator for fire adaptations in *P. pinaster*.

Seventeen SNPs, out of 251 SNPs tested, were potentially associated to serotiny reflecting a strong fire adaptive signal throughout the genome. “Ecological syndrome” traits, including several correlated adaptive characters, such as serotiny, make the identification of specific loci, coding for a single trait, extremely difficult. However, they facilitate a first approximation to identify loci or markers linked to loci underlying adaptive traits associated to the “phenotype syndrome” with relatively low genotyping effort. As ecological syndromes are widespread in nature (i.e. pollination syndromes, Fenster *et al.* 2004) or plant defense syndromes (Agrawal & Fishbein 2006), this approach seems promising to better understand the molecular basis on ecologically-relevant quantitative traits.

A predictive model based on 17 loci could predict 29% of the phenotypic variability in serotiny phenotypes (see Chapter 5). These moderate levels of predictive power are typical of association studies in conifer species (González-Martínez *et al.* 2007; Holliday *et al.* 2010). Parchman *et al.* (2012) used a genotyping by sequencing approach and detected 11 SNPs (out of over 95,000) potentially associated to serotiny (evaluated as a binary phenotype, i.e. trees are categorized as serotinous or non-serotinous) in *Pinus contorta*. Based on these 11 loci, they were able to predict 50% of the phenotypic variability. In the future, through second- and third- generation sequencing approaches, a great number of genetic markers will become available, including in non-model species,

which will facilitate association studies further. Additionally, future research should aim to identify the genomic pathways and physiological function in which possibly associated genes might be involved.

Conclusions

Section I. Phylogeography

1. Geographical barriers such as the Cameroon Volcanic Line, the Bight of Bonny and the Dahomey Gap shaped nuclear gene pools in *Symphonia globulifera*. This genetic structure agreed with phylogeographic patterns of other tropical tree species in Atlantic Equatorial Africa.

2. *Symphonia globulifera* was not restricted to the proposed Pleistocene forest refugia (*sensu* Maley 1996). Instead, this species persisted in more places throughout Lower Guinea, most likely along rivers or close to swamps and mangroves.

3. African tropical rainforest species responded differently to Pleistocene climatic oscillations, probably conditioned by their life-history traits.

Section II Spatial genetic structure

1. SGS is mostly shaped by limited dispersal. Distinct disperser guilds affect the SGS of animal pollinated and dispersed trees. In *Symphonia globulifera* long distance seed dispersal by bats possibly caused weaker SGS in the Neotropics than in Africa, where seed dispersal mediated by bats is lacking for this species.

2. If spatially explicit selection has an effect on SGS in addition to dispersal limitation, SNPs from relevant functional regions of the genome exhibit increased SGS compared to neutral microsatellites. SNPs can also provide more precise SGS analysis because stochasticity decreases with increasing number of markers.

3. Fire regime can induce differences in SGS even in species with similar life- history traits. This was apparent in the increased SGS in *Pinus halepensis* populations under high fire recurrence (HiFi) while no fire related effect was found in *Pinus pinaster*.

4. High fire recurrence does not cause lower genetic diversity or stronger bottlenecks in fire-adapted species such as *Pinus pinaster* and *Pinus halepensis*. This should hold true as long as gene flow connects forests and/or fire-free intervals of sufficient duration assure the recovery of genetically diverse canopy seed banks.

5. Environmental factors, such as steep altitudinal gradients, can cause genetic differentiation among individuals in the presence of gene flow. This was

demonstrated in *Pinus pinaster* trees along a 300 m altitudinal gradient in eastern Spain.

Section III. Molecular basis of local adaptation in nature

1. Genetic association studies are feasible *in situ* if the trait under study i) is highly heritable, ii) shows high variability in a region lacking population genetic structure and iii) can be phenotyped accurately in a high number of individuals.

2. Fire is a strong selective pressure that shapes adaptive plant traits and causes a strong adaptive signal throughout the genome. Syndrome phenotypes comprising several correlated adaptive traits, such as the “fire syndrome”, improve the chances of finding associated marker variation, even with relatively low genotyping effort.

3. Genetic associations valid in one maternal lineage do not necessarily hold true in other maternal lineages. This is possibly due to parallel or lineage specific adaptations.

Conclusiones

Sección I. Filogeografía

1. Las barreras geográficas, como la Línea Volcánica de Camerún, el Golfo de Biafra y el corredor de Togo-Dahomey, contribuyen a separar grupos genéticos nucleares en *Symphonia globulifera*. Esta estructura genética concuerda con algunos patrones filogeográficos de otras especies tropicales arbóreas en el África Ecuatorial Atlántica.

2. *Symphonia globulifera* no se limitó a los refugios forestales del Pleistoceno propuestos (*sensu* Maley 1996). En su lugar, esta especie sobrevivió en lugares adicionales de la Baja Guinea, probablemente a lo largo de ríos o cerca de pantanos y manglares.

3. Las especies de la selva tropical africana respondieron de manera diferente a las oscilaciones climáticas del Pleistoceno posiblemente en función de sus rasgos de historia vital.

Sección II. Estructura genética espacial

1. La estructura genética espacial a escala fina (SGS) está determinada principalmente por la existencia de dispersión limitada. Distintos grupos de polinizadores y dispersores afectan de forma diferente a la SGS. En *Symphonia globulifera* la dispersión de semillas a largas distancias por murciélagos podría haber causado una SGS más débil en los Neotrópicos que en África, donde los murciélagos no dispersan semillas de esta especie.

2. Aparte de los efectos de la restricción de la dispersión, la selección natural podría tener un impacto significativo sobre la SGS que estaría reflejado en marcadores funcionales, como los SNPs, pero no en marcadores neutrales, como los microsatélites. Los SNPs también pueden proporcionar análisis más precisos de la SGS porque la estocasticidad disminuye cuando el número de marcadores aumenta.

3. El régimen de fuego puede provocar diferencias de SGS, incluso en especies con rasgos de historia vital similares. Esto fue evidente en el aumento de la SGS en las poblaciones bajo alta recurrencia de fuegos (HiFi) de *Pinus halepensis*, mientras que no se encontró ningún efecto relacionado con el fuego en *Pinus pinaster*.

4. La recurrencia alta de incendios no causa menor diversidad genética o cuellos de botella demográficos más pronunciados en las especies de pinos adaptadas al fuego, como *Pinus pinaster* y *Pinus halepensis*. Esto debe ser cierto, siempre y cuando el flujo genético conecte los bosques y/o haya intervalos libres de incendios de duración suficiente para asegurar la recuperación de bancos de semillas aéreos que sean genéticamente diversos.

5. Los factores ambientales, tales como los gradientes altitudinales pronunciados, pueden causar la diferenciación genética entre individuos de la misma población en presencia de flujo genético. Esto se demostró en árboles de *Pinus pinaster* a lo largo de un gradiente altitudinal de 300 m en el este de la Península Ibérica.

Sección III. Bases moleculares de la adaptación local en condiciones naturales

1. Los estudios de asociación genética son viables en condiciones naturales (es decir, sin necesidad de establecer ensayos de campo) si el rasgo de estudio i) es altamente heredable, ii) muestra una alta variabilidad en una región que carece de estructura genética poblacional, y iii) se puede fenotipar con precisión en un gran número de individuos.

2. El fuego es una fuerte presión selectiva que forma rasgos adaptativos en las plantas y causa una señal fuerte de adaptación que se refleja en una gran parte del genoma. Los fenotipos resultantes de síndromes, es decir que comprenden varios rasgos adaptativos correlacionados, tales como el "síndrome de fuego", mejoran las probabilidades de encontrar variación genética asociada a nivel molecular, incluso con un esfuerzo de genotipado relativamente pequeño.

3. Las asociaciones genéticas que son válidas en un linaje materno no son necesariamente válidas en otros linajes maternos. Esto se debe posiblemente a adaptaciones paralelas o a adaptaciones específicas de linajes concretos.

References

Abdul-Salim K (2002) Systematics and biology of Symphonia globulifera L.f. (Clusiaceae). PhD thesis, Harvard University, Boston.

Agrawal A, Fishbein M (2006) Plant defense syndromes. *Ecology*, **87**, 132–149.

Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J (2008) Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Molecular Ecology*, **17**, 5177–5188.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-mclane S (2008) Adaptation , migration or extirpation : climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Aldrich PR, Hamrick JL, Chavarriaga P, Kochert G (1998) Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology*, **7**, 933–44.

Van Andel T (2003) Floristic composition and diversity of three swamp forests in northwest Guyana. *Plant Ecology*, **167**, 293–317.

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, **21**, 2078–2091.

Anhuf D, Ledru M-P, Behling H *et al.* (2006) Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **239**, 510–527.

Armbruster W (2006) Evolutionary and ecological perspectives on specialization: from the Arctic to the tropics. In: *Plant-pollinator interactions: from specialization to generalization.* (eds Waser N, Ollerton J), pp. 260–282. University of Chicago Press, Chicago.

Atwell S, Huang YS, Vilhjálmsson BJ *et al.* (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, **465**, 627–631.

Audigeos D, Brousseau L, Traissac S, Scotti-Saintagne C, Scotti I (2013) Molecular divergence in tropical tree populations occupying environmental mosaics. *Journal of Evolutionary biology*, **26**, 529–44.

Avise JC (2000) *Phylogeography: The history and formation of species*. Harvard University Press, Cambridge.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774.

Bergmann F (1978) The allelic distribution at an acid phosphatase locus in Norway spruce (*Picea abies*). *Theoretical and Applied Genetics*, **52**, 57–64.

Bonnefille R (2007) Rainforest responses to past climatic changes in tropical Africa. In: *Tropical rainforest responses to climate change.* (eds Bush MB, Flenley JR, Gosling WD), p. 488. Springer, Berlin/ Heidelberg.

Born C, Alvarez N, McKey D *et al.* (2011) Insights into the biogeographical history of the Lower Guinea Forest Domain: evidence for the role of refugia in the intraspecific differentiation of Aucoumea klaineana. *Molecular Ecology*, **20**, 131–42.

Bucci G, Anzidei M, Madaghiele A *et al.* (1998) Detection of haplotypic variation and natural hybridization in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. *Molecular Ecology*, **7**, 1633–1643.

Bucci G, González-Martínez SC, Le Provost G *et al.* (2007) Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology*, **16**, 2137–2153.

Budde KB, González-Martínez SC, Hardy OJ, Heuertz M (2013) The ancient tropical rainforest tree *Symphonia globulifera* L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa. *Heredity*, 1–11.

Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology*, **12**, 1487–1495.

Cavers S, Dick CW (2013) Phylogeography of Neotropical trees. *Journal of Biogeography*, **40**, 615–617.

Cavers S, Navarro C, Lowe AJ (2003) Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L., in Mesoamerica. *Molecular Ecology*, **12**, 1451–1460.

Conover DO, Duffy TA, Hice LA (2009) The covariance between genetic and environmental influences across ecological gradients: reassessing the evolutionary significance of countergradient and cogradient variation. *Annals of the New York Academy of Sciences*, **1168**, 100–129.

Daily G (1997) *Nature's services: societal dependence on natural ecosystems*. Island, Washington DC.

Daïnou K, Bizoux J-P, Doucet J-L *et al.* (2010) Forest refugia revisited: nSSRs and cpDNA sequences support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (Moraceae). *Molecular Ecology*, **19**, 4462–77.

Dauby G, Duminil J, Heuertz M *et al.* (2014) Congruent phylogeographic patterns of eight tree species in Atlantic Central Africa provide insights on the past dynamics of forest cover. In: *Scripta Botanica Belgica 52, AETFAT XX, Stellenbosch, South Africa* , p. 111. Stellenbosch.

Dauby G, Duminil J, Heuertz M, Hardy OJ (2010) Chloroplast DNA polymorphism and phylogeography of a Central African tree species widespread in mature rainforests: *Greenwayodendron suaveolens* (Annonaceae). *Tropical Plant Biology*, **3**, 4–13.

Degen B, Bandou E, Caron H (2004) Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity*, **93**, 585–91.

Degen B, Caron H, Bandou E *et al.* (2001) Fine-scale spatial genetic structure of eight tropical tree species as analysed by RAPDs. *Heredity*, **87**, 497–507.

de-Lucas A, González-Martínez SC, Vendramin GG, Hidalgo E, Heuertz M (2009) Spatial genetic structure in continuous and fragmented populations of *Pinus pinaster* Aiton. *Molecular Ecology*, **18**, 4564–4576.

Dick CW, Abdul-Salim K, Bermingham E (2003) Diversification of a widespread tropical rain forest tree. *The American Naturalist*, **162**, 691–703.

Dick CW, Hardy OJ, Jones FA, Petit RJ (2008) Spatial Scales of Pollen and Seed-Mediated Gene Flow in Tropical Rain Forest Trees. *Tropical Plant Biology*, **1**, 20–33.

Dick CW, Heuertz M (2008) The complex biogeographic history of a widespread tropical tree species. *Evolution*, **62**, 2760–74.

Dick CW, Lewis SL, Maslin M, Bermingham E (2013) Neogene origins and implied warmth tolerance of Amazon tree species. *Ecology and Evolution*, **3**, 162–169.

Duminil J, Heuertz M, Doucet J-L *et al.* (2010) CpDNA-based species identification and phylogeography: application to African tropical tree species. *Molecular Ecology*, **19**, 5469–83.

Eckert AJ, Bower AD, González-Martínez SC *et al.* (2010a) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology*, **19**, 3789–3805.

Eckert AJ, Bower AD, Wegrzyn JL *et al.* (2009) Association genetics of coastal Douglas fir (*Pseudotsuga menziesii var. menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics*, **182**, 1289–302.

Eckert AJ, van Heerwaarden J, Wegrzyn JL *et al.* (2010b) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, **185**, 969–982.

Epperson BK (2003) *Geographical Genetics*. Princeton University Press, Princeton, NJ.

Etheridge A (2011) *Some mathematical models of population genetics. Ecole d'Ete de Probabilites de Saint-Flour XXXIX-2009 (Lecture Notes in Mathematics*. Springer, Berlin/ Heidelberg.

FAO (2010) *Global Forest Resources Assessment 2010*. Rome.

Fenster C, Armbruster W, Wilson P, Dudash M, Thomson J (2004) Pollination sydromes and floral specialization. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 375–403.

Fernandes PM, Rigolot E (2007) The fire ecology and management of maritime pine (*Pinus pinaster* Ait.). *Forest Ecology and Management*, **241**, 1–13.

Fernandes PM, Vega J, Jimenez E, Rigolot E (2008) Fire resistance of European pines. *Forest Ecology and Management*, **256**, 246–255.

Fournier GR (1982) *Palynostratigraphic analysis of cores from site 493, Deep Sea Drilling Project leg 66. Initial reports of the Deep Sea Drilling Project 66*.

Fournier-Level A, Korte A, Cooper M (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to conservation genetics*. Cambridge University Press, Cambridge.

Freedman ML, Reich D, Penney KL *et al.* (2004) Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, **36**, 388–393.

Gapare WJ, Aitken SN (2005) Strong spatial genetic structure in peripheral but not core populations of Sitka spruce [*Picea sitchensis* (Bong.) Carr.]. *Molecular Ecology*, **14**, 2659–2667.

Garroway CJ, Radersma R, Sepil I *et al.* (2013) Fine-scale genetic structure in a wild bird population: the role of limited dispersal and environmentally based selection as causal factors. *Evolution*, **67**, 3488–3500.

Gauzere J, Klein EK, Oddou-Muratorio S (2013) Ecological determinants of mating system within and between three *Fagus sylvatica* populations along an elevational gradient. *Molecular Ecology*, **22**, 5001–5015.

Gómez A, González-Martínez SC, Collada C, Climent J, Gil L (2003) Complex population genetic structure in the endemic Canary Island pine revealed using chloroplast microsatellite markers. *Theoretical and Applied Genetics*, **107**, 1123–31.

Gómez A, Vendramin GG, Alía R (2005) Genetic diversity and differentiation of two Mediterranean pines (*Pinus halepensis* Mill . and *Pinus pinaster* Ait .) along a latitudinal cline using chloroplast microsatellite markers. *Diversity and Distributions*, 257–263.

González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008) Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity*, **101**, 19–26.

González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics*, **175**, 399–409.

Hamilton A, Taylor D (1991) History of climate and forests in tropical Africa during the last 8 million years. *Climatic Change*, **19**, 65–78.

Hardy O, Born C, Budde KB *et al.* (2013) Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geoscience*, **345**, 284–296.

Hardy OJ, Maggia L, Bandou E *et al.* (2006) Fine-scale genetic structure and gene dispersal inferences in 10 neotropical tree species. *Molecular Ecology*, **15**, 559–71.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution*, **16**, 700–706.

Van Heerwaarden J, Ross-Ibarra J, Doebley JF *et al.* (2010) Fine scale genetic structure in the wild ancestor of maize (*Zea mays ssp. parviglumis*). *Molecular Ecology*, **19**, 1162–1173.

Hernández-Serrano A, Verdú M, González-Martínez SC, Pausas JG (2013) Fire structures pine serotiny at different scales. *American Journal of Botany*, **100**, 1–8.

Heuertz M, Carnevale S, Fineschi S *et al.* (2006) Chloroplast DNA phylogeography of European ashes, Fraxinus sp. (Oleaceae): roles of hybridization and life history traits. *Molecular Ecology*, **15**, 2131–40.

Heuertz M, Duminil J, Dauby G, Savolainen V, Hardy OJ (2014) Comparative phylogeography in rainforest trees from lower Guinea, Africa. *PloS one*, **9**, e84307.

Heuertz M, Fineschi S, Anzidei M *et al.* (2004) Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior L.*) in Europe. *Molecular Ecology*, **13**, 3437–52.

Heuertz M, Vekemans X, Hausman JF, Palada M, Hardy OJ (2003) Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology*, **12**, 2483–2495.

Heywood J (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics*, **22**, 335–355.

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in Plant Science*, **15**, 675–83.

Holliday J a, Ritland K, Aitken SN (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist*, **188**, 501–514.

Holling CS (1973) Resilience and stability of ecological systems. *Annual Review of Ecology, Evolution, and Systematics*, **4**, 1–23.

Jan du Chêne R, Salami M (1978) Palynology and micropaleontology of the Upper Eocene of the well nsukwa-1 (Niger Delta, Nigeria). *Archives des Sciences*, **13**, 5–9.

Jump AS, Peñuelas J (2007) Extensive spatial genetic structure revealed by AFLP but not SSR molecular markers in the wind-pollinated tree, Fagus sylvatica. *Molecular Ecology*, **16**, 925–36.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Keeley JE, Pausas JG, Rundel PW, Bond WJ, Bradstock R a (2011) Fire as an evolutionary pressure shaping plant traits. *Trends in Plant Science*, **16**, 406–411.

Kettle CJ, Hollingsworth PM, Burslem DFRP *et al.* (2010) Determinants of fine-scale spatial genetic structure in three co-occurring rain forest canopy trees in Borneo. *Perspectives in Plant Ecology, Evolution and Systematics*, **13**, 47–56.

Kimura M, Weiss GH (1964) The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics*, **49**, 561–576.

Knowles LL (2009) Statistical Phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 593–612.

Koffi KG, Hardy OJ, Doumenge C, Cruaud C, Heuertz M (2011) Diversity gradients and phylogeographic patterns in *Santiria trimera* (Burseraceae), a widespread African tree typical of mature rainforests. *American Journal of Botany*, **98**, 254–64.

Körner C (2007) The use of “altitude” in ecological research. *Trends in Ecology & Evolution*, **22**, 569–574.

Lamont B, Maitre D Le, Cowling RM, Enright NJ (1991) Canopy seed storage in woody plants. *The Botanical Review*, **5**, 277–317.

Leal ME (2004) The African rain forest during the Last Glacial Maximum an archipelago of forests in a sea of grass. PhD thesis, University of Wageningen, Netherlands.

Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PloS one*, **3**, e4010.

Li H, Peng Z, Yang X *et al.* (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature Genetics*, **45**, 43–50.

Linhart YB (2000) Variation in woody plants: molecular markers, evolutionary processes and conservation biology. In: *Molecular biology of woody plants. Series Forest Sciences 64* (eds Jain SM, Minocha SC), pp. 341–374. Kluwer Academic, Dordrecht.

Lowe AJ, Harris D, Dormontt E, Dawson IK (2010) Testing putative African tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Tropical Plant Biology*, **3**, 50–58.

Malécot G (1968) *The mathematics of heredity*. Freeman, San Francisco.

Maley J (1996) The African rain forest - main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of Edinburgh*.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Maruyama T (1971) Analysis of population structure. II. Two-dimensional stepping stone models of finite length and other geographically structured populations. *Annals of Human Genetics*, **35**, 179–196.

Mayr E, O'Hara RJ (1986) The biogeographic evidence supporting the pleistocene forest refuge hypothesis. *Evolution*, **40**, 55–67.

Médail F, Quézel P (1999) Biodiversity hotspots in the Mediterranean Basin: setting global conservation priorities. *Conservation Biology*, **13**, 1510–1513.

Mittermeier R, Myers N (1998) Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. *Conservation Biology*, **12**, 516–520.

Moreira B, Castellanos MC, Pausas JG (2014) Genetic component of flammability variation in a Mediterranean shrub. *Molecular Ecology*, doi: 10.1111/mec.12665.

Morgante M, Felice N, Vendramin GG (1997) Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck. In: *Molecular Tools for Screening Biodiversity* (eds Karp A, Isaac PG, Ingram DS), pp. 407–412. Springer Netherlands, Dordrecht.

Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, **51**, 238–254.

Myers N (1988) Threatened biotas: "hot spots" in tropical forests. *The Environmentalist*, **8**, 187–208.

Myers N (1990) The biodiversity challenge: expanded hot-spots analysis. *The Environmentalist*, **10**, 243–56.

Myers N, Mittermeier R a, Mittermeier CG, da Fonseca G a, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Ne'eman G, Goubitz S, Nathan R (2004) Reproductive traits of Pinus halepensis in the light of fire – a critical review. *Plant Ecology*, **171**, 69–79.

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science*, **9**, 325–30.

Nieto Feliner G (2011) Southern European glacial refugia : A tale of tales. *Taxon*, **60**, 365–372.

Nosil P, Egan SP, Funk DJ (2007) Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Orr HA, Coyne J (1992) The genetics of adaptation: a reassessment. *The American Naturalist*, **140**, 725–742.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, **22**, 5983–99.

Oyen L (2005) *Symphonia globulifera* L.f. (Internet) Record from Protabase. In: Louppe D, Oteng-Amoako AA, Brink M (eds.) PROTA (Plant Resources of Tropical Africa/ Ressources vegetales de l'Afrique tropicale). *http//database.prota.org/search.htm*, Wageningen, Netherlands.

Palmer JD (1990) Contrasting modes and tempos of genome evolution in land plant organelles. *Trends in Genetics*, **6**, 115–120.

Pandey M, Rajora OP (2012) Higher fine-scale genetic structure in peripheral than in core populations of a long-lived and mixed-mating conifer-eastern white cedar (*Thuja occidentalis* L.). *BMC Evolutionary Biology*, **12**, 48.

Parchman TL, Gompert Z, Mudge J *et al.* (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

Pausas JG, Bradstock RA, Keith DA, Keeley JE (2004) Plant functional traits in relation to fire in crown-fire ecosystems. *Ecology*, **85**, 1085–1100.

Petit J, Aguinagalde I, Beaulieu J De *et al.* (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Petit J, Brewer S, Cheddadi R *et al.* (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.

Petit RJ, Duminil J, Fineschi S *et al.* (2005a) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Petit RJ, Hampe A (2006) Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 187–214.

Petit RJ, Hampe A, Cheddadi R (2005b) Climate changes and tree phylogeography in the Mediterranean. *Taxon*, **54**, 877–885.

Petit RJ, Hu FS, Dick CW (2008) Forests of the past: a window to future changes. *Science*, **320**, 1450–2.

Piotti A, Leonardi S, Heuertz M *et al.* (2013) Within-population genetic structure in beech (*Fagus sylvatica* L.) stands characterized by different disturbance histories: Does forest management simplify population substructure? (PK Ingvarsson, Ed,). *PLoS ONE*, **8**, e73391.

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, **23**, 564–571.

Raven PH (1976) Ethics and attitudes. In: *Conservation of threatened plants* (eds Simmons JB, Beyer RI, Brandham PE, Lucas GL, Parry UTH), pp. 155–179. Plenum Press, New York.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230–237.

Salvador L, Alía R, Agúndez D, Gil L (2000) Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait) in the Iberian peninsula. *Theoretical and Applied Genetics*, **100**, 89–95.

Samis KE, Murren CJ, Bossdorf O *et al.* (2012) Longitudinal trends in climate drive flowering time clines in North American *Arabidopsis thaliana*. *Ecology and Evolution*, **2**, 1162–1180.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.

Schwilk DW, Ackerly DD (2001) Flammability and serotiny as strategies: correlated evolution in pines. *Oikos*, **94**, 326–336.

Sexton JP, Hangartner SB, Hoffmann A a (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, **68**, 1–15.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.

Stevens PF (2007) Clusiaceae-Guttiferae. In: *The families and gernera of vascular plants. Volume IX Flowering plants - Eudicots* (ed Kubitzki K), pp. 48–66. Springer, Berlin.

Storfer A, Murphy M a, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–514.

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

Tapias R, Climent J, Pardos JA, Gil L (2004) Life histories of Mediterranean pines. *Plant Ecology*, **171**, 53–68.

Thomas CD, Cameron A, Green RE *et al.* (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.

Tian F, Bradbury PJ, Brown PJ *et al.* (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, **43**, 159–162.

Troupin D, Nathan R, Vendramin GG (2006) Analysis of spatial genetic structure in an expanding *Pinus halepensis* population reveals development of fine-scale genetic clustering over time. *Molecular Ecology*, **15**, 3617–3630.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Verdú M, Pausas JG (2007) Fire drives phylogenetic clustering in Mediterranean Basin woody plant communities. *Journal of Ecology*, **95**, 1316–1323.

Vinson CC, Amaral AC, Sampaio I, Ciampi AY (2005) Characterization and isolation of DNA microsatellite primers for the tropical tree, Symphonia globulifera Linn. f. *Molecular Ecology Notes*, **5**, 202–204.

Volis S, Zaretsky M, Shulgina I (2010) Fine-scale spatial genetic structure in a predominantly selfing plant: role of seed and pollen dispersal. *Heredity*, **105**, 384–93.

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, **67**, 3403–3411.

Webb T, Bartlein PJ (1992) Global changes during the last 3 million years: Climatic controls and biotic responses. *Annual Review of Ecology and Systematics*, **23**, 141–173.

Weng J, Xie C, Hao Z *et al.* (2011) Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) inbred lines. *PloS one*, **6**, e29229.

White F (1979) The Guineo-Congolian region and its relationships to other phytochoria. *Bulletin du Jardin Botanique national de Belgique*, **49**, 11–55.

White F (1983) *The vegetation of Africa. A descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa.* (Unesco/AETFAT/UNSO, Ed,). Copedith, Paris.

Whittaker RH (1975) *Communities and Ecosystems*. Macmillian, London.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution*, **11**, 413–418.

Yu J, Pressoir G, Briggs WH *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, **38**, 203–208.

SECTION I: Phylogeography

Chapter 1

The ancient tropical rainforest tree *Symphonia globulifera* L.f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa

This chapter reproduces explicitly the following publication:

Budde KB, González-Martínez SC, Hardy OJ, HeuertzM (2013) The ancient tropical rainforest tree *Symphonia globulifera* L. f.(Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa, *Heredity*, **111**, 66-76.

Heredity (2013), 1–11

www.nature.com/hdy

ORIGINAL ARTICLE

The ancient tropical rainforest tree *Symphonia globulifera* L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa

KB Budde[1], SC González-Martínez[1], OJ Hardy[2] and M Heuertz[1,2,3]

Understanding the history of forests and their species' demographic responses to past disturbances is important for predicting impacts of future environmental changes. Tropical rainforests of the Guineo-Congolian region in Central Africa are believed to have survived the Pleistocene glacial periods in a few major refugia, essentially centred on mountainous regions close to the Atlantic Ocean. We tested this hypothesis by investigating the phylogeographic structure of a widespread, ancient rainforest tree species, *Symphonia globulifera* L. f. (Clusiaceae), using plastid DNA sequences (chloroplast DNA [cpDNA], *psbA-trnH* intergenic spacer) and nuclear microsatellites (simple sequence repeats, SSRs). SSRs identified four gene pools located in Benin, West Cameroon, South Cameroon and Gabon, and São Tomé. This structure was also apparent at cpDNA. Approximate Bayesian Computation detected recent bottlenecks approximately dated to the last glacial maximum in Benin, West Cameroon and São Tomé, and an older bottleneck in South Cameroon and Gabon, suggesting a genetic effect of Pleistocene cycles of forest contraction. CpDNA haplotype distribution indicated wide-ranging long-term persistence of *S. globulifera* both inside and outside of postulated forest refugia. Pollen flow was four times greater than that of seed in South Cameroon and Gabon, which probably enabled rapid population recovery after bottlenecks. Furthermore, our study suggested ecotypic differentiation—coastal or swamp vs *terra firme*—in *S. globulifera*. Comparison with other tree phylogeographic studies in Central Africa highlighted the relevance of species-specific responses to environmental change in forest trees.
Heredity advance online publication, 10 April 2013; doi:10.1038/hdy.2013.21

Keywords: forest refuge theory; phylogeography; *psbA-trnH*; SSR; tropical Africa; approximate bayesian computation

INTRODUCTION

Forest refugia denote regions where forest taxa are believed to have persisted and evolved through adverse climatic conditions. The general premise of the forest refuge theory, which was developed with reference to the Pleistocene climatic oscillations, is that climate change can lead to isolation of conspecific populations, which will then become genetically differentiated and eventually speciate (Haffer, 1969; Prance, 1982). Forest refuge theory became a popular model for explaining biological diversification in the tropics in the 1980s when palaeoecological and geomorphological data challenged the hypothesis of long-term environmental stability and a resulting slow accumulation of species over time (Prance, 1982; Mayr and O'Hara, 1986). The identification of forest refugia is a promising approach to the conservation of biodiversity under climate change, as refugia represent locations that have avoided regional climate extremes and large-scale disturbance and therefore should have accumulated substantial biodiversity (Millar *et al.*, 2007). Also, genetic approaches to identifying refugia can reveal evolutionary processes that are highly relevant for predicting plant species' response to future climate change (McMahon *et al.*, 2011).

In temperate regions, refuge locations and postglacial migration routes were suggested for many woody taxa based on palaeoecological data and further confirmed by geographical patterns of genetic diversity at neutral markers (for example, Petit *et al.*, 2003; Heuertz *et al.*, 2006). These studies indicated latitudinal and altitudinal range shifts with, typically, glacial persistence in southern refugia followed by northward postglacial recolonization for northern hemisphere thermophilous species (Petit *et al.*, 2003). In tropical areas, refuge theory has been rejected in the Amazonian basin but remains an active hypothesis in tropical Africa and Australia (Ghazoul and Sheil, 2010). Climate-induced altitudinal, but not latitudinal, shifts and postglacial population expansion have recently been suggested in Africa (Born *et al.*, 2011).

The Guineo-Congolian phytogeographic region represents the largest regional centre of endemism for the African rainforest, being subdivided into three subcentres: Upper Guinea (West Africa), Lower Guinea and Congolia (White, 1979; Figure 1). Refuge theory in this region is supported by palaeoecological data, which indicates that the forest was fragmented during cool and dry glacial periods of the Pleistocene and that it expanded into grasslands under warmer and wetter conditions (Maley, 1996; Bonnefille, 2007). As refugia are thought to have led to speciation in isolation, refuge locations embedded in the current rainforest distribution have been proposed based on high species richness or endemism, especially in recently radiated or poorly dispersing groups (Sosef, 1994; Droissart, 2009). A review of palaeoecological and plant and animal species distribution

[1]INIA, Forest Research Centre, Department of Forest Ecology and Genetics, Carretera A Coruña km 7.5, Madrid, Spain; [2]Université Libre de Bruxelles, Faculté des Sciences, Evolutionary Biology and Ecology—Brussels, Belgium and [3]Real Jardín Botánico, CSIC, Plaza de Murillo 2, Madrid, Spain
Correspondence: KB Budde or Dr M Heuertz, INIA, Forest Research Centre, Department of Forest Ecology and Genetics, Carretera A Coruña km 7.5, 28040 Madrid, Spain.
E-mail: katha.budde@gmail.com (KBB) or heuertzm@gmail.com (MH)
Received 12 April 2012; revised 1 February 2013; accepted 25 February 2013

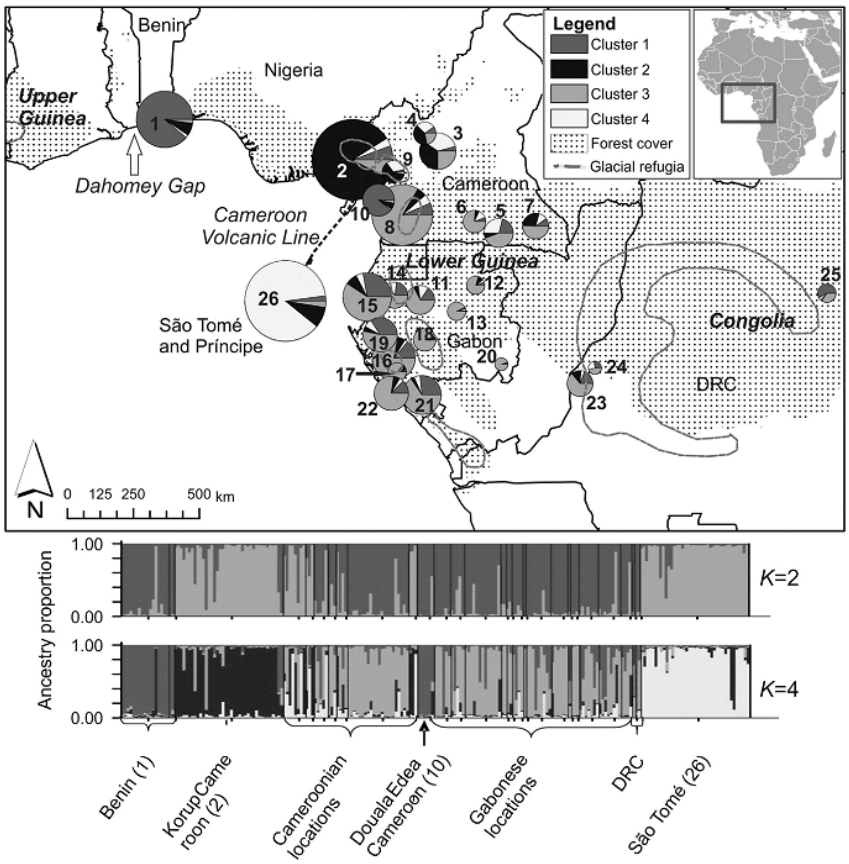

Figure 1 Above: Map of the geographical distribution of GPs identified by STRUCTURE ($K=4$ clusters) for *S. globulifera* in Atlantic Equatorial Africa and postulated refuge locations (areas limited with grey lines) during the last glacial maximum (LGM, about 18 000 years BP) adapted from Maley (1996). Numbers correspond to sampling locations in Table 1. Chart size increases with sample size (ranging from 1 to 39 individuals). Regional names are given in italics. Below: Bar plots showing STRUCTURE ancestry proportions in $K=2$ and $K=4$ clusters. Each individual is represented as a line segment, which is vertically partitioned into K coloured components representing the individual's estimated proportions of ancestry in the K clusters. DRC, Democratic Republic of the Congo.

data suggested three upland refugia in Upper Guinea, six in Lower Guinea (one on the Cameroon Volcanic Line (CVL) in West Cameroon, one in Southwest Cameroon, three in Gabon and one in the Mayombe massif in the Republic of Congo, Democratic Republic of Congo and Angola) and one large lowland refuge in the Congo Basin (Maley, 1996; Figure 1). Limitations for identifying refugia based on palaeoecological and/or species distribution data in Central Africa are threefold: (1) the palaeoecological record is still scarce and pollen profiles typically show temporal discontinuities and poor taxonomic resolution (for example, see Bonnefille, 2007). (2) Species richness and endemism may be poor proxies for refugia because the high habitat heterogeneity of upland areas favours high biodiversity and can trigger speciation due to diversifying selection without requiring historical isolation (for example, see Moritz *et al*., 2000). (3) Species distribution data in many plant and animal groups are inappropriate for indicating refugia because sister taxa diverged before the Pleistocene (Moritz *et al*., 2000; Plana, 2004).

Phylogeographic studies should allow for a broader test of the refuge theory in Africa and have indeed begun providing insights into responses of tree species to past climatic oscillations. For example, the two Cameroonian refuges proposed by Maley (1996) are also supported by the distribution of endemic alleles in *Irvingia gabonensis* (Lowe *et al*., 2010) and *Santiria trimera* (Koffi *et al*., 2011). In *Aucoumea klaineana*, genetic diversity is structured in four differentiated gene pools (GPs; identified using nuclear microsatellites), the location of which coincides with refugia postulated from species distribution data in Gabon (Born *et al*., 2011). However, genetic patterns in three pioneer (*Distemonanthus benthamianus, Erythrophleum suaveolens* and *Milicia excelsa*) and one non-pioneer (*Greenwayodendron suaveolens*) species were not in agreement with postulated refuges but showed a common north–south divide in Lower Guinea, which was suggested to reflect the seasonal inversion near the equator (Daïnou *et al*., 2010; Dauby *et al*., 2010; Debout *et al*., 2011; Duminil *et al*., 2010). These emerging trends allow formulation of hypotheses on biogeographic scenarios and evolutionary processes operating within species that can be evaluated with multilocus genetic data (preferably using loci with contrasting inheritance) using statistical phylogeographic approaches (Knowles, 2009).

In the present study, we chose an ancient widespread tropical tree species, *S. globulifera* L. f. (Clusiaceae; originated ca 45–30 Ma ago, see below), typical of mixed humid and freshwater swamp forests in Africa and America, to test the Pleistocene refuge theory and assess population demographic processes in Central Africa. The gradual

nature of morphological variation in *S. globulifera* has prevented its subdivision (Abdul-Salim, 2002; Oyen, 2005) and the species represents an interesting contrast to recently radiated, species-rich families, such as Begoniaceae, Orchidaceae or Rubiaceae, which are typically used to support the refuge theory in Africa (Sosef, 1994; Droissart, 2009). Furthermore, the genus *Symphonia* has a distinctive pollen morphology providing solid evidence that it was widespread in tropical Africa long before the Pleistocene (reviewed in Dick *et al.*, 2003). Therefore, if this species harboured localized endemic alleles and distinct GPs coinciding with proposed refuges, this would corroborate the refuge theory in Africa and give further support for the relevance of this model for biodiversity organization at the within-species level.

Specifically, we searched for signals of past demographic events in the spatial genetic constitution of *S. globulifera* in the Guineo-Congolian region using plastid DNA sequences, which essentially inform on colonization processes, and nuclear microsatellites, which provide more robust multi-locus information on the distribution of diversity. We included samples from a comprehensive set of locations through the region with a strong emphasis on Atlantic Equatorial Africa (Lower Guinea), addressing the following questions:

(1) Is the genetic diversity geographically structured in the study region? If so, are cpDNA haplotypes and SSR gene pools structured according to geographic features (mountain or ocean barriers and the forest-savannah zone of the Dahomey gap, see below) or can their distribution be explained in terms of origin from distinct forest refugia *sensu* Maley (1996)?
(2) Do current GPs or regions harbour signs of a recent demographic bottleneck, such as would be expected if they were still affected by last glacial maximum (LGM) forest fragmentation, or of population expansion, such as would be expected if rapid colonization occurred from refugia? What is the order of divergence of the distinct GPs?
(3) What is the pollen-to-seed dispersal distance ratio (σ_p/σ_s) estimated from markers with contrasting inheritance and how does this quantity help to explain the colonization history of the species?

MATERIALS AND METHODS

Study species

S. globulifera L.f. (synonym *S. gabonensis* [Vesque] Pierre) is a medium-sized tree (25 – 40 m tall) occurring in Africa from Guinea Bissau to Tanzania and in America from Mexico to Brazil. It has a wide ecological amplitude and grows in forests from sea level to 2600 m altitude (in East Africa) with an annual rainfall of 650–2100 mm and a mean annual temperature of 23–27 °C (Oyen, 2005). It is a late successional species in evergreen mixed humid forests where it regenerates by seeds, requiring shade for germination (Oyen, 2005). The species can grow on a variety of soils (Lescure and Boulet, 1985) and can be locally common along rivers, in swamp forests and on the inner edges of mangroves, where regeneration occurs mostly by root suckers (Van Andel, 2003; Oyen, 2005). The oldest fossil pollen records of *Symphonia* (fossil taxon *Pachydermites diederexi*) were found in Nigeria and date to the mid-Eocene (ca 45 Ma, Jan du Chêne and Salami, 1978) and there are records through the Oligocene and Miocene from Angola (Dick *et al.*, 2003). American fossils of the taxon are younger, the oldest dating to the early/mid-Miocene (ca 18–15 Ma, reviewed in Dick *et al.*, 2003). *S. globulifera* has 16–23 congeners, all endemic to Madagascar where *S. globulifera* does not occur (Abdul-Salim, 2002). The distribution, palaeoecological and genetic data favour Africa or Madagascar as the geographic origin of the genus (Dick *et al.*, 2003; M Heuertz, OJ Hardy and CW Dick, unpublished data).

S. globulifera has showy red flowers and is pollinated by butterflies, hummingbirds and perching birds in the Neotropics (Bittrich and Amaral, 1996) and by various insects and sun birds in Africa (Oyen, 2005). Seeds are dispersed by bats, tapirs, rodents, primates and deer in the Neotropics and by hornbills, primates and duikers in Africa (Forget *et al.*, 2007). *S. globulifera* is a predominantly outcrossing species but in some populations from the Neotropics significant levels of biparental inbreeding have been detected (Aldrich and Hamrick, 1998; Degen *et al.*, 2004; da Silva Carneiro *et al.*, 2007). Bird-mediated pollen dispersal has been shown to be limited in the Neotropics (mean 27–54 m, Degen *et al.*, 2004; 154–444 m, da Silva Carneiro *et al.*, 2007). No information on pollen (or seed) dispersal distances are available for the African range.

Plant material and DNA extraction

Leaf or cambium samples ($n = 251$) were collected from *S. globulifera* individuals in coastal forests in Benin (forest-savannah mosaic of the Dahomey gap, between Upper and Lower Guinea, 19 samples), and almost exclusively in *terra firme* forests of Cameroon and Gabon (Lower Guinea, 182 samples), the Democratic Republic of Congo (Congolia, 9 samples) and from the Island of São Tomé in the Gulf of Guinea (41 samples, Table 1, Figure 1). The sampling was conducted between 2006 and 2010 jointly with several other species as part of a series of field missions led by the Université Libre de Bruxelles. The forest of each locality or village was prospected for 1 or 2 days, which, in combination with a variable density of *S. globulifera* (from presumed absence of the species to ca 6 individuals per ha, with diameter at breast height >20 cm), resulted in variable sample sizes per locality. The plant material was immediately dried with silica gel. DNA was extracted using the Invisorb DNA Plant HTS 96 Kit (Invitek, Berlin, Germany).

Plastid DNA sequences

A subsample of 94 individuals from 18 sampling locations (Figure 2) was sequenced at the *psbA-trnH* plastid DNA (cpDNA) intergenic spacer using the psbAF and trnHR primers (Sang *et al.*, 1997). The total PCR volume of 25 µl contained approximately 20 ng of template DNA, 1 × HF PCR reaction buffer, 0.1 µM of each primer, 200 µM of each deoxyribonucleotide triphosphate (dNTP) and 0.25 U of Phusion polymerase (Finnzymes, Espoo, Finland). The PCR profile was 30 s at 98 °C, 35 cycles of 5 s at 98 °C, 30 s at 50 °C and 45 s at 72 °C, and a final elongation of 3 min at 72 °C. PCR products were purified on filter columns (QIAquick96 kit, Qiagen, Hilden, Germany or MSB HTS PCRapace/C[96]kit, Invitek) or using the enzymes Exonuclease I and Shrimp Alkaline Phosphatase (GE Healthcare, Waukesha, WI, USA) and quantified on agarose gels (1%). Sequencing reactions were performed in both directions using BigDye v.3.1 chemistry (Applied Biosystems, Lennik, Belgium), purified with an ethanol-sodium acetate protocol and analysed on an ABI3100 or an ABI3730 sequencer (Applied Biosystems). Sequences were edited and aligned in CodonCode Aligner 3.0.1. (CodonCode Corporation, Dedham, MA, USA) including base-calling with Phred (Ewing *et al.*, 1998) and alignment with Phrap (Green, 2009). A nucleotide site was considered polymorphic if different variants had at least a Phred quality value of 25, corresponding to an error probability of 3/1000.

Microsatellite markers

Five nuclear microsatellites (SSRs) developed for neotropical *S. globulifera* were optimized for our African samples: SG03 and SG18 (Degen *et al.*, 2004), SGC4 and SG19 (Aldrich *et al.*, 1998), and SG10 (Vinson *et al.*, 2005). Forward primers were fluorescently labelled with IRDye700 or IRDye800 (see Supplementary Information 1). The total PCR volume of 10 µl contained approximately 25 ng of template DNA, 1 × PCR Reaction Buffer (Bioline, London, UK), locus-specific concentrations of $MgCl_2$ (Supplementary Table 1), 0.2 mM of each primer, 0.2 mM of each dNTP and 1 U of Taq polymerase (Ecotaq, Bioline). The PCR conditions were 3 min at 94 °C, a locus-specific number of cycles for denaturation (30 s at 94 °C), annealing (30 s at locus-specific temperatures) and elongation (45 s at 72 °C), and a final elongation for 7 min at 72 °C (Supplementary Information 1).

PCR products were diluted with a formamide-loading buffer and separated on acrylamide gels on a 4300 DNA analyzer (Li-Cor Biosciences, Lincoln, NE, USA). Microsatellite allele sizes were determined using the genotyping automation software SAGA^{GT} (Li-Cor Biosciences) in comparison with the

Table 1 Plant material analysed in *Symphonia globulifera*

No.	*Location*	*Latitude*	*Longitude*	*Localities included*	n	n_{cp}	d_{mean} *(km)*	d_{max} *(km)*
1	Benin	6.42585° N	2.47969° E	Niaouli, Porto Novo, Ahozon, Kraké	19	13	29.10	72.44
2	*Cameroon Korup*	5.05685° N	8.83818° E	Korup National Park and 50 ha plot of the Smithsonian Tropical Research Institute	39	14	2.44	6.33
3	Cameroon CN	5.34846° N	11.72144° E	Yoko, Mégan, Nyafianga, Djendjou cliffs	8		40.49	116.43
4	Cameroon N	5.95185° N	11.28245° E	Mandah, Monkoin	3		30.73	42.46
5	Cameroon S1	2.57775° N	13.78582° E	Lélé, Mbalam, Mindourou	5	5	70.42	103.69
6	Cameroon S2	2.96657° N	12.965° E	Dja Biosphere Reserve camp Mamma, Mbouma	3	3	51.44	77.04
7	Cameroon SE	2.80969° N	15.03372° E	Ngato, Moloundou, management unit UFA10	4	3	46.23	90.36
8	*Cameroon SW, south-west;*	3.19157° N	10.52427° E	Akom, Bipindi, Ngovayang massif, management unit UFA 09-028 (EFFA),	22	13	23.53	80.72
9	*Cameroon Volcanic Line*	4.64713° N	10.22126° E	Mount Koupé, Yingui region	3	1	63.71	93.15
10	Cameroon W coast	3.69213° N	9.71888° E	RF Douala-Edéa, Limbe	6	1	29.58	86.23
11	Gabon CNE	0.31051° N	11.1281° E	Mékié, Ndjolé North	5		31.54	77.07
12	Gabon NE1	0.81011° N	13.00449° E	Makokou, Bélinga	2		82.55	82.55
13	Gabon NE2	0.0697° S	12.36367° E	Ivindo	2		0.36	0.36
14	*Gabon Crystal*	0.4726° N	10.27115° E	Crystal Mountains	4	4	0.63	1.00
15	Gabon coast	0.41459° N	9.33374° E	Cap Estérias, Pointe Denis, Pongara	14	14	18.07	46.16
16	Gabon CW	1.66178° S	10.23387° E	Niambo-kamba, Mandji W	11		32.91	77.10
17	Gabon CS	2.12277° S	10.35227° E	Moukalaba-Doudou National Park	2		78.37	78.37
18	*Gabon Waka*	1.03788° S	11.27908° E	Waka, Lopé Reserve	3	2	39.93	59.90
19	Gabon W	0.86243° S	9.77248° E	Petite Silang, Lake Avanga, Lake Anengue, Enyonga	7		52.97	95.51
20	Gabon SE	1.86288° S	13.87708° E	Ossélé	1	1	NC	NC
21	Gabon SW	2.90301° S	11.18371° E	Douano, Eastern slope of the Mayombe massive, Bikamba	9	2	36.23	72.18
22	*Gabon Gamba-Doudou*	2.84693° S	10.1492° E	Gamba, Luango	7	6	52.46	149.40
23	DRC W1	2.51542° S	16.53403° E	MbouMon, Mbanzi	4	4	12.78	24.57
24	DRC W2	1.98474° S	17.0428° E	Boleke	1	1	NC	NC
25	DRC N	0.53791° N	24.88835° E	Bamboussoko, Yangambi	2		106.84	106.84
26	São Tomé	0.27857° N	6.56685° E	São Tomé	39	5	2.81	8.38
	Overall sample						586.74	2672.05

Abbreviations: CN, centre-north; CNE, centre-north-east; CS, centre-south; CW, centre-west. DRC, Democratic Republic of the Congo; d_{mean} and d_{max}, mean and maximum pairwise distance between individuals sampled; *n*, number of individuals sampled and genotyped at nSSR markers; n_{cp}, number of individuals sequenced at psbA-trnH; NC, not calculated; SE, south-east; SW, south-west.
Location, sampling location (in italics and underlined for locations coinciding with refuges *sensu* Maley (1996); latitude and longitude, central geographic coordinates of sampling locations in decimal degrees. Numbers correspond to locations in Figure 1.

SequaMark DNA size marker (Invitrogen, San Diego, CA, USA). Length variants that could not unequivocally be distinguished were pooled into the same allele-class.

Plastid DNA data analysis

CpDNA haplotypes were defined taking into account only point mutations (single-nucleotide polymorphisms, SNPs) omitting insertions and deletions (indels), because they contained complex repetitive regions, which were difficult to code and added noise to the haplotype analysis. A haplotype network was constructed using a maximum parsimony method implemented in the median-joining algorithm of the software NETWORK version 4.600 (Fluxus Technology Ltd, Suffolk, UK; Bandelt *et al*., 1999). The geographical distribution of haplotypes was illustrated in a GIS (Geographic information system) map built in ArcMap9.3.1 (ArcGIS 9, ESRI, Redlands, CA, USA).

The genetic diversity of all samples was estimated as the number of haplotypes and as haplotypic diversity based on unordered (h) or ordered (v) haplotypes (Pons and Petit, 1996). For the latter, a genetic distance matrix was computed among haplotypes, defining the distance between two haplotypes as the number of polymorphic sites differing between them. Differentiation among locations and between pairs of locations with at least three individuals sampled was assessed with SPAGeDi1.2g (Hardy and Vekemans, 2002) using the corresponding statistics, G_{ST} for unordered and N_{ST} for ordered alleles. Permutation (10 000) of individuals among populations was used to test for deviation of G_{ST} or N_{ST} from zero. The impact of mutations on among-population differentiation (test for $N_{ST} > N_{ST[permuted]}$), that is, the presence of phylogeographic structure, was assessed by 10 000 permutations of pairwise haplotype distances among pairs of haplotypes.

Microsatellite data analysis

To identify the GP composition of the sample, we used the Bayesian clustering method implemented in STRUCTURE 2.2 (Pritchard *et al*., 2000). We ran an admixture model with correlated allele frequencies between clusters. Ten runs were performed for each number of clusters $K=1$ to $K=10$ with a burn-in length of 100 000 and a run length of 200 000 iterations. The K that best described the data was determined from the graph of posterior log likelihood of the data, LnP(D|K), against K, and from delta K (Evanno *et al*., 2005). For graphical representation of the clustering results, samples were grouped into 26 locations based on geographic proximity (Table 1, Figure 1) and a GIS map was built in ArcMap9.3.1. Clustering results were verified with TESS 2.3.1 (Chen *et al*., 2007), which infers ancestry proportions of individuals in K user-defined clusters, similar to STRUCTURE, and that includes the sampling location of each individual as prior information. We ran TESS using a conditional auto-regressive admixture model with burn-in of 10 000 and run length of 50 000 iterations, a spatial interaction parameter of 0.99 and a trend degree surface of 1. Ten repetitions were carried out for each $K=1$ to $K=8$.

For analyses at the regional level, regions were defined based on the geographic location of GPs. Overall and pairwise genetic differentiation among regions were assessed computing F_{ST} (Weir and Cockerham, 1984) using FSTAT 2.9.3.2 (Goudet, 1995) and R_{ST}, an analogue of F_{ST} accounting for allele size (Slatkin, 1995), using SPAGeDi1.2g. F_{ST} and R_{ST} were tested against the

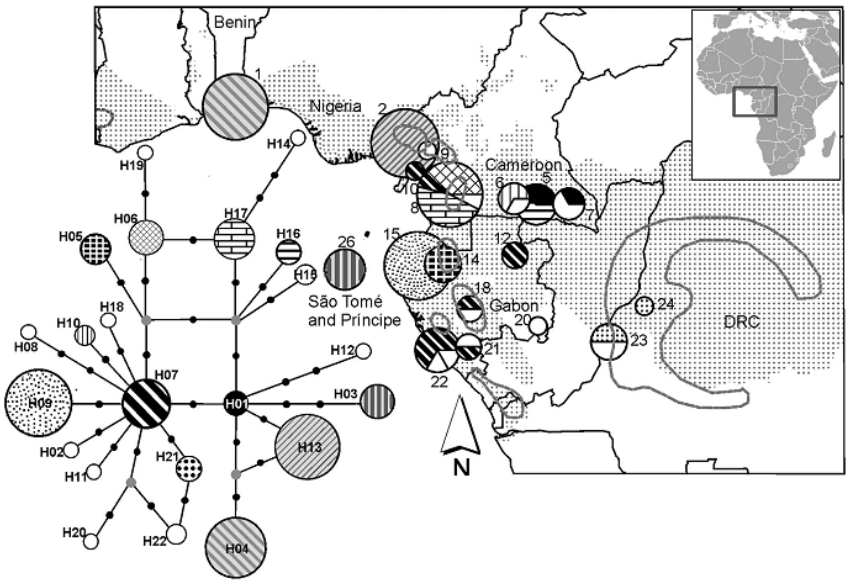

Figure 2 Distribution of *psbA-trnH* haplotypes of *S. globulifera*. Chart size increases with sample size (ranging from 1 to 14 individuals). Grey lines delimit postulated refuge locations during the LGM (about 18 000 years BP) adapted from Maley (1996). A grey background colour indicates haplotypes found in Benin, Cameroon Korup and in São Tomé, to facilitate comparison with the map of SSR GPs (Figure 1). Numbers correspond to sampling locations in Tables 1 and 2. Haplotypes detected in one or two samples in a single population only are marked in white. Lower left: Haplotype network with haplotype numbers in bold (small black circles indicate mutations; small grey circles indicate putative haplotypes not observed in this study).

null expectation of absence of population structure with permutation tests. A phylogeographic signal was tested for by comparing R_{ST} to its expectation ($R_{ST[permuted]}$) obtained from 10 000 permutations of allele sizes among alleles using SPAGeDi1.2g. The test is significant if stepwise-like mutations have contributed to differentiation, for example, in the case of ancient isolation (Hardy *et al.*, 2003). For each GP, the expected heterozygosity or gene diversity (H_E) and the allelic richness (R_S) were computed in FSTAT.

Demographic history analysis

The demographic history of each region was first explored computing the 'bottleneck' statistic T_2 with BOTTLENECK 1.2.02 (Cornuet and Luikart, 1996). T_2 represents the deviation of the sample gene diversity from the gene diversity expected for the number of alleles and is positive (heterozygosity excess) in the case of a recent bottleneck and negative in the case of recent population expansion (see Supplementary Information 2). As Approximate Bayesian Computation (ABC) methods have recently been shown to have increased power for bottleneck detection (Peery *et al.*, 2012), we used the ABC framework in DIYABC v1.0.4.46beta (Cornuet *et al.*, 2008; available at http://www1.montpellier.inra.fr/CBGP/diyabc) to compare four demographic scenarios in each region. Scenarios were defined based on the relative past pollen abundance of forest vs savannah species in Central Africa (for example, see Bonnefille, 2007; Ngomanda *et al.*, 2009; Dupont, 2011): (1) null model of constant population size, (2) a recent bottleneck coinciding approximately with the 1000–3000 BP rainforest disturbance, (3) a bottleneck coinciding approximately with the LGM, ca 15 000–22 000 BP and (4) an older bottleneck coinciding approximately with the previous glacial period (ca 130 000–190 000 BP). Initial runs were carried out to explore the parameter space. Parameter values were then drawn from informed broad prior distributions, assuming a generation time of 100 years (Supplementary Information 3), and 10^6 microsatellite data sets were simulated for each scenario. As summary statistics, we computed the mean number of alleles, mean genetic diversity, mean allele size variance and mean Garza–Williamson's M (Garza and Williamson, 2001). $M = k/r$ is the mean ratio of the number of alleles, k, to the allele size range, r (Garza and Williamson, 2001), and is believed to detect more ancient and severe bottleneck signals than methods based on the deficit of rare alleles such as T_2 (Williamson-Natesan, 2005). We estimated the posterior probability of each scenario using logistic regression on the 1% simulated data sets with summary statistics closest to the observed data in each region and compared the 95% highest posterior density intervals (HPD) of scenarios (Cornuet *et al.*, 2008). We then estimated the posterior distribution of parameters from the 1% best simulated data sets using local linear regression and a logit transformation of parameters.

We also used ABC to examine the order of divergence of the four identified GPs/regions (see Results, Figure 3 and Supplementary Information 4). We chose a representative location of each region to correct for a potential bias in coalescence times that may occur due to uneven sampling across regions (more recent coalescence in under-sampled GPs/regions, Städler *et al.*, 2009). Representative locations were defined as >80% STRUCTURE ancestry at the location level and similar sample size: location 1 (Benin) for GP1, 2 (Cameroon Korup) for GP2, 8 (Cameroon SW) for GP3 and 26 (São Tomé) for GP4 (see Table 1 and Figure 1). In each location, all individuals of the location were included in the analysis to account for migration events or uncertainty in STRUCTURE ancestry proportions. We computed the pairwise differentiation statistics R_{ST} and $(\delta\mu)^2$ (Goldstein and Pollok, 1997) between these locations with SPAGeDi1.3d and constructed UPGMA trees using Phylip v.3.69 (Felsenstein, 1989). $(\delta\mu)^2$ should retain the phylogenetic signal between GPs better than R_{ST} as it is less affected by drift (Hardy *et al.*, 2003), although in our case, the same topology was recovered. We therefore considered two scenarios (topologies) to assess the divergence history in the study area: (1) the $(\delta\mu)^2$-based population tree and (2) simultaneous divergence of all locations (Figure 3). We simulated 10^6 data sets in each scenario with DIYABC using broad informed parameter priors (Supplementary Information 4). Because we previously detected an LGM or penultimate glacial bottleneck in all regions (see Results) and because geographic features separating regions (CVL, ocean barrier in the Gulf of Guinea) are older than these events, a bottleneck timed between 15 000 and 200 000 years, similarly in all GPs to reduce model complexity, was simulated along each terminal branch after population divergence events (Supplementary Information 4). As summary statistics, we used the pairwise F_{ST}, the pairwise shared allele distance, the pairwise $(\delta\mu)^2$, the mean classification index for two samples (Rannala and Mountain, 1997) and the mean genetic diversity per population (for a total of 34 summary

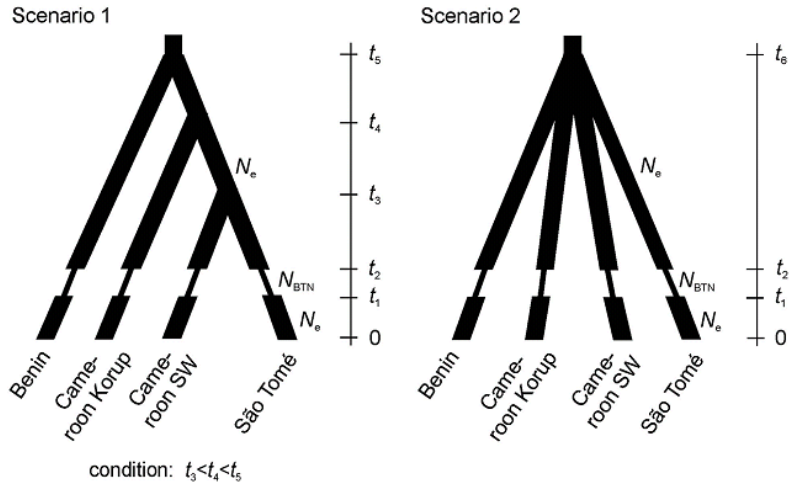

Figure 3 Two scenarios simulated in DIYABC to assess the population divergence history, (1) $(\delta\mu)^2$-based population tree and (2) simultaneous divergence of all locations. A bottleneck timed approximately between 15 000 and 200 000 years was simulated in terminal branches (for details see Supplementary Information 4).

statistics). Model choice and the estimation of posterior parameter distributions were done as described above. In addition, the goodness of fit of the models was checked by simulating 1000 data sets from the posterior predictive distribution of parameters, and comparing these to the observed data by means of summary statistics different from those chosen for model fitting: mean number of alleles, mean allele size variance and mean Garza-Williamson's *M*.

Pollen- vs seed-mediated gene flow

To estimate the relative contribution of seed dispersal to overall gene dispersal (σ_s/σ_g, where σ_s^2 and σ_g^2 are half the mean squared dispersal distances for seed and genes, respectively), we calculated $\sigma_s/\sigma_g = [2Sp(\text{nuclear})/Sp(\text{plastid})]^{1/2}$. To assess the pollen-to-seed dispersal distance ratio (σ_p/σ_s), we used $[(Sp(\text{plastid})/Sp(\text{nuclear}))\text{-}2]^{1/2}$ (formulas assuming hermaphrodite plants and purely maternal inheritance of plastids). The *Sp* statistics reflect the strength of the spatial genetic structure and were computed from the regression slope of pairwise kinship coefficients between individuals on the logarithm of the spatial distance (Vekemans and Hardy, 2004). These estimates were computed across the South Cameroon and Gabon region, which had sufficient sample size and was closer to demographic equilibrium than the other regions.

RESULTS

CpDNA diversity and structure

In a total of 94 *S. globulifera* samples from 18 locations from the Guineo-Congolian region, we identified 22 SNPs in the *psbA-trnH* intergenic region, resolving 22 haplotypes (Supplementary Information 5, Figure 2, GenBank accession numbers JQ996246-JQ996339). The haplotype network presented several reticulations reflecting homoplasy (that is, recurrent mutations at some sites) in the *psbA-trnH* region (Figure 2, Supplementary Information 5). Neighbouring haplotypes differed by a maximum of three mutations. Several geographic locations presented a single haplotype and most haplotypes (19 out of 22) were endemic to a single location. Endemic haplotypes occurred in proposed refuges (*sensu* Maley) and in other locations (Figure 2, Table 2 and Supplementary Information 5). Haplotype diversity was assessed in each location considering phylogenetic information using *v* statistics, or ignoring this information using *h* statistics. Population 'Cameroon SW' (8), located on a proposed refuge, was the most diverse location with four haplotypes ($h_s = 0.680$ and $v_s = 0.336$), whereas 'Cameroon S2' (6) was the most phylogenetically diverse harbouring two distantly related haplotypes (H10 and H14, $h_s = 0.667$ and $v_s = 1.343$; Table 2). The south of Lower Guinea (Gabon and Democratic Republic of the Congo) harboured haplotypes that were all closely related to the widespread H07 (maximum distance of three mutations, $h_s = 0.795$ and $v_s = 0.604$), whereas the north of Lower Guinea (South Cameroon) harboured more phylogenetically diverse haplotypes that were related

Table 2 Genetic diversity for *psbA-trnH* haplotypes in *S. globulifera*

	Location	n	A	A_P	h_S	v_S
1	Benin	13	1	1	0	0
2	*Cameroon Korup*	14	1	1	0	0
5	Cameroon S1	5	2	1	0.600	0.403
6	Cameroon S2	3	2	2	0.667	1.343
7	Cameroon SE	3	2	1	0.667	0.448
8	*Cameroon SW*	13	4	3	0.680	0.336
9	*Cameroon Volcanic Line*	1	1	1	NC	NC
10	Cameroon W coast	1	1	0	NC	NC
12	Gabon NE1	2	1	0	0	0
14	*Gabon Crystal*	4	1	1	0	0
15	Gabon coast	14	1	1	0	0
18	*Gabon Waka*	2	2	1	1	0.336
20	Gabon SE	1	1	1	NC	NC
21	Gabon SW	2	2	1	0	0.336
22	*Gabon Gamba-Doudou*	6	3	2	0.600	0.336
23	DRC W1	4	2	1	0.667	0.224
24	DRC W2	1	1	0	NC	NC
26	São Tomé	5	1	1	0	0
	North of Lower Guinea	26	10	9	0.895	0.850
	South of Lower Guinea	36	10	9	0.795	0.604
	Overall	94	22	22	0.968	0.968

Abbreviations: *A*, number of haplotypes; A_P, number of private or endemic haplotypes; DRC, Democratic Republic of the Congo; h_S and v_S, gene diversity per location based on unordered and ordered alleles, respectively; *n*, sample size; NC, not calculated.
Location, sampling location name (in italics and underlined for locations coinciding with refuges *sensu* Maley (1996). Numbers correspond to locations in Figure 2.

to H07 or H17 ($h_s = 0.895$ and $v_s = 0.850$; Table 2, Figure 2). G_{ST} among locations was 0.566 and N_{ST} was 0.722, significantly larger than $N_{ST[permuted]}$ ($P < 0.001$) and indicative of a phylogeographic signal at cpDNA.

Genetic diversity and structure at SSRs

We identified a total of 111 alleles and an average heterozygosity of $H_E = 0.860$ at the five SSRs (for summary statistics per locus and sampling location, see Supplementary Information 1). Log-likelihood values for the admixture model in STRUCTURE increased as a function of the number of clusters K, reaching a plateau for $K \geqslant 4$ and the delta K vs K graph (Evanno *et al.*, 2005) was bimodal, suggesting $K = 2$ or $K = 4$ as best describing the data (Supplementary Information 6). For $K = 2$, one cluster comprised Benin and the majority of the samples from Cameroon and Gabon, and the other cluster samples from West Cameroon (Korup) and from São Tomé (Figure 1). For $K = 4$, GP1 was essentially represented in Benin (location 1) and in a coastal region close to Douala (location 10) in Cameroon; GP2 was centred on West Cameroon (Korup, location 2); GP3 was centred on South Cameroon (east of the CVL) and Gabon; and GP4 was mostly represented in São Tomé (location 26; Figure 1). No substructure was identified in any of the four GPs. When geographic location of individuals was used as prior information in TESS 2.3.1 (Chen *et al.*, 2007), the same four GPs were identified (Supplementary Information 5) with slightly higher average maximum individual ancestry (87.9% ± 0.13% (s.d.) vs 83.7% ± 0.15% (s.d.) in STRUCTURE).

According to the primary location of the four GPs, we defined four main geographic regions; Benin (location 1); West Cameroon (Korup, location 2); South Cameroon (east of the CVL) and Gabon (locations 5–8 and 11–24); and São Tomé (location 26; Figure 1). Sampling locations 3, 4 and 9 of the CVL were not included because they were located in a contact zone between GPs and had small sample sizes (three to eight individuals). Also, location 10 (Cameroon W coast) was not included as this was a small sample (six individuals) from a distinct habitat (river bank) that we suspected to have distinct history (see Discussion). Location 25 was excluded because it had also small sample size (two individuals) and belonged to a distinct phytogeographic region (Congolia). Differentiation among regions was $F_{ST} = 0.135$ ($P < 0.0001$) and $R_{ST} = 0.185$ ($P < 0.0001$), and there was a phylogeographic signal ($R_{ST} > R_{ST[permuted]}$, $P < 0.01$) across the study area. Pairwise F_{ST} between regions ranged from 0.095 to 0.183 and phylogeographic signals were detected for all pairwise comparisons between Benin, West Cameroon and São Tomé (Table 3). Endemic alleles were found in all regions, 3 in Benin, 2 in West Cameroon, 31 in the geographically extended South Cameroon and Gabon, and 2 in São Tomé. Allelic richness was highest in South Cameroon and Gabon, whereas heterozygosity was similar in all regions (Table 4).

Table 3 Pairwise genetic differentiation between regions in *S. globulifera*

F_{ST}/R_{ST}	*Benin*	*West Cameroon*	*South Cameroon, Gabon*	*São Tomé*
Benin		0.109*	0.095*	0.183*
West Cameroon	0.410*		0.127*	0.128*
South Cameroon, Gabon	0.121	0.167		0.154*
São Tomé	0.415**	0.213*	0.135	

Above diagonal: F_{ST} (test with H1: $F_{ST} \neq 0$: $^*P < 0.05$); below diagonal: R_{ST} (test with H1: $R_{ST} > R_{ST[permuted]}$: $^*P < 0.05$; $^{**}P < 0.01$). Significance values refer to levels after Bonferroni correction.

Demographic history inference

Analyses with the Bottleneck program suggested signals of recent bottlenecks in Benin and West Cameroon (Supplementary Information 2). ABC analyses for population size change indicated that scenario 3 with a bottleneck timed approximately during the LGM had higher posterior probability than alternative scenarios in these regions, and its likelihood (51–58%, based on the 95% HPD) was about twice that of scenario 2 with a more recent bottleneck (Table 4). In São Tomé, the LGM bottleneck was also the most likely scenario, unequivocally (HPD 83–86%). In South Cameroon and Gabon, scenario 4 with a bottleneck during the penultimate glacial was the most likely, although its HPD (35–64%) overlapped with that of scenario 1 of constant population size (24–44%, Table 4). This older bottleneck was not an artefact because of a more extended sampling compared with other regions, as repeating the simulations in a representative location (8) confirmed scenario 4 (probability 90–92%, 95% HPD). Posterior estimates for scaled population sizes (by mutation rate) and bottleneck time events converged, although 95% confidence intervals were wide (Table 5, Supplementary Information 3). They indicated an older bottleneck in South Cameroon and Gabon, as expected from model choice results. The data were insufficient to conclude on any population size differences, either current or during the bottleneck, across regions.

The two divergence scenarios, (1) $(\delta\mu)^2$-based population tree and (2) simultaneous divergence of all locations (Figure 3), could not be distinguished using DIYABC: their respective probabilities were 0.460 and 0.540, with overlapping HPDs, (0.403,0.518) and (0.482,0.597). The estimates of divergence times scaled by mutation rate converged in both scenarios, but had broad confidence intervals (Supplementary Information 4). In scenario 1, the scaled median time of the oldest divergence event, divergence of Benin, $t_5\mu$, was 4.98 (1.26–20.2, 95% CI), which translated to ~3.32 Ma ago (0.84–13.47 Ma ago, 95% CI) when dividing by the estimated mutation rate and multiplying by generation time. Median estimates for subsequent divergence events $t_4\mu$ and $t_3\mu$ lay within the 95% CI of $t_5\mu$. In scenario 2, the scaled median time of simultaneous divergence of all locations, $t_6\mu$, was 3.01 (0.61–13.7, 95% CI), translating to ~1.31 Ma ago (0.27–5.98 Ma ago, Supplementary Information 4). Model checking revealed a relatively good fit: comparing our data to data sets simulated from the posterior predictive distribution, only two (Garza-Williamson's M for locations Cameroon SW (8) and São Tomé (26)) of 12 summary statistics lay outside of the 95% CI, but within the 99% CI for scenario 1, and one summary statistic (M for São Tomé (26)) lay between the 95 and the 99% CI for scenario 2. Therefore, the poor model discrimination and poor precision of divergence time estimates can rather be attributed to the information content of the data than to bad choice of model priors.

Pollen- vs seed-mediated gene flow

The relative contribution of seed to overall and to pollen-mediated gene dispersal distances was assessed in the South Cameroon and Gabon region. Significant spatial genetic structure occurred both for cpDNA, Sp(plastid) = 0.118, and for SSRs, Sp(nuclear) = 0.007, in this region. The ratio of seed to gene dispersal distance, σ_s/σ_g, was 0.35. The ratio of pollen to seed dispersal distance, σ_p/σ_s, was 3.84,

Table 4 Nuclear genetic diversity, allelic richness and demographic signatures in four geographic regions in *S. globulifera*

	R_S	H_E	*Scenario 1: constant-size population*	*Scenario 2: ca 3000–1000 BP bottleneck*	*Scenario 3: LGM bottleneck*	*Scenario 4: penultimate glacial bottleneck*
Benin	7.86 (0.91)	0.770 (0.061)	0.0631 (0.0590, 0.0672)	0.3081 (0.3001, 0.3161)	**0.5208 (0.5120, 0.5236)**	0.1080 (0.1023, 0.1137)
West Cameroon	8.33 (1.47)	0.803 (0.041)	0.0631 (0.0593, 0.0669)	0.2663 (0.2590, 0.2736)	**0.5783 (0.5703, 0.5864)**	0.0923 (0.0875, 0.0972)
South Cameroon and Gabon	11.36 (2.40)	0.769 (0.086)	0.3401 (0.2391, 0.4412)	0.0032 (0.0020, 0.0043)	0.1609 (0.1083, 0.2136)	**0.4957 (0.3527, 0.6388)**
São Tomé	8.16 (1.27)	0.772 (0.057)	0.0001 (0.0000, 0.0001)	0.1487 (0.1368, 0.1607)	**0.8461 (0.8341, 0.8581)**	0.0051 (0.0040, 0.0062)

Abbreviation: LGM, last glacial maximum.
R_S, allelic richness and its standard error (s.e.) for a sample of 16 individuals per population; H_E (SE), gene diversity; Scenario 1–4, posterior probabilities (95% highest posterior density intervals) for four demographic scenarios estimated with DIYABC from the 1% simulations closest to the observed data. The most likely scenario is indicated in bold type.

Table 5 Posterior parameter distribution (median and 95% confidence interval) for the most likely demographic scenario in each region (scenario 3 with an LGM bottleneck in Benin, West Cameroon and São Tomé and scenario 4 with a bottleneck during the penultimate glacial in South Cameroon and Gabon)

	Benin	*West Cameroon*	*South Cameroon and Gabon*	*São Tomé*
$N_e\mu$	8.1 (2.5, 27.7)	11.6 (3.7, 33.6)	27.1 (11.2, 46.0)	25.7 (10.9, 44.0)
$N_{BTN}\mu$	0.080 (0.014, 0.304)	0.114 (0.022, 0.352)	0.259 (0.110, 0.452)	0.095 (0.020, 0.295)
$t_a\mu$	0.042 (0.011, 0.131)	0.058 (0.017, 0.140)	0.767 (0.315, 1.210)	0.084 (0.034, 0.144)
$t_b\mu$	0.060 (0.014, 0.185)	0.082 (0.022, 0.192)	1.130 (0.462, 1.790)	0.131 (0.051, 0.204)

Abbreviation: LGM, last glacial maximum.
Current and bottleneck population sizes, N_e and N_{BTN}, and time of beginning t_a and end t_b of bottleneck are given, scaled by mutation rate μ.

indicating that gene flow via pollen was more wide-ranging than via seeds in South Cameroon and Gabon.

DISCUSSION

Genetic diversity and differentiation of *S. globulifera*, an ancient tropical rainforest tree

Genetic diversity in African *S. globulifera* was similar to that in America. Expected heterozygosity, H_E, at SSRs was 0.75 (Africa, this study) and 0.78 (Neotropics, Dick and Heuertz, 2008; with three SSRs overlapping with this study). These values are typical for widespread outcrossing long-lived plants growing at late succession stages (Nybom, 2004). Plastid DNA also showed similar polymorphism in both continents at the *psbA-trnH* fragment with 22 SNPs in Africa and 23 in America (Dick and Heuertz, 2008). This is a high level of variation compared with other species in the same African study region: 2 SNPs in *Milicia excelsa* (Daïnou *et al.*, 2010) or 14 SNPs in *Anthonotha macrophylla* s. l. (Leguminosae; M Heuertz and OJ Hardy, unpublished data), a taxon that potentially harbours several species (FJ Breteler, Personnel Communication, 2007). High diversity at the population and species level in *S. globulifera* suggests high effective population sizes, estimated to ~25 000 (GP1, Benin) to ~45 000 (GP3, South Cameroon and Gabon) from scaled N_e estimates. These estimates are somewhat smaller than estimates in the single *Pinus taeda* GP (~100 000, Willyard *et al.*, 2007), which has a wide distribution in the SE of the USA and grows at much higher density than *S. globulifera*. In *S. globulifera*, consistency with N_e can be attributed to a long history of effective breeding and dispersal strategies ensuring population connectivity across large parts of the range. Genetic differentiation was also similar on both continents in *S. globulifera*: $F_{ST} = 0.135$ among regions/GPs in Africa (this study) vs $F_{ST} = 0.138$ among neotropical populations. Our results suggest that the genetic structure found in Africa (ca 0.27–13.47 Ma ago, from ABC) originated probably posterior to the first colonization of America by *S. globulifera* (15–18 Ma ago, Dick *et al.*, 2003). A joint analysis of sequence variation in a phylogenetic framework should help clarifying the relationships between haplotypes from both continents and provide evidence for the number of colonization events (suggested to be three in Dick *et al.*, 2003) and possibly for the African GP of origin (or region, assuming absence of major range shifts) of the more recent colonization events. The *psbA-trnH* region might not be the most appropriate for this, as homoplasious point mutations and complex insertion/deletion structure may have erased part of the phylogenetic signal.

Geographic features shape the structure of genetic diversity

Geographic features explain the genetic structure obtained from STRUCTURE and TESS analyses on SSRs in *S. globulifera* in Atlantic Equatorial Africa: GP1 (Benin) was almost exclusive to coastal forests of the Dahomey gap, a forest-savannah mosaic separating Upper and Lower Guinea, GP2 (West Cameroon) and the large GP3 occurred predominantly on opposite sides of the CVL, and GP4 occurred mostly on São Tomé, which is separated from the African mainland by an ocean barrier within the Bight of Bonny. The same geographic pattern was also apparent in cpDNA data, where locations separated according to these geographical features did not share any haplotypes and, in some cases, were fixed for endemic haplotypes: H04 in Benin, H13 in West Cameroon (Korup) and H03 in São Tomé. This apparent fixation of haplotypes may result from pronounced genetic drift, as expected for cpDNA under ancient divergence; however, it could also reflect uneven sampling effort as cpDNA data have been obtained for only one or very few localities in GP1, GP2 and GP4.

The order of divergence events leading to the recognized SSR GPs could not be distinguished based on a modelling approach using ABC. This result allows the following interpretations: (1) distinct GPs

originated from a single ancestral population at similar times (that is, no hierarchical tree topology) or (2) improved data or an improved model may have led to different conclusions. Notably, we did not consider the possibility of gene flow between adjacent GPs, such as GP2 and GP3 separated by the CVL. Increasing sampling effort in regions where distinct GPs come into contact, increasing the number of markers and improving coalescent models should help to disentangle the population divergence history of *S. globulifera* in Africa in the future.

S. globulifera from Benin was clearly differentiated from Lower Guinea, supported by SSR divergence considering stepwise mutations (R_{ST}), by STRUCTURE and TESS results (GP1) and also by a long deletion (152 bp) at *psbA-trnH* in the Beninese samples (data not shown). This genetic distinctiveness could have two explanations, related either to geography or to ecology. Based on geography, a Beninese GP differentiated from Lower Guinea can be expected, as has been observed for another widespread forest tree, *Milicia excelsa* (Daïnou *et al.*, 2010). A biogeographic divide between Upper and Lower Guinea is reflected in both patterns of plant and animal endemism (White, 1979; Mayr and O'Hara, 1986; Linder, 2001) and, for some species, in the population genetic structure (for example, *Coffea canephora* (Gomez *et al.*, 2009); *Santiria trimera* (Koffi *et al.*, 2011)). This divide is presumably caused by recurrent connections and separations of both forest blocks since the end of the Pliocene, the latest opening of the Dahomey gap dating to ca 3700 BP (Maley, 1996). In *M. excelsa*, Beninese populations were interpreted to be of Upper Guinean origin because they showed the strongest pairwise differentiation in the study region, which comprised Benin and Lower Guinea (Daïnou *et al.*, 2010). A similar Upper Guinean origin is also plausible in *S. globulifera* from Benin. Analysis of plant material from Upper Guinea and from Nigeria would be necessary to test this hypothesis. With regard to ecology, Beninese *S. globulifera* trees were all sampled in coastal forests. A closer look to STRUCTURE ancestry proportions revealed that nine individuals outside Benin had ancestry $\geqslant$0.9 to GP1 characteristic of Benin (but had local cpDNA haplotypes). Five of these individuals were sampled on the bank of the Sanaga river (location 10), one in the coastal forest of Cap Estérias (location 15), one on the bank of the Ogooué river (location 19) and two in locations that were not in the immediate vicinity of rivers (locations 21 and 25 (at ca 10 km from the Congo river bank)). TESS recovered six of these individuals (from the Sanaga bank and Cap Estérias). These results suggest that there might be ecotype differentiation—coastal or swamp vs *terra firme*—in *S. globulifera*, which could be tested using more powerful markers and a more systematic sampling of coastal forests.

The CVL started forming at 30 Ma and represents a 1000 km chain of terrestrial volcanoes close to the Cameroon-Nigeria border and of seamounts in the Gulf of Guinea, including São Tomé (emerged ca 13 Ma, Meyers *et al.*, 1998). In the STRUCTURE analysis for $K=4$, confirmed by TESS results, GP2 typical of location 2 west of the CVL, GP3 typical of South Cameroon and Gabon, and GP4 typical of São Tomé were found together on the CVL (location 9) or east of the CVL (locations 3 and 4). In these locations, individual ancestry proportions were typically high in one of these three GPs (maximum average ancestry 0.86 for GP2, 0.76 for GP3 and 0.66 for GP4) and just a few individuals were admixed. These results suggest that distinct GPs persisted on the CVL or close to it, or that they came into secondary contact on the CVL. Persistence on the CVL is in agreement with the suggestion that the CVL harboured a Pleistocene forest refuge, based on its distinctive flora and palaeoecological data (Maley and Brenac, 1998). The presence of GP4 (São Tomé) on and close to the CVL suggests either gene flow from São Tomé, or, alternatively, a CVL origin of the São Tomé GP. The latter hypothesis is supported to some extent by the indistinctness of São Tomé and West Cameroon in the STRUCTURE analysis for $K=2$, suggesting a common origin of these GPs. Colonization of São Tomé along with the CVL has been proposed for *Drosophila santomea* (Cariou *et al.*, 2001), whereas studies on plants more generally proposed colonization of São Tomé from Lower Guinean forests (Stévart, 2003; Koffi *et al.*, 2011). The $(\delta\mu)^2$-based UPGMA tree in *S. globulifera* rather supported colonization of São Tomé from the South Cameroon and Gabon region, whereas cpDNA data concurred with both colonization scenarios as H03 from São Tomé was fairly closely related to H01 from South Cameroon (locations 5 and 7, east of the CVL) and to H13 from Korup (location 2, west of the CVL).

Demographic processes within regions/GPs and examination of the forest refuge hypothesis

The forest refuge theory (Haffer, 1969; Prance, 1982) presumes the isolation of populations in refugia during unfavourable climatic conditions of the Pleistocene. The expected effects are population size reductions (bottlenecks) and subsequent population expansions when conditions improve. These demographic changes leave signals in the genetic constitution of a species that can be examined with a variety of analytical tools (Knowles, 2009).

In *S. globulifera*, we detected bottleneck signals compatible with forest reduction during the LGM in Benin (GP1), in West Cameroon (GP2) and in São Tomé (GP4). These results should be interpreted keeping in mind that the dating of bottlenecks is imprecise as it was based on model choice using ABC, only few competing demographic models were tested, and the scenarios in these models were kept relatively simple. GP2 is located close to the proposed Pleistocene forest refuge on the CVL (Maley, 1996; Maley and Brenac, 1998). This refuge was supported by the discovery of endemic alleles in the rainforest trees *Santiria trimera* (Koffi *et al.*, 2011) and *Irvingia gabonensis* (Lowe *et al.*, 2010). Therefore, one interpretation is that *S. globulifera* in West Cameroon shared this refuge (see previous section) and has not yet recovered from an LGM bottleneck. *S. globulifera* from Benin and São Tomé also bore signals of an LGM bottleneck. This could be attributed to local persistence in reduced-size populations (for instance in gallery forests), or alternatively, to a recent colonization of Benin from Upper Guinea (see previous section). Our genetic divergence results are compatible with Pleistocene persistence of *S. globulifera* on São Tomé. The oceanic climate has probably buffered the Pleistocene climatic oscillations on the island, which is a recognized refuge for Pre-Pleistocene lineages of other forest plants (for example, Begonias, Plana *et al.*, 2004).

S. globulifera from South Cameroon and Gabon bore an older bottleneck signal than the other regions. It was detected with DIYABC but not with the BOTTLENECK software. This illustrates that using Garza-Williamson's M as summary statistic in an ABC approach improved the power of bottleneck detection compared with approaches based on heterozygosity (Garza and Williamson, 2001; Peery *et al.*, 2012). The absence of a recent bottleneck signal indicates that GP3 was either not as heavily affected by the LGM as other GPs, or that it recovered better, or both. Instead, allele size ranges conserved signals of (poorly dated) older demographic dynamics, which broadly coincide with late Pleistocene forest contraction–expansion dynamics in Lower Guinea and Congolia inferred from offshore pollen deposits (Dupont, 2011). Most *S. globulifera* plastid haplotypes in South Cameroon and Gabon had narrow ranges or were endemic at the location level. Endemic haplotypes were found in

all locations belonging to the proposed refuges examined (*sensu* Maley, 1996), as well as outside of them, suggesting wide-ranging long-term persistence of *S. globulifera* in the region. Narrow haplotype ranges contrast with the absence of substructure at the SSR level (established using STRUCTURE), which suggests much stronger genetic drift for cpDNA than for SSRs. This can be attributed to the lower historical effective population sizes of cpDNA due to the lack of recombination and the exclusively maternal inheritance of cpDNA. The four times further ranging pollen than seed dispersal detected in our study must have contributed to fast Holocene recovery of effective population sizes in South Cameroon and Gabon, whereas, on the other side, we did not find clear evidence for recolonization by seed from major postulated refuges in West Cameroon, SW Cameroon and the Crystal Mountains in Gabon (Maley, 1996). The only haplotype that ranged over hundreds of km, H07, occurred in the proposed refuges of SW Cameroon, and of the Chaillu massif (Waka) and the Gamba-Doudou complex in Gabon, suggesting that it has been widespread before the forest contraction of the LGM. Our limited sampling in the Congo Basin revealed endemic haplotypes but did not allow us to assess the importance of the proposed Congo refuge in terms of diversity or recolonization.

From these observations, we infer that our data supports partly the refuge hypothesis *sensu* Maley (1996), as endemic cpDNA haplotypes were found in all the proposed refugia examined. Our results indicated that *S. globulifera* survived unfavourable Pleistocene climatic conditions in additional locations, such as São Tomé and regions in Lower Guinea not previously proposed as refugia. Some of these locations may correspond to floodplain or gallery forests with assured water availability, similarly to the Pleistocene riverine persistence that has been proposed for some Caesalpinioideae species in Gabon (Leal, 2004). In *S. globulifera*, the survival in numerous, widespread forest fragments as well as high levels of pollen gene flow were probably key to a fast recovery of the species after glacial periods in Lower Guinea.

CONCLUSIONS

This study highlights the importance of geographic features in structuring the genetic constitution of the ancient rainforest tree species *S. globulifera* in the Guineo-Congolian phytogeographic region, in agreement with other population genetic studies of plant or animal taxa (Cariou *et al.*, 2001; Brouat *et al.*, 2004; Daïnou *et al.*, 2010). Recent (LGM) bottlenecks were inferred in Benin, West Cameroon and São Tomé, suggesting that *S. globulifera* was affected by Pleistocene cycles of forest contraction. In the GP located in South Cameroon and Gabon, wide-ranging LGM persistence of the species enabled rapid post-glacial recovery of population sizes, especially through pollen flow, and only an older bottleneck remained traceable. In this region, *S. globulifera* did not display SSR substructure, contrarily to other forest tree species (Born *et al.*, 2011; Debout *et al.*, 2011; Daïnou *et al.*, 2010). Furthermore, our study suggested ecotypic differentiation between coastal or swamp forest and *terra firme S. globulifera*. Our findings underline that species, and even GPs within species, responded individually to past disturbances. Therefore, the basis for making reliable predictions for communities in the face of current climate change should lie in gathering and jointly interpreting solid case studies on past demographic scenarios in species with a wide range of life history traits (Knowles, 2009).

DATA ARCHIVING

Microsatellite data have been deposited on Dryad (doi:10.5061/dryad.3j300) and cpDNA sequences have been deposited on GenBank (JQ996246-JQ996339).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Carmen García-Barriga (INIA-CIFOR, Madrid) for technical assistance in the laboratory and Armand Boudaby, Charlemagne Nguembou, Crépin Djopamdé, Elie Montchowui, Emerand Gassang, Fidel Baya, Gabriel Debout, Jean-François Gillet, Jérôme Duminil, Jérôme Laporte, Michel Arbonnier, Gilbert Todou, Nils Bourland, Paul Zok, Pierre-André Ntchandi-Otimbo, Pierre Agbani, Théophile Ayol, Charles Doumenge, Ingrid Parmentier, Peter Mambo, Gilles Dauby, Kasso Daïnou, Eben-Ezer Ewedje and Guillaume Koffi for their contribution to the sample collection. We also acknowledge the logging companies Pallisco, SFID, Wijma, CEB and CBG, the Missouri Botanical Garden (Central African Program), the CENAREST (Gabon), the Smithsonian Institution (Gabon Biodiversity Program), Patrice Ipandi (ENEF, Gabon), Bonaventure Sonké (Université de Yaoundé I, Cameroon) and Charles Doumenge (CIRAD) for facilitating field work and sample collection and thank Charles Doumenge and Katalin Csilléry for stimulating discussions. KB acknowledges a PhD scholarship (FPI) of the Spanish Ministry of Science and Innovation (MICINN). MH is a 'Ramón y Cajal' researcher of the MICINN and acknowledges funding from MICINN (JAE-Doc program), from the Belgian Fund for Scientific Research (FRS-FNRS projects 1.2.155.06F, FC66560-1.5.095.08F and F.4.519.10.F) and the Spanish Ministry for Economy and Competitiveness (CGL2012-40129-C02-02/AFFLORA).

Abdul-Salim K (2002). *Systematics and Biology of Symphonia L. f. (Clusiaceae)*. PhD thesis. Harvard University: Boston.
Aldrich PR, Hamrick JL (1998). Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. *Science* **281**: 103–105.
Aldrich PR, Hamrick JL, Chavarriaga P, Kochert G (1998). Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Mol Ecol* **7**: 933–944.
Bandelt HJ, Forster P, Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**: 37–48.
Bittrich V, Amaral MCE (1996). Pollination biology of *Symphonia globulifera* (Clusiaceae). *Plant Syst Evol* **200**: 101–110.
Bonnefille R (2007). Rainforest responses to past climate changes in tropical Africa. In: *Tropical rainforest responses to climate change*. Praxis Publishing: Chichester. pp. 117–170.
Born C, Alvarez N, McKey D, Ossari S, Wickings EJ, Hossaert-McKey M *et al.* (2011). Insights into the biogeographical history of the Lower Guinea Forest Domain: evidence for the role of refugia in the intraspecific differentiation of Aucoumea klaineana. *Mol Ecol* **20**: 131–142.
Brouat C, McKey D, Douzery EJP (2004). Differentiation in a geographical mosaic of plants coevolving with ants: phylogeny of the Leonardoxa africana complex (Fabaceae: Caesalpinioideae) using amplified fragment length polymorphism markers. *Mol Ecol* **13**: 1157–1171.
Cariou ML, Silvain JF, Daubin V, Da Lage JL, Lachaise D (2001). Divergence between *Drosophila santomea* and allopatric or sympatric populations of *D. yakuba* using paralogous amylase genes and migration scenarios along the Cameroon volcanic line. *Mol Ecol* **10**: 649–660.
Chen C, Durand E, Forbes F, François O (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Mol Ecol Notes* **7**: 747–756.
Cornuet JM, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.
Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ *et al.* (2008). Inferring population history with DIYABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**: 2713–2719.
da Silva Carneiro F, Sebenn AM, Kanashiro M, Degen B (2007). Low interannual variation of mating system and gene flow of *Symphonia globulifera* in the Brazilian Amazon. *Biotropica* **39**: 628–636.
Daïnou K, Bizoux JP, Doucet JL, Mahy G, Hardy OJ, Heuertz M (2010). Forest refugia revisited: SSRs and cpDNA sequences support historical isolation in a wide-spread African trees with high colonization capacity, *Milicia excelsia* (Moraceae). *Mol Ecol* **19**: 4462–4477.
Dauby G, Duminil J, Heuertz M, Hardy OJ (2010). Chloroplast DNA polymorphism and phylogeography of a Central African tree species widespread in mature rainforests: *Greenwayodendron suaveolens* (Annonaceae). *Trop Plant Biol* **3**: 4–13.
Debout GDG, Doucet JL, Hardy OJ (2011). Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, Distemonanthus benthamianus (Caesalpinioideae). *Heredity* **106**: 88–99.

Degen B, Bandou E, Caron H (2004). Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity* **93**: 585–591.
Dick CW, Abdul-Salim K, Bermingham E (2003). Molecular systematic analysis reveals cryptic Tertiary diversification of a widespread tropical rain forest tree. *Am Nat* **162**: 691–703.
Dick CW, Heuertz M (2008). The complex biogeographic history of a widespread tropical tree species. *Evolution* **62**: 2760–2774.
Droissart V (2009). *Étude taxonomique et biogéographique des plantes endémiques d'Afrique centrale atlantique: le cas des Orchidaceae*. PhD thesis. Université libre de Bruxelles: Brussels.
Duminil J, Heuertz M, Doucet JL, Bourland N, Cruaud C, Gavory F *et al*. (2010). CpDNA-based species identification and phylogeography: application to African tropical tree species. *Mol Ecol* **19**: 5469–5483.
Dupont L (2011). Orbital scale vegetation change in Africa. *Quat Sci Rev* **30**: 3589–3602.
Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
Ewing B, Hillier L, Wendl M, Green P (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res* **8**: 175–185.
Felsenstein J (1989). PHYLIP–Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164–166.
Forget PM, Dennis AJ, Mazer SJ, Jansen PA, Kitamura S, Lambert JE *et al*. (2007). Seed allometry and disperser assemblages in tropical rainforests: a comparison of four floras on different continents. In: Dennis AJ, Schupp EW, Green RJ, Westcott DA (eds) *Seed dispersal theory and its application in a changing world*. pp 5–36.
Garza JC, Williamson EG (2001). Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* **10**: 305–318.
Ghazoul J, Sheil D (2010). *Tropical Rain Forest Ecology, Diversity, and Conservation*. Oxford University Press: Oxford.
Goldstein DB, Pollock DD (1997). Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J Heredity* **88**: 335–342.
Gomez C, Dussert S, Hamon P, Hamon S, Kochko AD, Poncet V (2009). Current genetic differentiation of Coffea canephora Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evol Biol* **9**: 167.
Goudet J (1995). Fstat (Version 1.2): a computer program to calculate F-statistics. *J Heredity* **86**: 485–486.
Green P (2009). Phrap, version 1.090518. http://phrap.org
Haffer J (1969). Speciation in Amazonian Forest Birds. *Science* **165**: 131–137.
Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003). Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* **163**: 1467–1482.
Hardy OJ, Vekemans X (2002). SPAGeDI; a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.
Heuertz M, Carnevale S, Fineschi S, Sebastiani F, Hausman JF, Paule L *et al*. (2006). Chloroplast DNA phylogeography of European ashes, Fraxinus sp. (Oleaceae): roles of hybridization and life history traits. *Mol Ecol* **15**: 2131–2140.
Jan du Chêne RE, Salami MB (1978). Palynology and micropaleontology of the Upper Eocene of the well nsukwa-1 (Niger Delta, Nigeria). *Archives des Sciences* **13**: 5–9.
Knowles LL (2009). Statistical phylogeography. *Annu Rev Ecol Evol Syst* **40**: 593–612.
Koffi KG, Hardy JH, Doumenge C, Cruaud C, Heuertz M (2011). Diversity gradients and phylogeographic patterns in Santiria trimera (Burseraceae), a widespread African tree typical of mature rainforests. *Am J Bot* **98**: 254–264.
Leal ME (2004). *The African Rain Forest During the Last Glacial Maximum, an archipelago of Forests in a Sea of Grass*. PhD thesis. University of Wageningen.
Lescure JP, Boulet R (1985). Relationship between soil and vegetation in a Tropical Rain Forest in French Guiana. *Biotropica* **17**: 155–164.
Linder HP (2001). Plant diversity and endemism in sub-Saharan tropical Africa. *J Biogeogr* **28**: 169–182.
Lowe AJ, Harris D, Dormontt E, Dawson IK (2010). Testing putative African tropical refugia using chloroplast and nuclear DNA phylogeography. *Trop Plant Biol* **3**: 50–58.
Maley J (1996). The African rain forest: main characteristics of changes in vegetation and climate from the upper Cretaceous to the Quaternary. *Proc Roy Soc Edinburgh* **104B**: 31–73.
Maley J, Brenac P (1998). Vegetation dynamics, palaeoenvironments and climatic changes in the forests of western Cameroon during the last 28,000 years B.P. *Rev Paleobot Palyno* **99**: 157–187.
Mayr E, O'Hara RJ (1986). The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. *Evolution* **40**: 55–67.
McMahon SM, Harrison SP, Armbruster WS, Bartlein PJ, Beale CM, Edwards ME *et al*. (2011). Improving assessments of climate-change impacts on global biodiversity. *Trends Ecol Evol* **26**: 249–259.
Meyers J, Rosendahl B, Harrison C, Ding Z (1998). Deep-imaging seismic and gravity results from offshore Cameroon volcanic line, and speculation of African hotlines. *Tectonophysics* **284**: 31–63.
Millar CI, Stephenson NL, Stephens SL (2007). Climate change and forests of the future: managing in the face of uncertainty. *Ecol Appl* **17**: 2145–2151.
Moritz C, Patton JL, Schneider CJ, Smith TB (2000). Diversification of rainforest faunas: an integrated molecular approach. *Annu Rev Ecol Evol Syst* **31**: 533–563.
Ngomanda A, Neumann K, Schweizer A, Maley J (2009). Seasonality change and the third millennium BP rainforest crisis in southern Cameroon (Central Africa). *Quat Res* **71**: 307–318.
Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* **13**: 1143–1155.
Oyen LPA (2005). *Symphonia globulifera* L.f. [Internet] Record from Protabase. In: Louppe D, Oteng-Amoako AA, Brink M (eds) *PROTA (Plant Resources of Tropical Africa/Ressources végétales de l'Afrique tropicale)*. Wageningen: Netherlands <http://database.prota.org/search.htm> accessed on 29 March 2012.
Peery MZ, Kirby R, Reid BN, Stoelting R, Doucet-Bëer E, Robinson S *et al*. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Mol Ecol* **21**: 3403–3418.
Petit RJ, Aguinagalde I, Beaulieu JL, de Beaulieu JL, Bittkau C, Brewer S *et al*. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.
Plana V (2004). Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philos T Roy Soc B* **359**: 1585–1594.
Plana V, Gascoigne A, Forrest LL, Harris D, Pennington RT (2004). Pleistocene and pre-Pleistocene Begonia speciation in Africa. *Mol Phylogenet Evol* **31**: 449–461.
Pons O, Petit RJ (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**: 1237–1245.
Prance GT (1982). *Biological Diversification in the Tropics*. Colombia University Press: New York.
Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
Rannala B, Mountain JL (1997). Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA* **94**: 9197–9201.
Sang T, Crawford DJ, Stuessy TF (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). *Am J Bot* **84**: 1120–1136.
Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 1463–1463.
Sosef M (1994). Refuge begonias: taxonomy, phylogeny and historical biogeography of Begonia sect. Loasibegonia and sect. Scutobegonia in relation to glacial rain forest refuges in Africa. *Agricultural University Wageningen Papers* **94**: 1–306.
Stévart T (2003). *Étude taxonomique, écologie et phytogéographique des Orchidaceae en Afrique centrale atlantique*. PhD thesis. Université Libre de Bruxelles: Brussels.
Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P (2009). The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* **182**: 205–216.
Van Andel TR (2003). Floristic composition and diversity of three swamp forests in northwest Guyana. *Plant Ecol* **167**: 293–317.
Vekemans X, Hardy OJ (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* **13**: 921–935.
Vinson CC, Amaral AC, Sampaio I, Ciampi AY (2005). Characterization and isolation of DNA microsatellite primers for the tropical tree, *Symphoniaglobulifera* Linn. f. *Mol Ecol Resources* **5**: 202–204.
Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
White F (1979). The Guineo-Congolian region and its relationships to other phytochoria. *Bull Jard Bot Nat Belgique* **49**: 11–55.
Willyard A, Syring J, Gernandt DS, Liston A, Cronn R (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for Pinus. *Mol Biol Evol* **24**: 90–101.
Williamson-Natesan EG (2005). Comparison of methods for detecting bottlenecks from microsatellite loci. *Conserv Genet* **6**: 551–562.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)

SECTION II: Spatial genetic structure

Chapter 2

Fine-scale spatial genetic structure within *Symphonia globulifera* L.f. populations from different continents as related to dispersal vector species

This chapter is based upon the manuscript:

Budde KB, Heer K, González-Martínez SC, Dick CW, Heuertz M. Fine-scale spatial genetic structure within *Symphonia globulifera* L.f. populations from different continents as related to dispersal vector species. *In preparation*.

Fine-scale spatial genetic structure within *Symphonia globulifera* L.f. populations from different continents as related to dispersal vector species

Katharina B. Budde[1], Katrin Heer[2,3], Santiago C. González-Martínez[1], Christopher W. Dick[4], Myriam Heuertz[1,5]

[1]INIA Forest Research Centre, Department of Forest Ecology and Genetics, Carretera de A Coruña km 7.5, 28040 Madrid, Spain

[2]University of Ulm, Institute of Experimental Ecology, Albert Einstein Allee 11, 89069 Ulm, Germany

[3]University of Marburg, Karl-von-Frisch-Straße 8, 35043 Marburg, Germany

[4]University of Michigan, Department of Ecology and Evolutionary Biology and Herbarium, 830 North University Avenue, Ann Arbor, MI 48109, USA

[5]Université Libre de Bruxelles, Faculté des Sciences, Evolutionary Biology and Ecology cp160/12, 1050 Brussels, Belgium

Abstract

Fine scale spatial genetic structure (SGS) analysis typically reveals a decay of genetic relatedness as spatial distance increases between plants in a population. This spatial relatedness pattern can have impacts on the evolution of mating systems, genetic diversity and local adaptation. Pollen and seed dispersal modes determine gene dispersal and hence the fine scale spatial genetic structure within plant populations.

Symphonia globulifera is a widespread and ancient tropical tree species with disjunct distribution in Africa and the Neotropics. The divergence between the continents is several million years old, as *S. globulifera* appeared in the Neotropics around 17 MA, while pollen fossils from Africa date back to ~40 MA. Flowers are mainly pollinated by birds and insects and fruits are typically dispersed by a variety of small mammals. Differences between pollinator and seed dispersal vector species from both continents raised the question how disperser guilds might affect SGS patterns in natural populations of *S. globulifera.*

Our results indicated a slightly stronger SGS in African populations, while Neotropical populations displayed more leptokurtic gene dispersal. Furthermore, self-fertilization contributed to total inbreeding in most study populations but no difference between continents was detected. *Symphonia globulifera* flowers show a typical bird pollination syndrome and are pollinated by hummingbirds and perching birds in the Neotropics and sunbirds in Africa. This fact, in combination with our results might point to similar pollinator behavior on both continents. By contrast, bats are typical seed dispersers in America and might provide long distance seed dispersal but have never been described as vector species of *S. globulifera* seeds in Africa. This apparent absence of an analogous seed disperser, i.e. of the same functional type, could be the key to the differences observed in SGS patterns from Africa and America.

Key words: fine scale spatial genetic structure, *Symphonia globulifera*, gene flow, pollination, seed dispersal

Introduction

Fine scale spatial genetic structure (SGS), the non-random distribution of genotypes within plant populations, is shaped jointly by gene flow, drift and selective processes (Epperson 1990; Rousset 2004). The strongest determinant of the genetic structure at such a small spatial scale are typically gene dispersal abilities (Vekemans & Hardy 2004). Plants are sessile organisms and limited gene flow leads to stronger divergence between more distantly separated individuals. This pattern can have strong ecological and

evolutionary impacts, as it might decrease genetic diversity through increased bi-parental inbreeding between neighbouring trees or increase selection and improve local adaptation (Heywood 1991).

The peculiar conditions of tropical lowland rain forests, which are among the most species rich ecosystems worldwide (Myers *et al.* 2000) affect SGS of tree species. These rain forests typically harbour a high species richness of woody taxa, while single species often occur at low densities. The relatively large distances between congeneric individuals in combination with the humid conditions hamper wind pollination and consequently most tropical tree species rely on animal pollination (Regal 1977; Turner 2001). A review showed that pollination in tropical tree species occurs mainly between neighbours, although some studies reported more wide ranging pollen flow (Dick *et al.* 2008). Seed dispersal is also mainly performed by animals, as about 70% of tropical lowland tree species are animal dispersed and even higher percentages in more humid forests (Howe & Smallwood 1982). Usually, multiple seed disperser species act as vectors for the same tree species, with variable efficiency, while specialisations are rare (Schupp 1993). Life-history traits, such as pollination and seed dispersal modes, condition gene flow and hence SGS. Anemophilous species (more common in temperate regions) have wide ranging gene flow and therefore lower or non-significant SGS compared to animal pollinated ones (Vekemans & Hardy 2004). Furthermore, the behavior of different frugivores is known to affect dispersal distances and microhabitat of seed deposition (Jordano *et al.* 2007). In tropical trees, seed dispersal by gravity, wind (in the case of heavy seeds) or scatter-hoarding rodents is typically less far ranging than seed dispersal by monkeys, birds or bats and results in stronger SGS (Hardy *et al.* 2006; Dick *et al.* 2008).

A typical pollinator syndrome in tropical regions is that of “bird flowers”. These flowers are mainly scentless and with narrow tubular labellum. The conspicuous colours, mostly red, yellow or orange and the abundant nectar attract birds (Faegri & van der Pijl 1965). In the Neotropics hummingbirds often act as pollinators (Grant & Grant 1968), while sunbirds are common pollinators in Africa (Willmer 2011). The behavior of pollinator species can influence the evolution of plant traits and mating system and vice versa (Mitchell *et al.* 2009). For instance, foraging pollinator species tend to visit nearby flowers (Levin & Kerster 1968). Pollen dispersal might be limited due to these short pollinator flights and hummingbird pollinated plants are typically self-compatible (Wolowski *et al.* 2013), which can have implications for SGS (Wright 1931; Vekemans & Hardy 2004). Although weak selection pressures

of seed dispersers on plants are assumed (Herrera 1985), different general fruit syndromes can be observed that attract distinct disperser types. A characteristic fruit syndrome in tropical forests is that of “bat fruits”. These fruits are usually green or whitish, odorless or musky, fleshy, frequently rich in starch or lipids and often pendant (van der Pijl 1972). In the Neotropics frugivorous bats of the family Phyllostomidae (Microchiroptera) act as seed dispersers while in Africa megabats (Pteropodidae, Megachiroptera) are abundant seed dispersers. However, frugivorous bats are much less important in Africa compared to the Neotropics (Gautier-Hion *et al.* 1985). In general, bats are assumed to provide relatively wide ranging seed dispersal (Nathan 2006; Dick *et al.* 2008).

Symphonia globulifera L. f. (Clusiaceae) is a widespread and ancient tropical tree species. It occurs in tropical rain forests of Africa and America from sea level up to 2600 m asl. A dated phylogeny and fossil pollen records indicated colonization of the Neotropics through sweepstakes dispersal from Africa during the mid-Miocene (ca. 17 million years ago, Dick *et al.* 2003). In the Neotropics a phylogeograhic analysis revealed a cis- (Central America + west Ecuador) vs. trans-Andean (French Guiana + Amazonia) divergence (Dick *et al.* 2003; Dick & Heuertz 2008). Furthermore, a phylogeographic study in Atlantic Equatorial Africa exposed four gene pools based on microsatellites (one in Benin, one in West Cameroon, one in South Cameroon and Gabon and one in São Tomé; Budde *et al.* 2013). Gene diversity and *psbA-trnH* haplotype diversity were similar on both continents (Dick & Heuertz 2008; Budde *et al.* 2013). In Atlantic Equatorial Africa numerous endemic chloroplast haplotypes with narrow ranges indicated survival inside and outside of postulated forest refugia during glacial periods (Budde *et al.* 2013). *Symphonia globulifera* is a late successional species, requiring shade for germination. It is often found in the vicinity of rivers and mangroves, although it shows a wide ecological amplitude (Oyen 2005). A possible ecotype divergence between trees from swamp and *terra firme* forests has been suspected in sites from both continents (Baraloto *et al.* 2007; Budde *et al.* 2013) and distinct growth forms exist in Costa Rica (C.W. Dick, pers. obs.) and in Atlantic Equatorial Africa (C. Doumenge, pers. com.). A study of morphological characters revealed only gradual trait variation throughout the distribution range, insufficient to justify a splitting of the species (Abdul-Salim 2002).

The bright red flowers of *S. globulifera* are mainly pollinated by birds - sunbirds in Africa (Oyen 2005) and hummingbirds (Bawa *et al.* 1985; Bittrich & Amaral 1996; Degen *et al.* 2004) and perching birds (different species of Thraupidae, Bittrich & Amaral 1996; Gill *et al.* 1998) in the Neotropics (Table 1). However, the role of

flower visitors is not always clear and different species of bees (Bittrich & Amaral 1996; Oyen 2005), butterflies (Bittrich & Amaral 1996), hummingbirds (Gill *et al.* 1998) and monkeys (Bittrich & Amaral 1996) might be pollinators or nectar robbers. The fruits are usually dispersed by small mammals and some birds. In the Neotropics, dispersers include bats (Aldrich & Hamrick 1998; Degen *et al.* 2004; Hardy *et al.* 2006), nocturnal arboreal mammals and scatter hoarding rodents (Hardy *et al.* 2006). In contrast, in Africa, monkeys (Gautier-Hion *et al.* 1985; Oyen 2005), hornbills and ruminants (Gautier-Hion *et al.* 1985) have been described as fruit consumers, although their function as seed dispersers is not always clear. *Symphonia globulifera* is a predominantly outcrossing species but significant levels of bi-parental inbreeding and also selfing have been described (Degen *et al.* 2004; da Silva Carneiro *et al.* 2009). Degen *et al.* (2004) reported limited pollen flow with average effective pollen dispersal distances of 27 to 53 m in a site in French Guiana. By contrast, da Silva Carneiro *et al.* (2009) found more far ranging pollen dispersal in a continuous population in Brazil. However, in Paracou (French Guiana) significant SGS was detected in adult trees and seed dispersal, mainly mediated by bats in this site, was probably more far ranging than pollen flow (Degen *et al.* 2004). Studies describing SGS within *S. globulifera* populations have so far focused on the Neotropics revealing, essentially, a stronger SGS in fragmented forest patches than in continuous forest (Aldrich *et al.* 1998).

As mentioned before, this species is confronted with different guilds of vector species acting as pollinators or seed dispersers throughout its distribution range (Table 1). By comparing SGS patterns from African and American populations of *Symphonia globulifera* we aimed at identifying differences in SGS in populations from both continents and relating them to groups of associated species. We hypothesised that SGS might be less pronounced in Neotropical populations, where bat mediated long distance seed dispersal might weaken the relatedness of neighbouring trees.

Material and methods

Sample sites and plant material

A total of five populations from Africa and America were included in this study, two located in Cameroon, Mbikiliki (3º11.4'N, 10º31.8E) with n=83 individuals and NkongMekak (2º45.6'N, 10º32.4'E) with n=68 individuals, one in Brazil, Ituberá (13º47.4'S, 39º10.8'W) with n=83 individuals, one in French Guyana, Paracou (5º18'N, 52º53'W) with n=148 individuals and one in Panama, Barro Colorado Island (BCI, 9º9'N, 79º51'W) with n=148 individuals (Figure 1). The Neotropical populations BCI and Paracou

show strong divergence as they are separated by the Andes (Dick & Heuertz 2008). Ituberá in the Atlantic forest is separated by dry habitat from Paracou and probably also significantly differentiated. Both African populations, so far included in this study, are located in the south of Cameroon at a distance of 40 km and belong to the same gene pool (Budde *et al.* 2013). Leaf or cambium samples were collected in the study sites. Plant material was stored in a conservation buffer (Paracou, see Degen *et al.* 2004) or dried with silica gel (BCI, African plots and Ituberá). DNA was extracted using a CTAB protocol (samples from Paracou, Degen *et al.* 2004), the Qiagen DNeasy plant kit (Qiagen Corporation, Valencia, CA, samples from BCI) or the Invisorb DNA Plant HTS 96 Kit (Invitek, Berlin, Germany, samples from Mbikiliki, Nkong Mekak and Ituberá).

For Paracou, the genetic data was published by Degen *et al.* (2004) and contained three nuclear microsatellite (simple sequence repeat, SSR) loci: SgC4 from Aldrich *et al.* (1998), and Sg03 and Sg18 developed specifically in Degen *et al.* (2004). For BCI, genotyping followed methods described in (Dick & Heuertz 2008) for five SSRs: SgC4 and Sg19 (Aldrich *et al.* 1998), Sg03 and Sg92 [=Sg18], and Sg06 (Vinson *et al.* 2005). For Mbikiliki, NkongMekak and Ituberá, five SSRs were genotyped: Sg03, SgC4, Sg18 and Sg19, and in addition Sg10 from Vinson *et al.* (2005). PCR conditions were used as described by Budde *et al.* (2013). Forward primers were 5'end-labeled with fluorochromes (Sg03 – PET, SgC4 – FAM, Sg10 – PET, Sg18 – FAM and Sg19 – VIC). Amplified fragments were separated using an ABI 3730 genetic analyzer (Applied Biosystems, Carlsbad, USA) and fragment sizes were determined with reference to the GeneScan™ –500 LIZ® Size Standard (Applied Biosystems) using GeneMapper software version 4.0 (Applied Biosystems). Not all microsatellites described for *S. globulifera* amplify in all populations as it was already mentioned by Dick & Heuertz (2008), especially problematic was e.g. Sg06 that did not amplify in African populations and was impossible to read in Ituberá. These differences reveal a strength of genetic divergence between populations of *S. globulifera*, which is normally found at genus level.

Data analysis

For each sample location the expected heterozygosity or gene diversity (H_E) was calculated in GENEPOP 4.0 (Raymond & Rousset 1995). The fixation index (F_{IS}) was assessed and deviation from zero (Hardy-Weinberg genotypic proportions) was tested by 10,000 permutations of alleles within populations in SPAGeDi1.3d (Hardy & Vekemans 2002).

The fine scale spatial genetic structure (SGS) was analyzed for each sample

location respectively following Vekemans & Hardy (2004). Within-population pairwise kinship coefficients were calculated in SPAGeDi 1.3d (Loiselle *et al.* 1995) for all samples and a kinship-distance-plot was created to display SGS graphically. For this purpose, the average kinship coefficient per distance class was calculated and plotted against the logarithm of the distance. With the aim to estimate the strength of SGS, pairwise kinship coefficients were regressed on the logarithm of pairwise distances between individuals. Permutation tests (permuting the spatial position of the individuals 10,000 times) were used to assess the significance of the regression slope. The strength of the SGS was estimated as $Sp = -b/(F_{ij(1)}-1)$ (Vekemans & Hardy 2004) where b is the regression slope and $F_{ij(1)}$ is the averaged kinship coefficient of individuals in the first distance class (<20m). The analysis of within-population SGS was performed for each population and across all populations from Africa and the Neotropics, respectively. To evaluate differences in SGS between continents, we compared jackknife 95% confidence intervals (CIs) of b between continents and populations (a steeper slope indicating stronger SGS, Vekemans & Hardy 2004).

The shape of the regression curve can give hints on the relative contribution of pollen and seed dispersal to overall gene dispersal, conditional on a set of assumptions (Heuertz *et al.* 2003). Assuming Gaussian dispersal kernels, a concave curve shape indicates that the short distance component of gene dispersal is restricted, while a convex shaped curve indicates no restriction of this component (Heuertz *et al.* 2003; Vekemans & Hardy 2004). In *S. globulifera* pollen flow is typically short-ranging (Degen *et al.* 2004) while seed dispersal mediated by bats can be far ranging (Nathan 2006; Dick *et al.* 2008). Therefore we hypothesize that seed dispersal may be the long distance component of gene dispersal. To infer the shape of the regression curves of the kinship-distance-plot at short spatial distance, the residuals of pairwise kinship coefficients were fitted to a polynomial function of third power of the form $f(r) = a + b\ln(r) + c[\ln(r)]2 + d[\ln(r)]^3$, where r is the inter-individual geographic distance using R version 2.15.0 (R Development Core Team 2008) following Vekemans & Hardy (2004). The shape of this polynomial function was inspected at the level of the first distance class using its second derivative k which depends on the shape parameter d. Positive values of k indicate concavity while negative values of k indicate convexity (Heuertz *et al.* 2003; Vekemans & Hardy 2004).

Furthermore, a comparison of the cumulative frequency distribution (CFD) of F_{IS}-*intra* (estimates of intra-individual F_{ij}) and $F_{ij(1)}$ was used to estimate the contribution of biparental inbreeding to total inbreeding. A one-sided non-

parametric Kolmogorov-Smirnov test with alternative hypothesis that the CFD of F_{IS}-*intra* lay below that of $F_{ij(1)}$ was performed in R version 2.15.0. A significant test result points out that F_{IS}-*intra* is larger than $F_{ij(1)}$ and that therefore inbreeding is mainly a result of selfing. If both statistics have similar CFDs, the contribution of biparental inbreeding to total inbreeding is enhanced (Barbará *et al.* 2008).

Results

Heterozygosity estimates were quite similar in American and African populations with the exception of Ituberá, where slightly lower gene diversity was found. Significant inbreeding indices were detected in all populations (lowest in Ituberá, F_{IS}=0.094*** and highest in Mbikiliki, F_{IS}=0.178***, Table 2).

Significant SGS was observed in all populations and the strength of SGS ranged from *Sp*=0.009 in Ituberá to *Sp*=0.027 in NkongMekak (Table 3). The SGS within stands across the Neotropics was *Sp*= 0.013***, while SGS within stands across Africa was *Sp*= 0.032***. Jackknife 95% confidence intervals of the regression slopes of all populations overlapped (Figure 3). The fit of the polynomial function of third power revealed significant values for the shape parameter *d* in Mbikiliki and Paracou. Curve shapes were convex in Mbikiliki, Paracou and NkongMekak (although not significant in the latter case). The other populations had concave shapes, although not significant (Table 3, Figure 2). Furthermore, selfing contributed strongly to total inbreeding in all populations except Ituberá where the CFDs of F_{IS}-*intra* and of $F_{ij(1)}$ were similar (Table 3).

Discussion

Genetic diversity and inbreeding estimates were similar in populations from America and Africa, indicating connatural life history traits despite long-term geographic separation. These results were in agreement with previous studies from both continents that reported similar levels of genetic diversity for nuclear microsatellites and chloroplast haplotypes (Dick & Heuertz 2008; Budde *et al.* 2013). However, slight differences in SGS were found in African and American *S. globulifera* populations, indicating a possible trend of stronger SGS in Africa, although jackknife CIs of the SGS regression slopes overlapped in all populations. Furthermore, both African populations showed a convex polynomial regression slope, pointing to less leptokurtic gene dispersal in these populations and in Paracou than in the other Neotropical populations. The *S. globulifera* population in Paracou is characterized by the presence of two ecotypes (ca. 90 and 10% frequency), that are genetically very closely related and do not demonstrate significant allele

frequency differences (Degen *et al.* 2004, C. Scotti-Saintagne, pers. com.). The SGS pattern in this stand is therefore likely due to a complex combination of non-random mating between the ecotypes and dispersal limitation of each ecotype. The Neotropical populations from BCI and Ituberá displayed a concave shape of the kinship-ln(distance) plot, indicating a limitation of the short distance component of gene dispersal in relation to overall gene dispersal (Heuertz *et al.* 2003; Vekemans & Hardy 2004). If the short distance component of gene dispersal reflected pollen-mediated gene flow, this would indicate a restriction of gene dispersal via pollen in American *S. globulifera* populations, and therefore, distinct pollination behavior of hummingbirds and perching birds (in America) vs. sunbirds (in Africa). If confirmed, this result would be surprising as we a priori considered these pollination vectors to belong to the same functional type (due to similar morphological features and behavior, see also Armbruster 2006). The comparison of the CFDs of F_{IS}-*intra* and $F_{ij(1)}$ indicated that selfing was the main factor responsible for total inbreeding in four of the five study populations. Only in Ituberá, where also a lower SGS was detected, biparental inbreeding contributed to the increased inbreeding coefficient. This finding pointed to similar degrees of selfing in populations from both continents, which would be coherent with our a priori expectation of pollinators belonging to the same functional type.

In contrast, the functional types of seed dispersers differ in populations from both continents. So far, no bat mediated seed dispersal has been reported in African populations (see Table 1). As seed dispersal by bats is potentially far ranging (Nathan 2006; Dick *et al.* 2008), these small nocturnal mammals could be the key to the difference in the shape of kinship-ln(distance) curves in Africa and America. Even in Paracou where our results indicated less leptokurtic gene dispersal (based on three microsatellites), a previous study on the mating system and SGS reported limited pollen dispersal relative to bat mediated seed dispersal (Degen *et al.* 2004) coinciding with the general trend observed for Neotropical populations. As *S. globulifera* is assumed to have evolved in the Palaeotropics and later spread to the Neotropics (Dick *et al.* 2003) its fruits could be an example for exaptation to bat dispersal (Forget *et al.* 2007). In fact, they are not pendulant which is often the case for “bat fruits”, possibly pointing to an evolution in the presence of other vector species.

Although the strength of SGS varied between populations, *Sp* values found in *S. globulifera* populations were typical for species with similar life-history traits. Low populations density, bird pollination and animal mediated seed dispersal in predominantly outcrossing species produce similar SGS patterns (Vekemans

& Hardy 2004; Hardy *et al.* 2006). However, especially the shape of the dispersal curve and the relative contribution of pollen and seed dispersal to overall gene dispersal might allow a more precise evaluation of specific life-history traits on the built up of SGS.

So far our results remain speculative. However, this study is still ongoing and in the future more populations from Africa, which have recently been sampled, will be included. More replicates, also from other regions in Africa, should improve the comparison of populations from both continents and enable us to draw improved conclusions on SGS patterns. Furthermore, it is foreseen to include *psbA-trnH* chloroplast haplotypes. If polymorphic enough, these maternally inherited markers will enable us to estimate the contribution of gene flow mediated by seeds in relation to overall gene dispersal in each population. As a consequence, pollen- and seed-mediated gene flow are expected to be differentiable, so that the distinct shapes of kinship-ln(distance) curves will become interpretable in terms of specific pollen- vs. seed dispersal distances.

Acknowledgements

We dedicate this publication to Elisabeth K. V. Kalko, the PhD supervisor of K. Heer who passed away during field work in Tanzania in 2011. We are very grateful to Dr. Bernd Degen (BFH, Institute for Forest Genetics and Forest Tree Breeding, Grosshansdorf, Germany) for providing data from the Paracou population. Funding for K. B. Budde was provided by a PhD scholarship from the Spanish Ministry of Science and Innovation (MICINN) (FPI: BES-2009 015443), and for K. Heer by a PhD scholarship of the German National Academic Foundation. Fieldwork in Brazil was financed by Michelin. MH is a 'Ramón y Cajal' researcher of the MICINN and acknowledges funding from the Spanish Ministry of Economy and Competitiveness (CGL2012-40129-C02-02/ AFFLORA).

References

Abdul-Salim K (2002) Systematics and biology of *Symphonia globulifera* L.f. (Clusiaceae). PhD thesis, Harvard University, Boston.

Aldrich PR, Hamrick JL (1998) Reproductive Dominance of Pasture Trees in a Fragmented Tropical Forest Mosaic. *Science*, **281**, 103–105.

Aldrich PR, Hamrick JL, Chavarriaga P, Kochert G (1998) Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology*, **7**, 933–44.

Armbruster W (2006) Evolutionary and ecological perspectives on specialization: from the Arctic to the tropics. In: *Plant-pollinator interactions: from specialization to generalization.* (eds Waser N, Ollerton J), pp. 260–282. University of Chicago Press, Chicago.

Baraloto C, Morneau F, Bonal D, Blanc L, Ferry B (2007) Seasonal water stress tolerance and habitat association. *Ecology*, **88**, 478–489.

Barbara T, Lexer C, Martinelli G *et al.* (2008) Within-population spatial genetic structure in four naturally fragmented species of a neotropical inselberg radiation , *Alcantarea imperialis* , *A . geniculata* , *A . glaziouana* and *A . regina*. *Heredity*, **101**, 285–296.

Bawa K, Perry D, Coville R, Grayum M (1985) Reproductive biology of tropical lowland rain forest trees. II. Pollination Systems. *American Journal of Botany*, **72**, 346–356.

Bittrich V, Amaral MCE (1996) Pollination biology of *Symphonia globulifera* (Clusiaceae). *Plant Systematics and Evolution*, **200**, 101–110.

Budde KB, González-Martínez SC, Hardy OJ, Heuertz M (2013) The ancient tropical rainforest tree *Symphonia globulifera* L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa. *Heredity*, 1–11.

Charles-Dominique P (1986) Inter-relations between frugivorous vertebrates and pioneer plants: Cecropia, birds and bats in French Guyana. In: *Frugivores and seed dispersal* (eds Estrada A, Fleming TH), pp. 119–135. Dr W. Junk Publishers, Dordrecht.

Clark C, Poulsen J (2001) The role of arboreal seed dispersal groups on the seed rain of a lowland tropical forest. *Biotropica*, **33**, 606–620.

Degen B, Bandou E, Caron H (2004) Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity*, **93**, 585–91.

Dick CW, Abdul-Salim K, Bermingham E (2003) Diversification of a widespread tropical rain forest tree. *The American Naturalist*, **162**, 691–703.

Dick CW, Hardy OJ, Jones FA, Petit RJ (2008) Spatial Scales of Pollen and Seed-Mediated Gene Flow in Tropical Rain Forest Trees. *Tropical Plant Biology*, **1**, 20–33.

Dick CW, Heuertz M (2008) The complex biogeographic history of a widespread tropical tree species. *Evolution*, **62**, 2760–74.

Epperson BK (1990) Spatial autocorrelation of genotypes under directional selection. *Genetics*, **124**, 757–771.

Faegri K, van der Pijl L (1965) *The principles of pollination ecology*. Pergamon Press, Oxford.

Forget P, Dennis A, Mazer S *et al.* (2007) Seed allometry and disperser assemblages in tropical rainforests: A comparison of four floras on different continents. In: *Seed dispersal. Theory and its Application in a Changing World* (eds Dennis A, Schupp EW, Green R, Westcott D), pp. 5–37. CAB International, Oxfordshire.

Gautier-Hion A, Duplantier J-M, Quris R *et al.* (1985) Fruit characters as a basis of fruit choice and seed dispersal in a tropical forest vertebrate community. *Oecologia*, **65**, 324–337.

Gill GE, Fowler RT, Mori SA (1998) Pollination biology of *Symphonia globulifera* (Clusiaceae) in Central French Guiana. *Biotropica*, **30**, 139–144.

Grant K, Grant V (1968) *Hummingbirds and their flowers*. Columbia University Press, New York.

Hardy OJ, Maggia L, Bandou E *et al.* (2006) Fine-scale genetic structure and gene dispersal inferences in 10 neotropical tree species. *Molecular Ecology*, **15**, 559–71.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Henry O, Feer F, Sabatier D (2006) Diet of the Lowland Tapir (*Tapirus terrestris* L.) in French Guiana. *Biotropica*, **32**, 364–368.

Herrera C (1985) Determinants of plant-animal coevolution: the case of mutualistic dispersal of seeds by vertebrates. *Oikos*, **44**, 132–141.

Heuertz M, Vekemans X, Hausman JF, Palada M, Hardy OJ (2003) Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology*, **12**, 2483–2495.

Heywood J (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics*, **22**, 335–355.

Howe HF, Smallwood J (1982) Ecology of seed dispersal. *Annual Review of Ecology and Systematics*, **13**, 201–228.

Jordano P, Garcia C, Godoy JA, Garcia-Castano J (2007) Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences*, **104**, 3278–3282.

Levin D, Kerster H (1968) Local gene dispersal in *Phlox*. *Evolution*, **22**, 130–139.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Mitchell RJ, Irwin RE, Flanagan RJ, Karron JD (2009) Ecology and evolution of plant – pollinator interactions. *Annals of Botany*, **103**, 1355–1363.

Myers N, Mittermeier R a, Mittermeier CG, da Fonseca G a, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nathan R (2006) Long-distance dispersal of plants. *Science*, **313**, 786–788.

Oyen L (2005) *Symphonia globulifera* L.f. (Internet) Record from Protabase. In: Louppe D, Oteng-Amoako AA, Brink M (eds.) PROTA (Plant Resources of Tropical Africa/ Ressources végeétales de l'Afrique tropicale). *http//database.prota.org/search.htm*, Wageningen, Netherlands.

Pascarella JB (1992) Notes on flowering phenology, nectar robbing and pollination of *Symphonia globulifera* l.f. (Clusiaceae) in a lowland rain forest in Costa Rica. *Brenesia*, **38**, 83–86.

Van der Pijl L (1972) *Principles of dispersal in higher plants*. Springer-Verlag, Berlin.

Poulsen JR, Clark CJ, Smith TB (2001) Seed dispersal by a diurnal primate community in the Dja Reserve, Cameroon. *Journal of Tropical Ecology*, **17**, 787–808.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *The Journal of Heredity*, **86**, 248–249.

Regal P (1977) Ecology and evolution of flowering plant dominance. *Science*, **196**, 622–629.

Rousset F (2004) *Genetic structure and selection in subdivided populations*. Princeton University Press, Princeton.

Schupp EW (1993) Quantity, quality and the effectiveness of seed dispersal by animals. *Vegetatio*, **107**, 15–29.

Da Silva Carneiro F, Degen B, Kanashiro M *et al.* (2009) Forest Ecology and Management High levels of pollen dispersal detected through paternity analysis from a continuous Symphonia globulifera population in the Brazilian Amazon. *Forest Ecology and Management*, **258**, 1260–1266.

Turner I (2001) *The ecology of trees in the tropical rain forest*. Cambridge University Press, Cambridge.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Vinson CC, Amaral AC, Sampaio I, Ciampi AY (2005) Characterization and isolation of DNA microsatellite primers for the tropical tree, Symphonia globulifera Linn. f. *Molecular Ecology Notes*, **5**, 202–204.

Willmer P (2011) *Pollination and floral ecology*. Princeton University Press, Princeton.

Wolowski M, Saad CF, Ashman T, Freitas L (2013) Predominance of self-compatibility in hummingbird-pollinated plants in the Neotropics. *Naturwissenschaften*, **100**, 69–79.

Wright S (1931) Evolution in Mendelian Populations. *Genetics*, **16**, 97–159.

Author contributions

KBB conducted sampling in Mbikiliki and NkongMekak, performed SSR genotyping, data analyses and wrote the manuscript. KH did the sample collection in Ituberá and contributed to SSR genotyping. CD, SCGM and MH designed the study. MH conducted sampling in Africa and contributed to data analyses. All co-authors critically read and provided comments to the MS.

Tables

Table 1 Animals that have been reported as seed dispersers and pollinators of *Symphonia globulifera* in populations from Africa and the Neotropics. p, pollinator; sd, seed disperser.

Visitors	Function	References
Africa		
sunbirds	p	(Oyen 2005)
monkeys	sd	(Gautier-Hion *et al.* 1985; Clark & Poulsen 2001; Poulsen *et al.* 2001)
ruminants	sd	(Gautier-Hion *et al.* 1985; Forget *et al.* 2007)
hornbills	sd	(Gautier-Hion *et al.* 1985)
Neotropics		
hummingbirds	p	(Bawa *et al.* 1985; Bittrich & Amaral 1996; Degen *et al.* 2004)
perching birds	p	(Bittrich & Amaral 1996; Gill *et al.* 1998)
lepidoptera	p	(Pascarella 1992; Bittrich & Amaral 1996)
bees	p	(Bittrich & Amaral 1996)
bats	sd	(Charles-Dominique 1986; Aldrich & Hamrick 1998; Degen *et al.* 2004)
scatter-hoarding rodents	sd	(Hardy *et al.* 2006)
nocturnal arboreal mammals	sd	(Hardy *et al.* 2006)
tapirs	sd	(Henry *et al.* 2006)

Table 2 Estimates of population genetic parameters of the study populations of *Symphonia globulifera*. *N*, number of samples; loci, number of loci genotyped; H_E, expected heterozygosity; F_{IS}, fixation index and significance for F_{IS}-observed >F_{IS}-expected: ***, $P < 0.001$)

Population	*N*	*loci*	H_E	F_{IS}
NkongMekak	68	5	0.793	0.157***
Mbikiliki	83	5	0.750	0.178***
BCI	148	5	0.833	0.147***
Paracou	148	3	0.881	0.172***
Ituberá	83	5	0.594	0.094***

Table 3 Estimates of mating system and fine scale spatial genetic structure in each sample location of *Symphonia globulifera*, respectively. F_{IS}-*intra*, intra individual inbreeding coefficient; $F_{ij(1)}$, average kinship coefficient of the first distance class; $P(F_{IS}\text{-}intra > F_{ij(1)})$, significance of a one-sided Kolmogorov–Smirnov test with alternative hypothesis that the cumulated frequency distribution of the intra individual inbreeding coefficient (F_{IS}-*intra*) lies under that of the corresponding pairwise kinship coefficients of the first distance class; *Sp*, intensity of SGS; significance of SGS as indicated by *P*-values of the regression slope (***, $P<0.001$); *d*, shape fit parameter of the polynomial function of third power (n.s., $P>0.1$; **.**, $P<0.1$;*, **, $P<0.01$; ***, $P<0.001$); *k*, indicator of curve shape as second derivate of polynomial function ($k>0$ concave; $k<0$ convex)

Population	F_{IS}-*intra*	$F_{ij(1)}$	$P(F_{IS}\text{-}intra > F_{ij(1)})$	*Sp*	*d*	*k*
NkongMekak	0.1551	0.0612	0.0003	0.0156***	0.0010 n.s.	-0.0139
Mbikiliki	0.1831	0.0675	0.0170	0.0266***	0.0033 ***	-0.0460
BCI	0.1441	0.0434	0.0000	0.0158***	-0.0002 n.s.	0.0018
Paracou	0.1709	0.0619	0.0000	0.0094***	0.0015 **	-0.0164
Ituberá	0.0938	0.0960	0.1084	0.0086***	-0.0012 .	0.0300

Figures

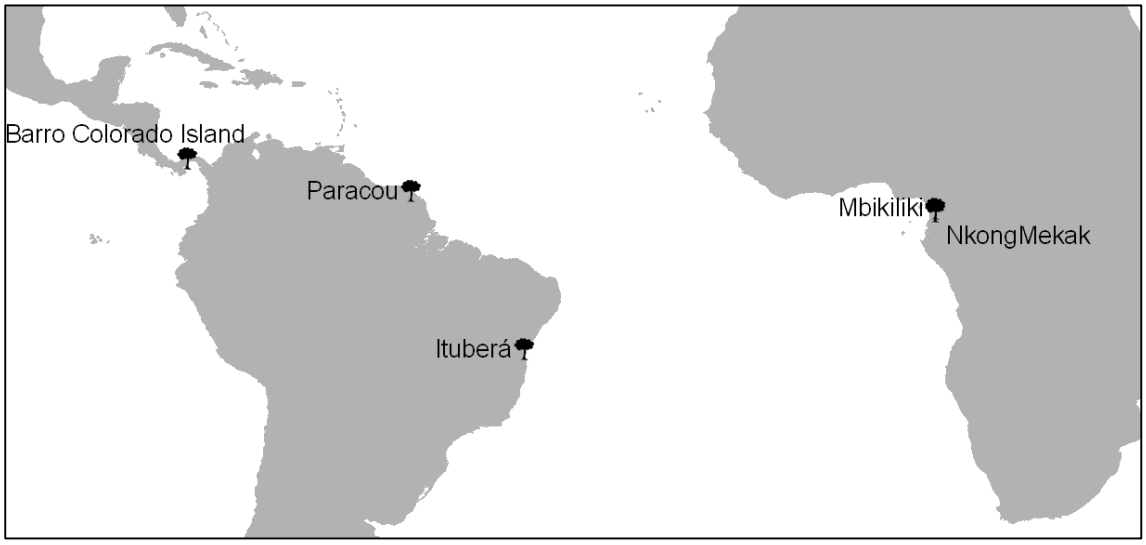

Figure 1 Sample locations of *Symphonia globulifera* in Africa and America.

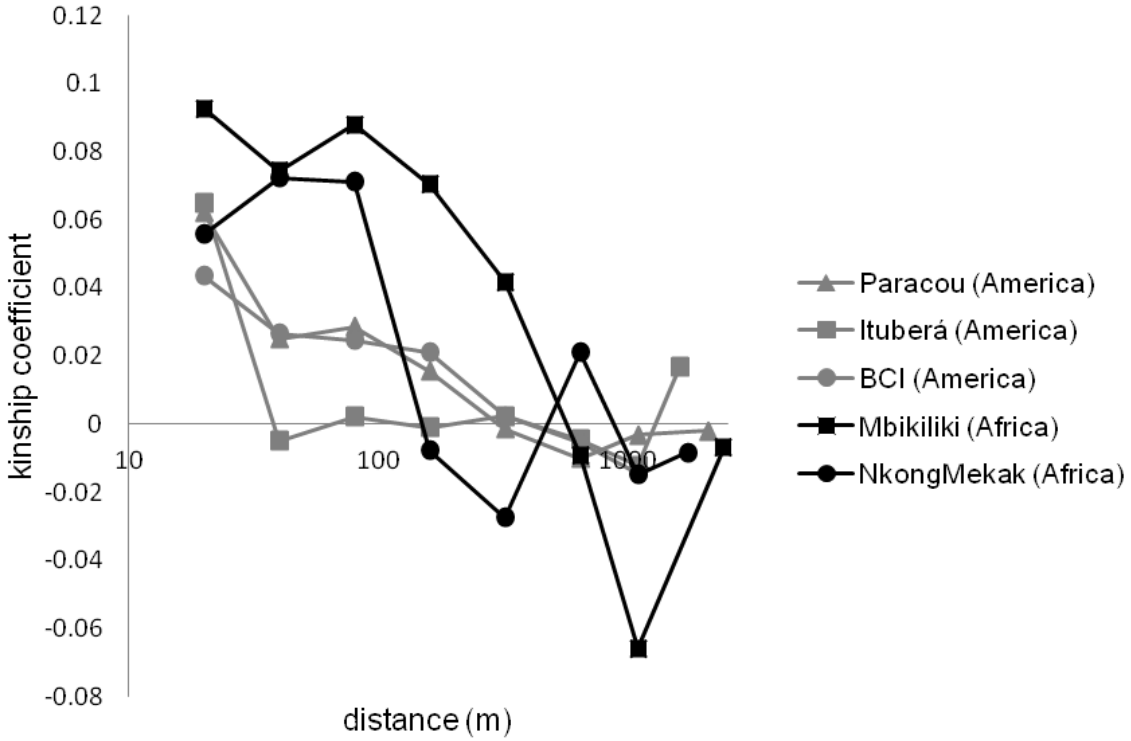

Figure 2 Average kinship coefficient plotted against the logarithm of the mean geographical distance between individuals in each distance class for each *Symphonia globulifera* population, respectively. Black lines indicate African populations and grey lines Neotropical populations.

.

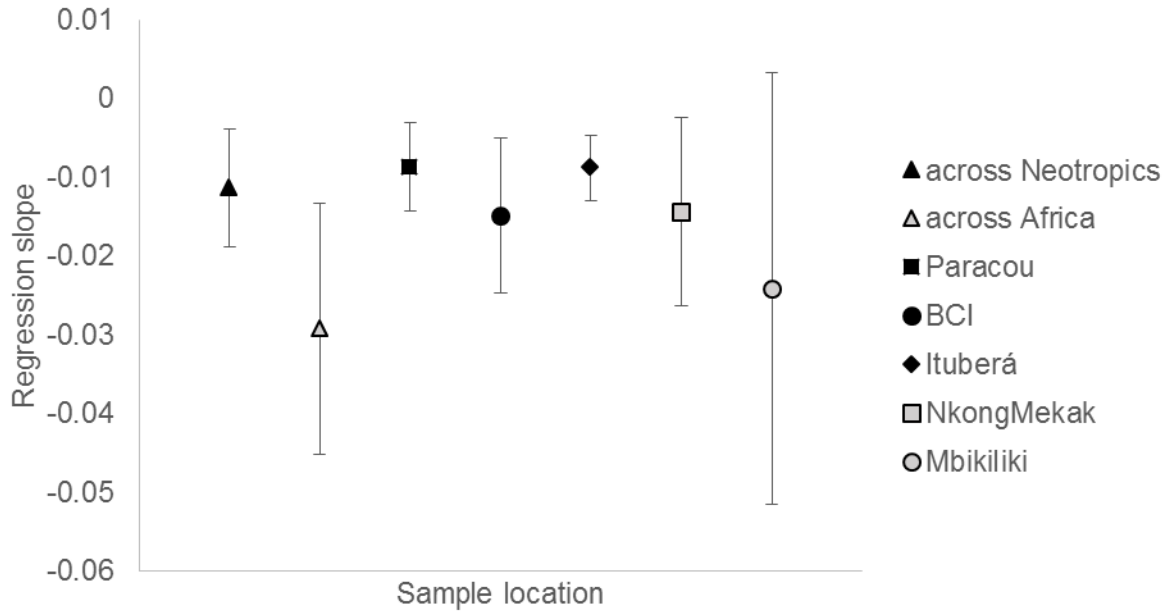

Figure 3 Mean jackknife regression slopes and 95% confidence intervals of within stand SGS for Neotropical (black symbols) and African (grey symbols) populations of *Symphonia globulifera*. Analyses were performed across populations from the same continent ("across Neotropics" and "across Africa") and for each population, respectively.

Chapter 3

Effects of fire regime on the population genetics of natural pine stands

This chapter is based upon the manuscript:

Budde KB, González-Martínez SC, Navascués M, Burgarella C, Herández-Serrano A, Unger GM, Lorenzo Z, Zabal-Aguirre M, Vendramin GG, Pausas JG, Verdú M, Heuertz M. Effects of fire regime on the population genetics of natural pine stands. *Submitted to Molecular Ecology.*

Effects of fire regime on the population genetics of natural pine stands

Katharina B. Budde[1], Santiago C. González-Martínez[1], Miguel Navascués[2], Concetta Burgarella[3], Ana Hernández-Serrano[4], Gregor M. Unger[1], Zaida Lorenzo[1], Mario Zabal-Aguirre[1,5], Giovanni G. Vendramin[6], Miguel Verdú[4], Juli G. Pausas[4], Myriam Heuertz[1,7]

[1]INIA Forest Research Centre, Department of Forest Ecology and Genetics, Carretera A Coruña km 7.5, 28040 Madrid, Spain

[2]INRA, UMR1062 CBGP, F-34988 Montferrier-sur-Lez, France

[3]Amélioration génétique et adaptation des plantes, Institut National de la Recherche Agronomique UMR AGAP, 2, Place P. Viala, 34060 Montpellier Cedex 1, France

[4]Centro de Investigaciones sobre Desertificación (CSIC/UV/GV), 46113 Moncada (Valencia), Spain.

[5]Instituto de Ciencias Agrarias (CSIC/ICA), Grupo de Interacciones Beneficiosas Planta-Microrganismo, Calle Serrano 115-bis, 28006 Madrid, Spain

[6]Plant Genetics Institute, National Research Council, Via Madonna del Piano 10, 50019 Sesto Fiorentino (Florence), Italy

[7]University of Fribourg, Department of Biology, Unit of Ecology and Evolution, Chemin du Musée 10, 1700 Fribourg, Switzerland

Abstract

The recurrence of wildfires is predicted to increase worldwide due to climate change, resulting in severe impacts on biodiversity and ecosystem functioning. We used simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers to examine the effects of fire regime on genetic diversity, demographic history and fine-scale spatial genetic structure (SGS) of *Pinus pinaster* and *P. halepensis*, two conifers with similar adaptations to fire in the eastern Iberian Peninsula. Stands growing under high (HiFi) or low (LoFi) frequency of crown fires had similar levels of genetic diversity and similar demographic history, with bottlenecks detected in all stands in both species. HiFi populations were not genetically depleted, suggesting that adaptations such as a diverse canopy seed bank due to serotinous cones, an early age of first flowering and high gene flow buffer against possible reductions of genetic diversity. Significantly stronger SGS at SNPs in HiFi than LoFi stands of *P. halepensis* suggested fire-related altered dispersal and/or micro-environmental selection in this fire-sensitive "seeder" species. In contrast, SGS at SNP markers was unrelated to fire regime in *P. pinaster*. This could be a consequence of more pronounced fire-resistance in this species enabling some adults to survive fire, hence causing a lower dependence on post-fire regeneration from seeds. Our results highlight that the impact of fire differs in species with similar life-history traits. Therefore, species-specific studies are needed to understand the role of wildfires for the evolution of future forests.

Keywords: Spatial genetic structure, SSRs, SNPs, demographic history, fire ecology

Introduction

Wildfires influence plant species composition in natural environments and play an important role in the distribution of biomes worldwide (Pausas & Keeley 2009). Changes in fire regimes can have complex and severe impacts on biodiversity and ecosystem functioning (Lavorel *et al.* 2006). In Mediterranean ecosystems, wildfires are a common feature (Keeley *et al.* 2012), but the predicted rise in temperatures and reduction in precipitation (De Castro *et al.* 2005) are expected to increase fire recurrence in the coming decades (e.g., Piñol *et al.* 1998; Pausas 2004). Anthropogenic land use change may aggravate the situation because it has caused fire regimes to be more drought-driven than limited by fuel availability (low productivity, Pausas & Fernández-Muñoz 2012). Therefore, studies aiming to understand the effects of fire events on

biodiversity are especially important in the face of global change.

Wildfires are major selection drivers shaping adaptive plant traits (Keeley *et al.* 2011). Plant species are adapted to specific fire regimes, which exert selection pressure through their features, including 1) the amount of fuel consumption and fire spread patterns, 2) the intensity and severity of the fires, 3) their frequency, 4) the burn patch size and the distribution of patches, and 5) the seasonality of fire (Keeley *et al.* 2012). Plant adaptive strategies to fire regimes vary substantially in different plant groups and geographic regions (Pausas *et al.* 2004, Keeley *et al.* 2011). In Mediterranean-type vegetation, three types of adaptive strategies to fire regime can be distinguished: 1) individuals persist through resprouting (resprouters), 2) individuals die and populations persist through seeding (seeders), and 3) individuals are unaffected by fire (fire-escaping). Mediterranean pines respond to fire through variable levels of seeding (e.g., serotiny) and fire-escaping (e.g. having a thick bark, large and protected buds or self-pruning of lower branches) (Tapias, *et al.* 2004, He et al. 2012, Hernandez-Serrano et al. 2013).

The evolutionary impact of fire regime on the within-species genetic constitution is determined by the interaction of fire regime with the species' adaptive strategy and other life history traits, as well as by extrinsic biotic and abiotic factors. Fire regime can have consequences for 1) the species' level of genetic diversity, 2) fine-scale spatial genetic structure (SGS) and 3) may differentially affect neutral and adaptive diversity. First, mortality through fire and high fire frequency can produce population bottlenecks, i.e., reduce the effective population size and genetic diversity (Whelan 1995). This effect can be aggravated as the combined effects of drought, herbivory or fungal diseases increase fire-related mortality (Turner *et al.* 1999; Ayres & Lombardero 2000). Second, SGS, the non-random distribution of related individuals or gene copies in a population, is largely shaped by dispersal patterns, which will be altered when seed traits and dispersal capacity are under fire-mediated selection (Saracino *et al.* 1997). At the local level, post-fire seedling establishment and survival is influenced by shrub cover, the extent of erosion, slope, soil depth and ectomycorrizal distribution patterns (Buscardo *et al.* 2011), as well as site aspect and the amount of dead branches on the floor (Pausas *et al.* 2004). We expect that these interactions may result in patchy patterns of relatedness and heterogeneous selection, which should affect SGS (Epperson 1995; Linhart & Grant 1996) and could leave distinct signatures at neutral and adaptive diversity.

Empirical studies indicate that in seeder species, seed banks preserve genetic diversity and buffer against demographic fluctuations (Templeton & Levin 1979; Barrett *et al.* 2005; Ayre *et al.* 2010). In agreement with this observation, levels of genetic diversity in the species *Pinus halepensis*, a typical seeder, were similar in unburnt stands and in stands regenerated after fire on Mt. Carmel in Israel (Schiller *et al*. 1997). Seeders are nevertheless vulnerable to shifted fire regimes because a too short time span between fires may prevent the establishment of new seed banks (Bradstock *et al.* 1996; Odion & Tyler 2002). Comparison of seeders and resprouters indicated higher genetic diversity in seeder than resprouter populations of South African *Erica coccinea* (Segarra-Moragues & Ojeda 2010), whereas in Mediterranean *Erica*, a resprouter species held higher genetic diversity than a seeder (Segarra-Moragues *et al.* 2012).

Empirical studies that investigated SGS in relation with fire regime found variable patterns. SGS was observed in post-fire stands across three cohorts of *Populus trichocarpa*, which was partially explained by resprouting (clonality, Namroud *et al.* 2005). In *Nothofagus dombeyi,* which does not possess any particular adaptation to fire, young post-fire cohorts showed significant but weaker SGS than mature stands (Premoli & Kitzberger 2005). In *Pinus halepensis*, spatial autocorrelation for serotiny was believed to reflect an aggregated pattern of post-fire seed dispersal (Hernández-Serrano *et al.* 2013), in agreement with another study that found clumped seedling establishment near burnt adults of this species (Ne'eman & Izhaki 1998). In *Persoonia mollis*, a fire adapted species with a soil seed bank, seedlings under dead adults did not reflect simple seed shadows, but SGS in seedlings was increased after a fire event (Ayre *et al.* 2009).

In the present study, we investigated the effect of the fire regime on genetic diversity, demography and SGS in natural stands of two Mediterranean pine species, *Pinus halepensis* and *P. pinaster*. Pine species are appropriate models to study the evolutionary impact of fire because notable fire adaptations in several species have facilitated the wide distribution of the genus (Agee 1998; Schwilk & Ackerly 2001; He *et al.* 2012). In our study system, the long fire history in the Mediterranean basin has shaped a mosaic of populations adapted to the specific local fire regimes (Hernández-Serrano *et al.* 2013; Tapias *et al.* 2004). *Pinus halepensis* and *P. pinaster* are considered seeder species, that is, adults are killed by fire and viable seeds are preserved in canopy seed banks in so-called serotinous cones until seed release is triggered by high temperatures associated to fires (Tapias

et al. 2004; Lamont *et al.* 1991). In areas with frequent crown fires, serotiny level is higher in *Pinus halepensis* than in *P. pinaster* (Hernández-Serrano *et al.* 2013). In areas where crown-fires are infrequent, (e.g. mostly surface fires), *P. pinaster* exhibits more pronounced fire escaping traits than *P. halepensis* and adult trees are often able to survive fire (Fernandes *et al.* 2008). We examined genetic diversity and SGS at microsatellites (simple sequence repeats, SSRs) and single nucleotide polymorphisms (SNPs) in natural stands of both species under either high- or low crown fire frequency (HiFi vs. LoFi) in the eastern Iberian peninsula, hypothesizing 1) that HiFi populations may show stronger signs of demographic bottlenecks due to the negative impact of frequent fires on seed bank diversity, 2) that HiFi populations may show stronger SGS, in agreement with spatial autocorrelation observed for serotiny in *P. halepensis* (Hernández-Serrano *et al.* 2013), 3) that SNPs should display a stronger SGS signal than SSRs in HiFi populations if the studied SNPs were under fire-related micro-environmental selection and 4) that fire effects would be stronger in *P. halepensis* than in *P. pinaster* because its trait variation indicates that population turnover might *a priori* be more dependent on fire. To our knowledge, this is the first study addressing the long-term effects of fire regimes on the genetic constitution of natural pine populations.

Material and Methods

Study species

Pinus halepensis (Aleppo pine) and *P. pinaster* (cluster pine) are long-lived, monoecious conifer tree species native to the Mediterranean basin where ecological disturbances such as forest fires and human use have generated a mosaic of continuous and fragmented populations of both species (Pausas *et al*. 2004). Extensive wind mediated pollen and seed dispersal has been documented in both species (González-Martínez *et al.* 2006; de-Lucas *et al.* 2008; Steinitz *et al.* 2011). SGS in natural *P. halepensis* and *P. pinaster* stands was found to be weak or non-significant using SSR markers (González-Martínez *et al.* 2002; Troupin *et al.* 2006; de-Lucas *et al.* 2009). In regions with frequent stand-replacing crown fires, both species show seeder traits including an early age of first flowering, retention of dead branches on the stem which increases the likelihood of crown fires, and high levels of serotiny, i.e., a well-developed canopy seed bank crucial for post-fire regeneration (Ne'eman *et al.* 2004; Tapias *et al.* 2004; Fernandes & Rigolot 2007; Schwilk & Ackerly 2001). In regions with less frequent crown fires, recruitment is not

dependent on serotinous cones. *Pinus halepensis* is a typical fire-embracer (fire-sensitive) species which persists on drier sites than *P. pinaster*, which has a thinner bark and an earlier age of first flowering, smaller and more numerous cones and shows typically high serotiny (Fernandes *et al.* 2008; Ne'eman *et al.* 2004; Pausas *et al*. 2004; Tapias *et al*. 2004). *Pinus halepensis* also has a high colonization capacity from non-serotinous cones (Goubitz *et al.* 2004). *Pinus pinaster* is more variable in the expression of fire-related traits: some populations are adapted to endure fire events displaying thick bark, some resistance to crown scorch, low levels of serotiny and delayed reproduction (Fernandes & Rigolot 2007; Tapias *et al*. 2004). Serotinous cones of *P. pinaster* open readily at lower temperatures (Tapias *et al.* 2001); however, during fire-free time intervals, serotinous cones remain longer on *P. pinaster* trees (Hernández-Serrano *et al*. 2013) and maintain seed viability longer (30 years) than in *P. halepensis* (20 years, Catalán 1991; Tapias *et al.* 2004).

Study sites and sampling

A local scale was chosen for this study to enable comparison of fire regimes while controlling for demographic effects due to broad scale processes (e.g. glacial history). The study region is located in the east of the Iberian Peninsula and harbours a single gene pool in each species (see Supporting Information S1). Contrasting fire regimes affect populations in close vicinity in the study region (Pausas 2004). The western (inland) sub-region features mountainous Mediterranean vegetation (>800 m asl) with sub-humid climate where low intensity understory fires occur but where crown fires are rare; a fire regime labelled LoFi hereafter (i.e., low frequency of crown fires). Recruitment in these stands depends generally on gaps unrelated to fire and tree generations normally overlap. The eastern (coastal) sub-region (<800 m asl) has a warmer and drier climate where crown fires are common; labelled HiFi hereafter (i.e., high frequency of crown fires). Recruitment typically occurs after fire events resulting in more even-aged stands (Pausas *et al*. 2004). Although no long-term fire statistics are available for the study stands, recent fire history reveals that more than 50% of the study area at < 800 m asl (HiFi conditions) burned at least once during the 1978–2001 period, while the proportion was only about 15% for >800 m asl (LoFi conditions; Pausas 2004; Abdel Malak & Pausas 2006; Pausas & Fernández-Muñoz 2012).

Needles of 63-67 adult individuals were collected in 2009 and 2010 in three stands of *P. halepensis* (n = 197 trees) and in three stands of *P. pinaster* (n = 199 trees), choosing stand pairs (one stand per

species) in three locations: Serra Calderona and Eslida under HiFi regime, and Sinarcas under LoFi regime (geographic coordinates are reported in Table S2.1 in Supporting Information S2). Trees of similar diameter at breast height (DBH) were sampled, avoiding sub-canopy trees. Inter-tree sampling distances were comprised between 10 and ca. 1000 m (see also Hernández-Serrano *et al.* 2013, who assessed serotiny in the same stands), increasing sampling distances by an order of magnitude compared to previous studies to reduce the risk of overestimating SGS (Troupin *et al.* 2006; de-Lucas *et al.* 2009). All stands were mainly mono-specific with similar populations densities, of 312-649 trees/ha in *P. pinaster* and 285-441 trees/ha in *P. halepensis*. Needles were dried with silica gel and DNA was extracted using the Invisorb DNA Plant HTS 96 Kit (Invitek, Berlin, Germany). After first results indicated significant SGS with SNPs in HiFi stands of *P. halepensis*, we included two additional LoFi (Montán and Titaguas) and three HiFi stands (Alzira, Serra d'Irta, and Cabanes, Table S2.1 in Supporting Information) in this species to test the hypothesis of increased SGS at SNPs due to high fire recurrence. In these stands, 29-37 individuals were sampled at minimum distances of 20 m (see Hernández-Serrano *et al.* 2013 for details).

Genetic markers

Nuclear microsatellites

All *Pinus pinaster* samples were genotyped at 11 nuclear microsatellites (SSRs). The SSR primer sequences were obtained from Mariette *et al.* (2001): ITPH4516 and FRPP94; Chagné *et al.* (2004): rpTest11, Ctg4363, NZPR1078 and NZPR413; Guevara *et al.* (2005): A6F03; Steinitz *et al.* (2011): pEST2669; and F. Sebastiani and G.G. Vendramin (pers. com., June 2011): epi3, epi5 and gPp14. *Pinus halepensis* individuals were genotyped at nine SSR loci. Primer sequences were obtained from Mariette *et al.* (2001): FRPP94; Chagné *et al.* (2004): NZPR544; Guevara *et al.* (2005): B4F08; Steinitz *et al.* (2011): pEST8; F. Sebastiani and G.G. Vendramin (pers. com., June 2011): epi3; Pinzauti *et al.* (2012): Ppinea11; Elsik *et al.* (2000): PtTX3030; and González-Martínez *et al.* (2004): PtTX3116. In both species, forward primers were 5´end-labeled with fluorochromes (NED, FAM, VIC or PET, see Supporting Information S3) and loci were amplified using the Qiagen Multiplex PCR Kit (Qiagen, Venlo, The Netherlands) following manufacturer's instructions. In *P. halepensis*, pEST8 and PtTX3116 were amplified individually each in a final volume of 10 μl containing 10 ng of template DNA, 1x PCR buffer,

0.2 mM of each dNTP, 1 U of GoTaq polymerase (Promega, Madison, WI), 0.2 μM of each primer, and 1.5 mM and 4.5 mM $MgCl_2$ for pEST8 and PtTX3116, respectively. For pEST8 a touch-down PCR protocol was used: denaturation at 94°C for 3 min followed by 10 cycles at 94°C (30 s), 60°-50°C (30 s) lowering the annealing temperature by 1°C/cycle, 72°C (30 s), followed by 25 cycles at 94°C (30 s), 50°C (30 s), 72°C (30 s), and a final step at 72°C for 10 min. PtTX3116 was amplified using PCR protocol 3 described in Steinitz *et al.* (2011). Amplified fragments were separated using an ABI 3730 genetic analyzer (Applied Biosystems, Carlsbad, Ca) and fragment sizes were determined with reference to the GeneScan™ –500 LIZ® Size Standard (Applied Biosystems) using GeneMapper software version 4.0 (Applied Biosystems).

Single nucleotide polymorphisms

Single nucleotide polymorphism genotyping was conducted on *P. pinaster* samples using IlluminaVeraCode® technology for a 384-plex SNP assay (OPA [Oligo Pool Assay] design file and further details in Budde *et al.* 2014) designed based on a subsample of the 1536-plex SNP array for *P. pinaster* by Chancerel *et al.* (2011) and additional SNPs from drought stress response candidate genes (Grivet *et al.* 2011). For *P. halepensis*, a 384-plex SNP assay was designed from the aligned transcriptomic sequences of two individuals with extreme phenotypes regarding fire adaptation (specifically, with high and low numbers of serotinous cones), sequenced with Illumina HiSeq2000 technology and assembled using a mixed reference-based/*de novo* strategy (S. Pinosio and G.G. Vendramin, pers. com.). This assay also included SNPs from candidate genes sequenced in previous studies (Grivet *et al*. 2009; 2011). SNP-typing was conducted with IlluminaVeraCode® technology (OPA design file as DRYAD entry and further details on SNPs in Supporting Information S3). Conversion rates for SNP assays were relatively high, ~66% for both *P. pinaster* and *P. halepensis*, and we obtained high quality genotypes for 251 polymorphic SNP loci in each species (GC50 score of ~81% in *P. pinaster* and of ~82% in *P. halepensis*).

Genetic diversity

For SSRs, genetic diversity (H_E) was calculated in each stand using GENEPOP 4.0 (Raymond & Rousset 1995). The allelic richness (R_S), computed as mean number of alleles per locus, was assessed in FSTAT 2.9.3.2 (Goudet 1995). The fixation index (F_{IS}) was computed and deviation from zero (Hardy-Weinberg genotypic proportions) was tested by

10,000 permutations of alleles within populations in SPAGeDi 1.3d (Hardy & Vekemans 2002). For biallelic SNPs, the genetic diversity (H_E) and the fixation index (F_{IS}) were computed as for SSRs.

Demographic history

SSRs were used to assess demographic history in each stand applying three complementary approaches.

The Cornuet & Luikart's (1996) T_2 statistic, reflecting the deviation of gene diversity from expectations at demographic equilibrium, was computed using the Two-Phase Model with default parameter settings in Bottleneck 1.2.02 (Piry *et al.* 1999). This method is suitable to detect recent bottlenecks of low magnitude (Williamson-Natesan 2005). Positive T_2 values (H_E excess) indicate a bottleneck, whereas negative values (H_E deficiency) are consistent with recent population expansions. The significance of T_2 was tested using Wilcoxon's signed rank test.

Garza & Williamson's (2001) *M* was calculated as the ratio of the number of alleles over the allele size range using arlsumstat 3.5.1.1 (Excoffier & Lischer 2010). This approach is suitable to detect older and longer-lasting bottlenecks (Williamson-Natesan 2005). After a severe bottleneck event, *M* is predicted to decline because the number of alleles should decrease faster than the allele size range. A value below the critical M_C=0.68 indicates a bottleneck event (Garza & Williamson 2001).

Finally, an Approximate Bayesian Computation (ABC) approach was used to examine historical population size changes and visualize them in skyline plots (Ho & Shapiro 2011). Analyses were run in R version 2.15.0 (R Development Core Team 2012) using the abc package (Csilléry 2012), following the same approach as Burgarella *et al.* (2012). One million demographic scenarios were simulated allowing population size changes at four different time points. The values of the population sizes and the time points were drawn from prior distributions. A log uniform distribution (10^{-3}, 10^{4}) was used for the prior for theta, $\theta=2N_e\mu$, i.e., the population scaled mutation rate. Time points scaled to mutation units ($\tau=t\mu$) were taken randomly in an interval from 0 to 10 using a Dirichlet distribution whose parameter values were taken from an hyperprior with uniform distribution between 0.5 and 2. A coalescent genealogy was simulated with the software fastsimcoal (Excoffier & Foll 2011) for each demographic scenario for 11 SSRs in *P. pinaster* and for 9 SSRs in *P. halepensis* using the same sample sizes as in our study stands. Summary statistics were calculated from simulated data using arlsumstat version 3.5 (Excoffier & Lischer 2010, see Supporting Information S4). The 0.1% of simulations with summary statistics closest to those

observed in each sample stand was retained to build the skyline plot. Posterior distributions of θ were calculated at 100 time points from which the median (point estimate) and the 95% highest density probabilities of the 0.1% best fitting simulations were used to build the skyline plot.

Fine-scale spatial genetic structure

Spatial genetic structure patterns

SGS was assessed in each stand separately for each marker type and for the combined data set of SSRs and SNPs as described by Vekemans & Hardy (2004). Kinship coefficients *F* (Loiselle *et al.* 1995) were calculated in SPAGeDi 1.3d (Hardy & Vekemans 2002) for all pairs of individuals and regressed on the logarithm of spatial distance. SGS was tested by comparing the regression slope *b* to its distribution obtained from 10,000 permutations of individual locations. To display SGS graphically, kinship coefficients were averaged in distance classes (upper bounds of 20, 40, 80, 160, 320, 640, >640 m for the main stands and <40 m for the first distance class of the added *P. halepensis* stands) and plotted against the logarithm of distance. The strength of SGS was estimated as $Sp = -b/(1-F_1)$ (Vekemans & Hardy 2004) where F_1 is the kinship coefficient in the first distance class. Furthermore, for *P. halepensis*, SGS within stands was assessed across all HiFi (5 stands) or LoFi (3 stands) stands, respectively. To assess differences in SGS between fire regimes, we used *t*-tests on per-locus *b* and compared jackknife 95% confidence intervals (CIs) of *b* between fire regimes (a steeper slope indicating stronger SGS, Vekemans & Hardy 2004).

Testing for micro-environmental selection

Selective processes that operate spatially may affect the SGS of (some) functional SNP markers. SGS at SNPs can be shaped by neutral (i.e. dispersal-related) and non-neutral processes, whereas SGS at SSRs is affected by neutral processes only. Therefore, a significantly stronger SGS signal at SNPs than SSRs should represent evidence for spatially explicit micro-environmental selection (see Van Heerwarden *et al.* 2010 for an analogous approach). We used two methods to compare SGS patterns between marker types, accounting for power differences: First, we compared the jackknife 95% CIs of *b* (as above) between makers. These CIs are approximate (SPaGeDi program documentation) as their precision depends on the number and power of markers; hence they take differences in marker power implicitly into account. Second, as each locus represents an independent estimate of SGS, we applied Welch's *t*-test for (unequal) samples with

unequal variances (Welch 1947) on per-locus *b* values obtained from SNPs and SSRs. Since the SGS computation with SNPs potentially confounded regional- and local-scale selective effects, we repeated all SGS analyses using a reduced SNP data set in which we controlled for regional-scale selection by excluding SNPs that were identified as F_{ST} outliers using FDIST2 (Beaumont & Nichols 1996).

Results

Fire regime (HiFi vs. LoFi) had no effect on levels of genetic diversity (Table 1, Table 2). *Pinus halepensis* had a lower SSR diversity than *P. pinaster*; the lowest genetic diversity was found in the *P. halepensis* stand from Eslida (Table 1). In each species, 251 SNP markers were genotyped successfully and displayed polymorphism in the study region. Levels of SNP genetic diversity were similar across stands within both species (Table 1).

A fire regime related demographic signal was not observed in any species (Table 3). Bottlenecks were detected in all three *P. halepensis* stands under the Two-Phase Model (TPM) in Bottleneck, and in Sinarcas and Serra Calderona using the Garza-Williamson *M* ratio test (Table 3). In *P. pinaster*, a bottleneck was detected using *M* in the three stands, but not using Bottleneck. The skyline-plots were visually similar across sample stands, showing a recent population size decline which was more pronounced in *P. halepensis* than in *P. pinaster* (Fig. 1).

SGS was detected with higher power using SNPs (251 markers in each species) than with nine or 11 SSRs, as demonstrated by the narrower jackknife CIs of kinship vs. ln(distance) regression slopes and using numerical simulations (Fig. 2, Supporting Information S5). At SSRs, only one stand (Eslida, HiFi, in *P. halepensis*) displayed significant SGS ($P<0.05$, Table 4, Fig. 3). At SNPs, both *P. halepensis* HiFi stands (Eslida and Serra Calderona) showed significant SGS ($P<0.001$, Table 4, Fig. 3) whereas in *P. pinaster*, one HiFi and one LoFi stand displayed significant SGS ($P<0.05$, Table 4, Fig. 3). Controlling for regional-scale selective effects (i.e., removing three F_{ST} outlier SNPs in *P. halepensis* while no F_{ST} outlier was detected in *P. pinaster*) did not change the SGS results (results not shown).

The SGS results tentatively supported our hypothesis of increased SGS for the HiFi regime in *P. halepensis.* This hypothesis was confirmed with additional *P. halepensis* stands wherein SNP data displayed significant (4 stands) or marginally-significant (1 stand) SGS in HiFi stands, but only in one LoFi stand. The combined SSR and SNP data sets displayed the same SGS pattern as the SNPs (Table 2). SGS at SNPs was

stronger across *P. halepensis* HiFi than across LoFi stands (non-overlapping 95% CIs, Fig. 2; Student's *t*-test: t = -6.067, df = 486, P = 1.315e-09). For SSRs, significant SGS was detected across HiFi or LoFi stands, but no difference in SGS was observed between fire regimes (Table 2).

The comparison of SGS between marker types to test for signals of micro-environmental selection revealed a stronger SGS at SNPs than at SSRs in two *P. halepensis* HiFi stands (Alzira and Serra Calderona) by means of non-overlapping 95% CIs (Supplementary Information S5). Welch's *t*-test confirmed this result in Serra Calderona (P<0.01) and suggested a trend for a stronger SGS at SNPs than SSRs with P<0.1 in three (out of five) HiFi stands and across HiFi stands, but not within or across LoFi stands (Table 2).

Discussion

We showed in this work that two Mediterranean pine species did not exhibit fire regime related differences in genetic diversity or inferred population genetic history. Aleppo pine stands growing under high frequency of crown fires displayed a stronger spatial genetic structure at SNP markers than those growing under low frequencies of crown fires. This enhanced genetic structure associated with intense fire regime suggested a combination of neutral and/or non-neutral processes increasing the relatedness between close-by individuals in this species. In cluster pine, SGS was not related to fire regime

.

Adaptations to fire regime buffer against loss of diversity

We expected a reduction in genetic diversity and stronger bottleneck signals in stands of Mediterranean pines exposed to frequent crown fires. However, genetic diversity and estimates of demographic history were not affected by fire regime in *P. halepensis* and *P. pinaster*. Our results suggest that any loss of genetic diversity due to frequent stand-replacing fires was prevented, similarly as in the fire-adapted obligate seeder *Ulex parviflorus* (Moreira *et al.* 2014). Diversity can be preserved because of 1) adaptations to fire such as a diverse canopy seed bank due to serotinous cones and an early age of first flowering, in combination with 2) life history traits, such as a high colonization capacity of seeds from non-serotinous cones in *P. halepensis* (Goubitz *et al.* 2004) and high levels of overall gene flow in both pine species (de-Lucas *et al*. 2009; Shohami & Nathan 2014). Seed banks are known to effectively maintain diversity and buffer against demographic fluctuations (Templeton & Levin 1979; Ayre *et al.* 2009). In Proteaceae species genetic diversity accumulates fast in seed banks under wide-ranging gene flow as it

depends more on the number of cones rather than on time since fire (Barrett *et al.* 2005; Ayre *et al.* 2010). In *P. halepensis* increased wind speeds enhanced pollen gene flow in the vegetation-free post-fire landscape (Shohami & Nathan 2014). These factors could contribute to a fast post-fire recovery of a diverse canopy seed bank in the study species.

Each species has its own demographic history

Bottleneck signals were detected in all stands of both species independently of fire regime. The ABC skyline-plots dated the bottlenecks to the latest tens or hundreds of generations, similarly as in *Taxus baccata,* a conifer that does not occur in fire prone environments (Burgarella *et al.* 2012). This may reflect that both pine species probably experienced a bottleneck during the last glacial maximum (LGM). The bottleneck dating obtained through ABC skyline-plots must however be interpreted with caution as it integrates the uncertainty from both mutation rates and generation times. Bottlenecks in *P. halepensis* were likely more severe or longer lasting than those in *P.* pinaster, a pattern that is consistent with broad scale demographic signals in both species (Grivet *et al.* 2009; Jaramillo-Correa *et al.* 2010). This differential signal could be due to 1) bottlenecks in *P. halepensis* during its Pleistocene recolonization of the Western Mediterranean basin from the Balkan peninsula (Morgante *et al.* 1997; Bucci *et al.* 1998; Grivet *et al.* 2009; Jaramillo-Correa *et al*. 2010) vs. long term persistence of *P. pinaster* in the Iberian Peninsula (Carrión et al. 2000; Carrión & Van Geel 1999, Gómez *et al.* 2005) or 2) a mid-Holocene phase of low fire incidence which negatively affected the abundance of *P. halepensis*. Indeed, there is evidence that oak forest replaced pine forest in the south-eastern Iberian Peninsula ca. between 8000 and 5500 BP, concomitantly with a reduction in fire frequency (Vannière *et al.* 2011). This could have caused a drastic bottleneck in pines, especially in the apparently more fire-dependent *P. halepensis*, which was near-absent during this period in the study region but abundant before and afterwards (Carrión & Van Geel 1999; Carrión *et al.* 2012).

Frequent fires enhance within-stand genetic structure in Aleppo pine

We expected an enhanced SGS under frequent crown fires because of previous indications of clumped post-fire recruitment near burnt adults (Ne'eman & Izhaki 1998) and significant spatial autocorrelation of serotiny associated to frequent fires in *P. halepensis* (Hernández-Serrano *et al.* 2013). We indeed observed a stronger SGS in

P. halepensis HiFi than in LoFi stands using SNPs, matching results of spatial autocorrelation for serotiny (Hernández-Serrano *et al.* 2013). In LoFi stands, recruitment occurs from non-serotinous cones, and our detection of very weak SGS suggests an efficient genetic mixing in pollination and seed dispersal. The enhanced SGS detected with SNPs in HiFi stands can be due to dispersal-related selectively neutral processes, possibly in combination with non-neutral processes. Hernández-Serrano *et al.* (2013) suggested that spatial autocorrelation for serotiny in HiFi stands may reflect restricted overlapping of post-fire seed rain shadows. Dispersal-related differences between fire regimes should similarly affect SGS at SNPs and SSRs. Therefore, our failure to observe significant SGS with SSRs in HiFi stands could be due to the lower power of these markers compared to SNPs.

We observed stronger SGS at SNPs than at SSRs across HiFi stands (P=0.081) suggesting that in addition to neutral processes, SNPs could bear a signal of heterogeneous, patch-dependent selection (Epperson 1995; Linhart & Grant 1996). Selection at short spatial scale can be caused by micro-environmental factors which influence post-fire seedling establishment and survival (Buscardo *et al.* 2011, Pausas *et al.* 2004). A robust signal for micro-environmental selection was only detected in Serra Calderona (P=0.009), a stand with a particularly rugged topography and steep slopes with variable aspect. Interestingly, in this stand spatial autocorrelation for serotiny was not detected (Table 2, Hernández-Serrano *et al.* 2013). If dispersal-related clumping were very weak in this stand, the interaction between neutral and selective processes may have been weak as well, facilitating the discrimination of non-neutral effects.

In general, our data lacked power to confidently distinguish the contributions of (neutral) dispersal effects and putative micro-environmental selection on SGS. This is largely due to a generally weak SGS pattern, on the order of Sp=0.0045 for SSRs in HiFi stands, which corresponds to the 10% weakest SGS patterns described for plants (Vekemans & Hardy 2004). Hence, although our SSR markers were powerful enough to detect a significant SGS across stands, the 95% CI for b was wide and did not allow differentiating SGS in HiFi vs. LoFi stands, nor in SSR vs. SNP markers. Future research into micro-environmental selection in this system may be a promising avenue and could proceed by comparing SGS at SNPs and neutral markers with higher statistical power, and/or by examining the effect of particular micro-environmental conditions on SGS.

Between-species differences in fire-dependence and SGS

In *P. pinaster*, significant SGS was detected in two out of three stands at SNPs only, and it was not related to the fire regime. SGS in the two *P. pinaster* stands was lower than in most *P. halepensis* stands. Weak or non-significant SGS has been described for wind pollinated trees with wind dispersed seeds (mean *Sp* = 0.012, n = 5 species, Vekemans & Hardy 2004). Our results illustrate that powerful markers such as SNPs are needed to confidently detect very weak SGS. The absence of a fire-related SGS signal in *P. pinaster* might relate to its propensity to exhibit fire resister traits. If these traits enable a significant proportion of adults to survive fire events, SGS will depend to a lesser extent on post-fire recruitment from the canopy seed bank than in *P. halepensis*. If so, the expected SGS in *P. pinaster* should be mainly influenced by stand-specific processes rather than fire regime. Conversely, the fire-related SGS signal of *P. halepensis* probably reflects its strong fire dependence with much greater risk of post-fire adult mortality and the more constant expression of fire-related traits (Fernandes *et al.* 2008). In our work it has not been possible to reveal the specific traits or functional genes responsible for enhanced fire-related SGS in *P. halepensis*. Only few individual SNPs retained significant SGS after multiple test corrections (40 SNPs in Eslida, 5 SNPs in Calderona, 1 SNP in Montán, results not shown) and these SNPs overlapped only minimally among stands. Future research into local selection processes should seek to combine SGS patterns with phenotypic trait measurements and biotic and abiotic micro-environmental site variation.

Conclusions

Predicting future fire regime changes is extremely challenging and complex (Keeley *et al.* 2012). During climate change, fire frequency is expected to increase, including areas where fire has not had an important impact so far (e.g., Piñol *et al.* 1998; Pausas 2004). *Pinus halepensis* and *P. pinaster* stands are well-adapted to stand replacing fires in a region with a long fire history. However, intervals between fire events must be sufficiently long to enable the build-up of a seed bank, i.e., longer than the minimum age of reproduction for post-fire recruitment, which is 4-8 years in *P. halepensis* and 4-10 years in *P.pinaster* (Tapias *et al.* 2004). Currently *P. halepensis* is expanding at the expense of *P. pinaster* in xeric and highly fire-prone habitats because it has higher fertility, a higher percentage of serotinous cones and higher cone-opening temperatures, which enable it to maintain a greater canopy seed bank (Tapias *et al.* 2001;

Tapias *et al*. 2004). Our results indicate that this replacement will favour the species with less genetic diversity and stronger SGS under recurrent crown fire regime. We demonstrate that the impact of fire can be different in species with similar fire syndromes and life-history traits. Therefore, species-specific studies are needed to understand the role of wildfires for the functioning and evolution of future forest ecosystems.

Acknowledgements

We thank Ana I. de-Lucas, Cristobal Odoñez, Yusuf Kurt, Fernando del Caño, Diana Barba and Francisco Auñon for field assistance and Carmen García-Barriga (INIA-CIFOR), Francesca Pinzauti (CNR-IGV), Sara Torre (CNR-IGV) and Federico Sebastiani (CNR-IGV) for assistance in the laboratory. Three anonymous reviewers and the editor Prof. Victoria Sork have provided useful comments to improve the quality of the manuscript. The present study was funded by the following projects (Spanish Ministry of Science and Innovation, MICINN): VAMPIRO (CGL2008-05289-C02-01/02), LinkTree (EUI2008-03713, under the ERAnet-BiodivERsA call), and TREVOL (CGL2012-39938-C02-00/01). GGV was also supported by the Italian MIUR project "Biodiversitalia" (RBAP10A2T4). KBB acknowledges a PhD scholarship (FPI: BES-2009-015443) of theMICINN. MH is a 'Ramón y Cajal' researcher of the MICINN and acknowledges funding from the Spanish Ministry for Economy and Competitiveness (CGL2012-40129-C02-02/AFFLORA) and a Marie-Curie Intra-European fellowship.

References

Abdel Malak D and Pausas JG (2006) Fire regime and post-fire Normalized Difference Vegetation Index changes in the eastern Iberian Peninsula (Mediterranean basin). *International Journal of Wildland Fire*, **15**, 407-413.

Agee JK (1998) Fire and pine ecosystems. In: *Ecology and biogeography of Pinus* (ed. Richardson DM), pp. 193–218, Cambridge University Press, Cambridge.

Ayre DJ, Ottewell KM, Krauss SL, Whelan RJ (2009) Genetic structure of seedling cohorts following repeated wildfires in the fire-sensitive shrub *Persoonia mollis ssp. nectens*. *Journal of Ecology*, **97**, 752–760.

Ayre DJ, O'Brien E, Ottewell KM, Whelan RJ (2010) The accumulation of genetic diversity within a canopy stored seed bank. *Molecular Ecology*, **19**, 2640–2650.

Ayres MP, Lombardero MJ (2000) Assessing the consequences of global change for forest disturbance from herbivores and pathogens. *The Science of the Total Environment*, **262**, 263–286.

Barrett LG, He T, Lamont BB, Krauss SL (2005) Temporal patterns of genetic variation across a 9-year-old aerial

seed bank of the shrub *Banksia hookeriana* (Proteaceae). *Molecular Ecology*, **14**, 4169–4179.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B*, **263**, 1619 –1626

Bradstock RA, Bedward M, Scott J, Keith DA (1996) Simulation of the effect of spatial and temporal variation in fire regimes on the population viability of a *Banksia* species. *Conservation Biology*, **10**, 776–784.

Bucci G, Anzidei M, Madaghiele A *et al.* (1998) Detection of haplotypic variation and natural hybridization in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. *Molecular Ecology*, **7**, 1633–1643.

Budde KB, Heuertz M, loz-Serrano A, Pausas J, Vendramin GG, Verdú M, González-Martínez SC (2014) *In situ* genetic association for fire phenotypes in Mediterranean maritime pine (*Pinus pinaster* Aiton). *New Phytologist,* **201**, 230–241.

Burgarella C, Navascués M, Zabal-Aguirre M *et al.* (2012) Recent population decline and selection shape diversity of taxol-related genes. *Molecular Ecology*, **21**, 3006–3021.

Buscardo E, Freitas H, Pereira JS, De Angelis P (2011) Common environmental factors explain both ectomycorrhizal species diversity and pine regeneration variability in a post-fire Mediterranean forest. *Mycorrhiza*, **21**, 549–558.

Carrión JS, Fernández S, González-Sampériz P *et al*. (2012) *Paleoflora y Paleovegetación de la Península Ibérica e Islas Baleares: Plioceno-Cuaternario.*Ministerio de Economía y Competitividad, Madrid.

Carrión JS, Van Geel B (1999) Fine-resolution Upper Weichselian and Holocene palynological record from Navarrés (Valencia, Spain) and a discussion about factors of Mediterranean forest succession. *Review of Palaeobotany and Palynology*, **106**, 209–236.

Carrión JS, Navarro C, Navarro J, Munuera M (2000) The distribution of cluster pine (*Pinus pinaster*) in Spain as derived from palaeoecological data: relationships with phytosociological classification. *The Holocene*, **2**, 243–252.

De Castro M, Martín-Vide J, Alonso S (2005) El clima de España: pasado, presente y escenarios de clima para el siglo XXI. In: *Impactos del Cambio Climático en España*, pp. 1–64.

Catalán G (1991) Semillas de árboles y arbustos forestales. Colección Técnica, ICOMA Madrid.

Chagné D, Chaumeil P, Ramboer A *et al.* (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics*, **109**, 1204–1214.

Chancerel E, Lepoittevin C, Le Provost G *et al.* (2011) Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine. *BMC Genomics*, **12**, 368.

Cornuet JM, Luikart G (1996) Power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC).

Methods in Ecology and Evolution, **3**, 475–479.

de-Lucas A, González-Martínez SC, Vendramin GG, Hidalgo E, Heuertz M (2009) Spatial genetic structure in continuous and fragmented populations of *Pinus pinaster* Aiton. *Molecular Ecology*, **18**, 4564–4576.

de-Lucas A, Robledo-Arnuncio JJ, Hidalgo E, González-Martínez SC (2008) Mating system and pollen gene flow in Mediterranean maritime pine. *Heredity*, **100**, 390–399.

Elsik CG, Minihan VT, Hall SE, Scarpa AM, Williams CG (2000) Low-copy microsatellite markers for *Pinus taeda* L. *Genome*, **43**, 550–555.

Epperson BK (1995) Spatial structure of two-locus genotypes under isolation by distance. *Genetics*, **140**, 365–75.

Excoffier L, Foll M (2011) fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, **27**, 1332–1334.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fernandes PM, Rigolot E (2007) The fire ecology and management of maritime pine (*Pinus pinaster* Ait.). *Forest Ecology and Management*, **241**, 1–13.

Fernandes PM, Vega J, Jimenez E, Rigolot E (2008) Fire resistance of European pines. *Forest Ecology and Management*, **256**, 246–255.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.

Gómez A, Vendramin GG, González-Martínez SC, Alía R (2005) Genetic diversity and differentiation of two Mediterranean pines (*Pinus halepensis* Mill. and *Pinus pinaster* Ait.) along a latitudinal cline using chloroplast microsatellite markers. *Diversity and Distributions*, **11**, 257-263.

González-Martínez SC, Burczyk J, Nathan R *et al.* (2006) Effective gene dispersal and female reproductive success in Mediterranean maritime pine (*Pinus pinaster* Aiton). *Molecular Ecology*, **15**, 4577–4588.

González-Martínez SC, Gerber S, Cervera MT *et al.* (2002) Seed gene flow and fine-scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. *Theoretical and Applied Genetics*, **104**, 1290–1297.

González-Martínez SC, Robledo-Arnuncio JJ, Collada C *et al.* (2004) Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines. *Theoretical and Applied Genetics*, **109**, 103–111.

Goubitz S, Nathan R, Roitemberg R, Shmida A, Ne'eman G (2004) Canopy seed bank structure in relation to: fire, tree size and density. *Plant Ecology*, **173**, 191–201.

Goudet J (1995) FSTAT (Version 1.2): A Computer Program to calculate F-statistics. *The Journal of Heredity*, **86**, 485–486.

Grivet D, Sebastiani F, Alía R *et al.* (2011) Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology and Evolution*, **28**, 101–116.

Grivet D, Sebastiani F, González-Martínez SC, Vendramin GG (2009) Patterns of polymorphism resulting from long-range colonization in the

Mediterranean conifer Aleppo pine. *New Phytologist*, **184**, 1016–1028.

Guevara MA, Chagné D, Almeida MH *et al.* (2005) Isolation and characterization of nuclear microsatellite loci in *Pinus pinaster* Ait. *Molecular Ecology Notes*, **5**, 57–59.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

He T, Pausas JG, Belcher CM, Schwilk DW, Lamont BB (2012) Fire-adapted traits of *Pinus* arose in the fiery Cretaceous. *New Phytologist*, **194**, 751–9.

van Heerwaarden J, Ross-Ibarra J, Doebley JF *et al.*(2010) Fine scale genetic structure in the wild ancestor of maize (*Zea mays* ssp. *parviglumis*). *Molecular Ecology*, **19**, 1162–1173.

Hernández-Serrano A, Verdú M, González-Martínez SC, Pausas JG (2013) Fire structures pine serotiny at different scales. *American Journal of Botany*, in press.

Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, **11**, 423–434.

Jaramillo-Correa JP, Verdú M, González-Martínez SC (2010) The contribution of recombination to heterozygosity differs among plant evolutionary lineages and life-forms. *BMC Evolutionary Biology*, **10**, 22.

Keeley JE, Bond WJ, Bradstock RA, Pausas JG, Rundel PW (2012) *Fire in Mediterranean Ecosystems. Ecology, Evolution and Management,* pp.515, Cambridge University Press, Cambridge.

Keeley JE, Pausas JG, Rundel PW, Bond WJ, Bradstock RA (2011) Fire as an evolutionary pressure shaping plant traits. *Trends in Plant Science*, **16**, 406–411.

Lamont B, Le Maitre D, Cowling R (1991) Canopy seed storage in woody plants. *The Botanical Review*, **5**, 277–317.

Lavorel S, Flannigan MD, Lambin EF, Scholes MC (2006) Vulnerability of land systems to fire: Interactions among humans, climate, the atmosphere, and ecosystems. *Mitigation and Adaptation Strategies for Global Change*, **12**, 33–53.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, **27**, 237–277.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Mariette S, Chagné D, Decroocq S *et al.* (2001) Microsatellite markers for *Pinus pinaster* Ait. *Annals of Forest Science*, **58**, 203–206.

Moreira B, Castellanos MC, Pausas JG (2014) Genetic component of flammability variation in a Mediterranean shrub. *Molecular Ecology*, doi: 10.1111/mec.12665.

Morgante M, Felice N, Vendramin GG (1997) Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck. In: *Molecular Tools for Screening Biodiversity* (eds Karp A, Isaac PG, Ingram DS), pp. 407–412. Springer Netherlands, Dordrecht.

Namroud M-C, Park A, Tremblay F, Bergeron Y (2005) Clonal and spatial genetic structures of aspen (*Populus tremuloides* Michx.). *Molecular Ecology*, **14**, 2969–2980.

Ne'eman G, Goubitz S, Nathan R (2004) Reproductive traits of *Pinus halepensis* in the light of fire – a critical review. *Plant Ecology*, 69–79.

Ne'eman G, Izhaki I (1998) Stability of pre- and post-fire spatial structure of pine trees in Aleppo pine forest. *Ecography*, **21**, 535–542.

Odion D, Tyler C (2002) Are long fire-free periods needed to maintain the endangered, fire-recruiting shrub *Arctostaphylos morroensis* (Ericaceae)? *Conservation Ecology*, **6**, 4.

Pausas JG (2004) Changes in fire and climate in the Eastern Iberian Peninsula (Mediterranean Basin). *Climatic Change*, **63**, 337–350.

Pausas JG, Fernández-Muñoz S (2012) Fire regime changes in the Western Mediterranean Basin: from fuel-limited to drought-driven fire regime. *Climatic Change*, **110**, 215–226.

Pausas JG, Keeley JE (2009) A Burning Story: The Role of Fire in the History of Life. *BioScience*, **59**, 593–601.

Pausas JG, Ribeiro E, Vallejo R (2004) Post-fire regeneration variability of *Pinus halepensis* in the eastern Iberian Peninsula. *Forest Ecology and Management*, **203**, 251–259.

Pinzauti F, Sebastiani F, Budde KB, Fady B, Vendramin GG (2012) Nuclear microsatellites for *Pinus pinea* (Pinaceae), a genetically depauperate tree, and their transferability to *P. halepensis*. *American Journal of Botany*, **99**, e362–365.

Piry S, Luikart G, Cornuet J-M (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *The Journal of Heredity*, **90**, 502–503.

Piñol J, Terradas J, Lloret F (1998) Climate warming, wildfire hazard, and wildfire occurrence in coastal eastern Spain. *Climatic Change*, **38**, 345–357.

Premoli AC, Kitzberger T (2005) Regeneration mode affects spatial genetic structure of *Nothofagus dombeyi* forests. *Molecular Ecology*, **14**, 2319–2329.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *The Journal of Heredity*, **86**, 248–249.

Saracino A, Pacella R, Leone V *et al.* (1997) Seed dispersal and changing seed characteristics in a *Pinus halepensis* Mill. forest after fire. *Plant Ecology*, **130**, 13–19.

Schiller G, Ne'eman G, Korol L (1997) Post fire vegetation dynamics in native *Pinus halepensis* Mill. forest on Mt. Carmel, Israel. *Israel Journal of Plant Sciences*, **45**, 297–308.

Schwilk DW, Ackerly DD (2001) Flammability and serotiny as strategies: correlated evolution in pines. *Oikos*, **94**, 326–336.

Segarra-Moragues JG, Ojeda F (2010) Postfire response and genetic diversity in *Erica coccinea*: connecting population dynamics and diversification in a biodiversity hotspot. *Evolution*, **64**, 3511–3524.

Segarra-Moragues JG, Torres-Díaz C, Ojeda F (2012) Are woody seeder plants more prone than resprouter to population genetic differentiation in

Mediterranean-type ecosystems? *Evolutionary Ecology*, **27**, 117–131.

Shohami D, Nathan R (2014) Fire-induced population reduction and landscape opening increases gene flow via pollen dispersal in *Pinus halepensis. Molecular Ecology*, doi: 10.1111/mec.12506

Steinitz O, Troupin D, Vendramin GG, Nathan R (2011) Genetic evidence for a Janzen–Connell recruitment pattern in reproductive offspring of *Pinus halepensis* trees. *Molecular Ecology*, **20**, 4152–4164.

Tapias R, Climent J, Pardos JA, Gil L (2004) Life histories of Mediterranean pines. *Plant Ecology*, **171**, 53–68.

Tapias R, Gil L, Fuentes-Utrilla P, Pardos JA (2001) Canopy seed banks in Mediterranean pines of south-eastern Spain: a comparison between *Pinus halepensis* Mill., *P. pinaster* Ait., *P. nigra* Arn. and *P. pinea* L. *Journal of Ecology*, **89**, 629–638.

Templeton AR, Levin DA (1979) Evolutionary consequences of seed pools. *The American Naturalist*, **114**, 232–249.

Troupin D, Nathan R, Vendramin GG (2006) Analysis of spatial genetic structure in an expanding *Pinus halepensis* population reveals development of fine-scale genetic clustering over time. *Molecular Ecology*, **15**, 3617–3630.

Turner MG, Romme WH, Gardner RH (1999) Prefire heterogeneity, fire severity, and early postfire plant reestablishment in subalpine forests of Yellowstone National Park, Wyoming. *International Journal Of Wildland Fire*, **9**, 21–36.

Vannière B, Power MJ, Roberts N *et al.* (2011) Circum-Mediterranean fire activity and climate changes during the mid-Holocene environmental transition (8500-2500 cal. BP). *The Holocene*, **21**, 53–73.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Whelan RJ (1995) *The ecology of fire*. Cambridge University Press, Cambridge.

Williamson-Natesan EG (2005) Comparison of methods for detecting bottlenecks from microsatellite loci. *Conservation Genetics*, **6**, 551–562.

Author contributions

SCGM, MH, JGP and MV designed the study; KB, AHS, MZA, MH and SCGM performed the sampling; KB performed the molecular analyses assisted by GMU under the supervision of SCGM, MH and GGV; KB and MH analysed the data; MN and CB contributed to the ABC skyline-plot analysis; AHS provided phenotype data and analysis; ZL and MZA created the SNP genotyping assay in *P. halepensis*; GGV facilitated microsatellite genotyping and contributed to generating transcriptome sequence data used for the SNP genotyping assay in *P. halepensis*; KB and MH wrote the first draft of the manuscript. All authors contributed to data interpretation and improvement of the manuscript.

Tables

Table 1 Genetic diversity of *Pinus halepensis* and *P. pinaster* stands. *n*, sample size; H_E, gene diversity and its standard error (SE); R_S, allelic richness standardized to a sample of 61 diploids in *P. halepensis* or 56 diploids in *P. pinaster*; F_{IS}, inbreeding coefficient (significance test for F_{IS} observed >F_{IS} expected: *, P<0.05; ***, P<0.001; n.s.: not significant).

	Pinus halepensis					
	SSRs				SNPs	
	n	H_E(SE)	R_S(SE)	F_{IS}	H_E(SE)	F_{IS}
Sinarcas (LoFi)	67	0.441 (0.019)	3.957 (1.037)	-0.073 n.s.	0.289 (0.001)	0.023 *
Serra Calderona (HiFi)	64	0.470 (0.014)	4.638 (1.315)	0.044 n.s.	0.300 (0.001)	0.071 ***
Eslida (HiFi)	67	0.359 (0.017)	3.069 (0.723)	-0.006 n.s.	0.266 (0.001)	0.054 ***
	Pinus pinaster					
	SSRs				SNPs	
	n	H_E(SE)	R_S(SE)	F_{IS}	H_E(SE)	F_{IS}
Sinarcas (LoFi)	67	0.580 (0.020)	6.282 (0.839)	0.031 n.s.	0.303 (0.001)	-0.007n.s.
Serra Calderona (HiFi)	67	0.587 (0.016)	5.781 (0.773)	0.056 *	0.292 (0.001)	0.017 *
Eslida (HiFi)	67	0.564 (0.018)	5.922 (0.712)	0.016 n.s.	0.296 (0.001)	-0.029n.s.

Table 2 Fine-scale spatial genetic structure (SGS) and serotiny phenotype autocorrelations (see Hernández-Serrano *et al.* 2013) in LoFi and HiFi stands of *Pinus halepensis*. *np,* number of pairs*, Sp,* intensity of the SGS, *P*-values of regression slope *b*: n.s., not significant. S. Calderona: Serra Calderona.

Location	Fire regime	*np*	9 SSRs	serotiny		9 SSRs		251 SNPs		251 SNPs & 9 SSRs		Welch test, H1: $b_{SNPs}<b_{SSRs}$
			H_E (SE)	Slope	*P*-value	*Sp*	*P*-value	*Sp*	*P*-value	*Sp*	*P*-value	*P*-value
Alzira	HiFi	630	0.385 (0.042)	-0.236	<0.001	-0.0070	n.s.	0.0068	0.0012	0.0070	0.0024	0.058
Cabanes	HiFi	465	0.344 (0.049)	-0.260	<0.001	-0.0062	n.s.	0.0160	0.0003	0.0150	0.0003	0.092
S. Calderona	HiFi	1953	0.471 (0.042)	-0.031	n.s.	0.0011	n.s.	0.0112	0.0000	0.0106	0.0000	0.009
Eslida	HiFi	2211	0.359 (0.052)	-0.052	0.027	0.0109	0.0027	0.0104	0.0000	0.0103	0.0000	n.s.
Serra d'Irta	HiFi	666	0.458 (0.045)	-0.099	0.021	0.0049	n.s.	0.0020	0.0951	0.0027	0.0557	n.s.
All HiFi	HiFi	6084				0.0045	0.0218	0.0094	<0.0001	0.0090	0.0000	0.081
Montán	LoFi	406	0.400 (0.064)	0.025	n.s.	0.0169	0.0377	0.0143	0.0000	0.0132	0.0000	n.s.
Sinarcas	LoFi	2145	0.441 (0.056)	-0.020	n.s.	0.0030	n.s.	0.0005	n.s.	0.0007	n.s.	n.s.
Titaguas	LoFi	496	0.428 (0.063)	0.026	n.s.	0.0048	n.s.	-0.0027	n.s.	-0.0017	n.s.	n.s.
All LoFi	LoFi	3113				0.0055	0.0332	0.0027	0.003	0.0027	0.0014	n.s.

Table 3 Bottleneck tests for SSR markers in *Pinus halepensis* and *P. pinaster* stands. T_2, bottleneck statistic of Cornuet and Luikart (1996), *P*-value of the Wilcoxon signed rank test (one tail for *H* excess) under the Two-Phase Model (TPM) calculated in Bottleneck software; *M*, ratio of the number of alleles over the allele size range; a bottleneck is indicated for values below the critical M_c=0.68.

	Pinus halepensis			*Pinus pinaster*		
	T_2	*P*-value	*M*	T_2	*P*-value	*M*
Sinarcas (LoFi)	4.265	0.001	0.585	1.703	0.139	0.492
Serra Calderona (HiFi)	3.955	0.001	0.640	1.385	0.071	0.433
Eslida (HiFi)	4.095	0.001	0.685	1.671	0.289	0.648

Table 4 Fine-scale spatial genetic structure (SGS) in *Pinus halepensis* and *P. pinaster* stands. np, number of pairs, F_1, average kinship coefficient between individuals separated by less than 20 m*; b,* regression slope and significance of SGS (n.s., not significant; *, *P*< 0.05; **, *P*<0.01, ***, *P*<0.001); *Sp,* intensity of SGS. S. Calderona: Serra Calderona.

			Pinus halepensis								
			SSRs			SNPs			SNPs & SSRs		
	Fire	*np*	F_1	*b*	*Sp*	F_1	*b*	*Sp*	F_1	*b*	*Sp*
Sinarcas	LoFi	2145	0.0101	-0.0030 n.s.	0.0030	0.0058	-0.0005 n.s.	0.0005	0.0008	-0.0007 n.s.	0.0007
S. Calderona	HiFi	1953	0.0135	-0.0011 n.s.	0.0011	0.0515	-0.0107 ***	0.0112	0.0234	-0.0098 ***	0.0100
Eslida	HiFi	2211	0.0517	-0.0103 *	0.0109	0.0336	-0.0100 ***	0.0104	0.0308	-0.0096 ***	0.0099
			Pinus pinaster								
			SSRs			SNPs			SNPs & SSRs		
	Fire	*np*	F_1	*b*	*Sp*	F_1	*b*	*Sp*	F_1	*b*	*Sp*
Sinarcas	LoFi	2211	0.0108	-0.0009 n.s.	0.0009	0.0234	-0.0036 *	0.0036	0.0226	-0.0034 **	0.0034
S. Calderona	HiFi	2145	-0.0014	-0.0017 n.s.	0.0017	0.0048	0.0001 n.s.	-0.0012	0.0044	<-0.0001 n.s.	<0.0001
Eslida	HiFi	2145	-0.0080	-0.0007 n.s.	0.0007	0.0145	-0.0021 *	0.0022	0.0128	-0.0020 *	0.0021

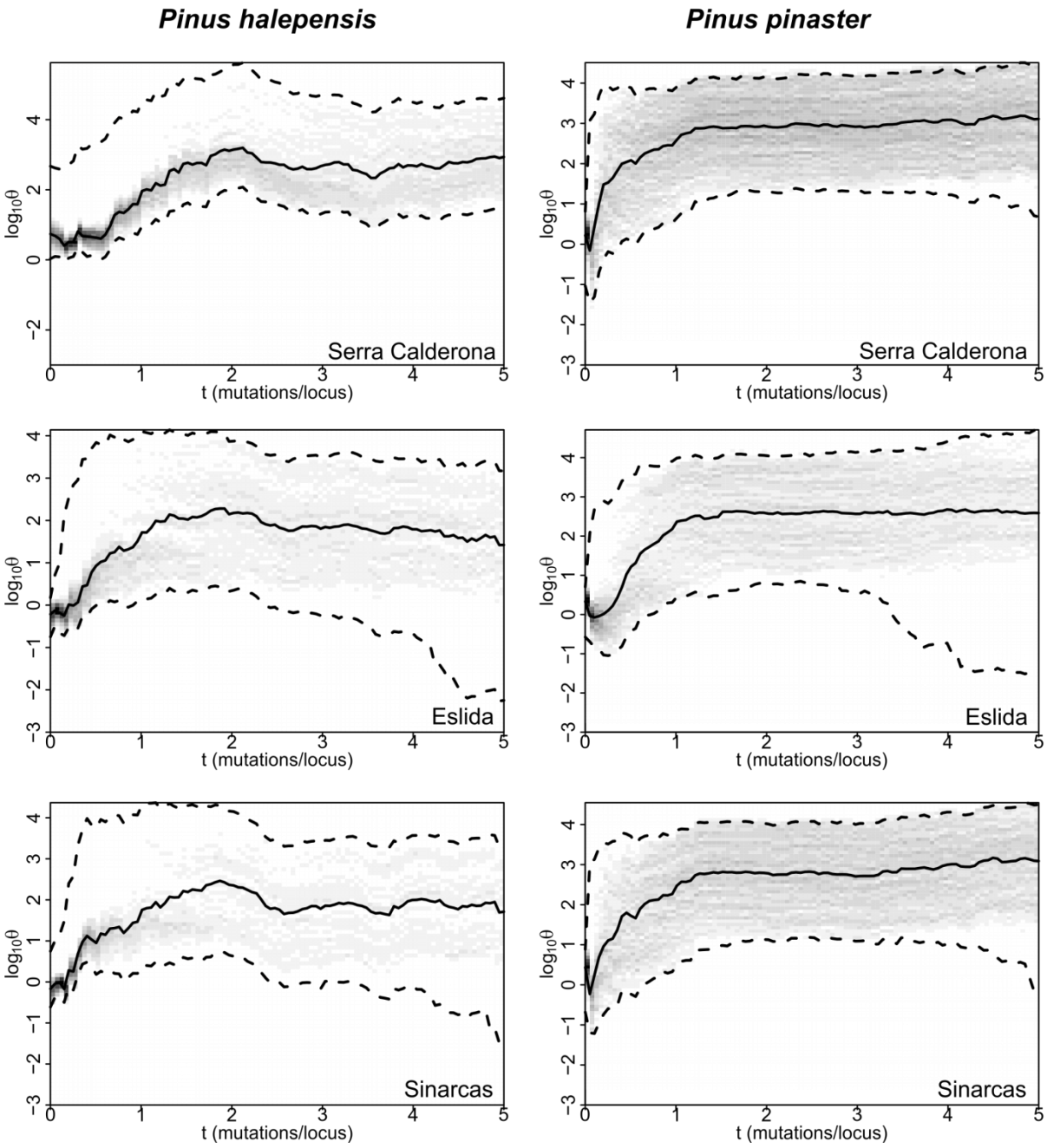

Figure 1 Posterior density distribution of the mutation-scaled population size ($\theta=2N_e\mu$) as a function of time (as mutation/locus), estimated using Approximate Bayesian Computation for each of the three *P. halepensis* and *P. pinaster* stands. Black continuous lines represent the medians of the posterior distribution and dotted lines the 95% highest posterior density intervals calculated for 0.1% simulated data sets closest to the observed data.

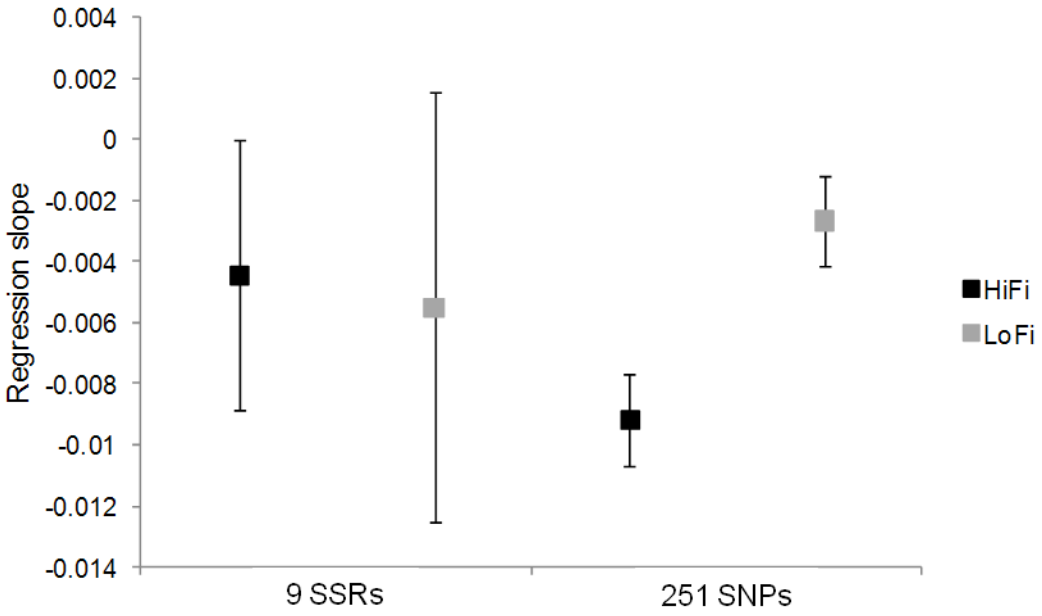

Figure 2 Mean jackknife regression slopes and 95% confidence intervals of within stand SGS across *P. halepensis* HiFi (black squares) or LoFi (grey squares) stands, respectively. HiFi stands had stronger SGS (more negative *b*) than LoFi stands at SNP but not at SSR markers.

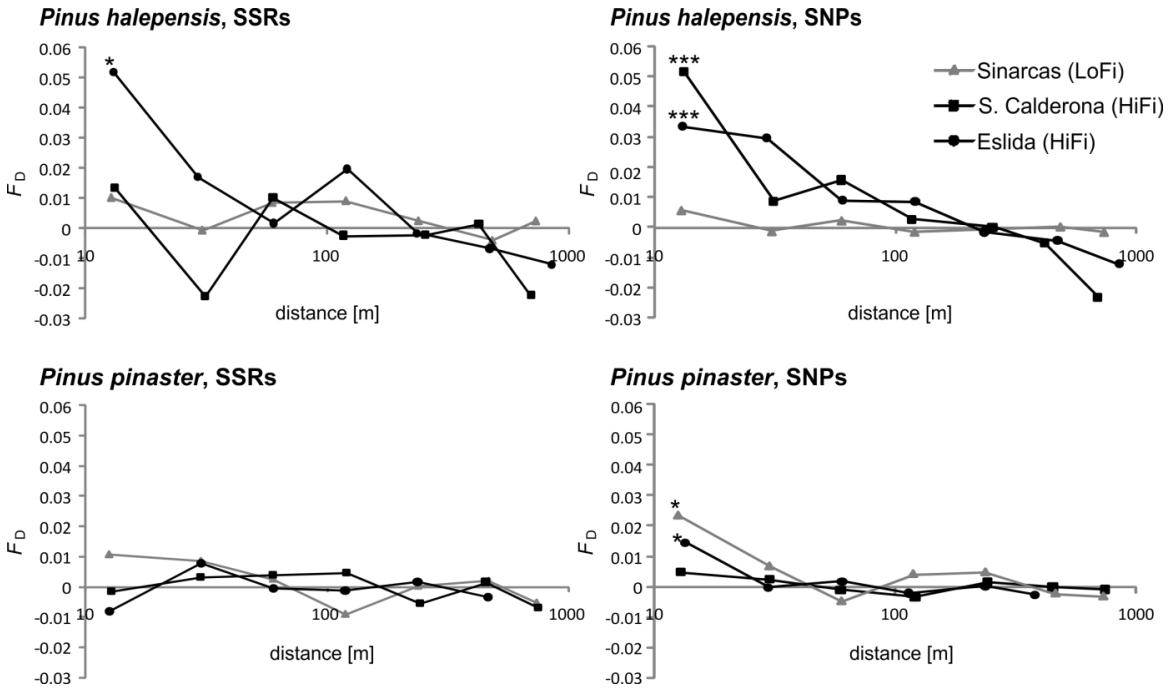

Figure 3 Spatial genetic structure in stands of *P. halepensis* and *P. pinaster* using SSRs and SNPs. Average kinship coefficient F_D was plotted against the logarithm of the mean geographic distance between individuals in each distance class. The colours of the curve indicate the fire recurrence, grey for low crown fire recurrence (LoFi) and black for high crown fire recurrence (HiFi); the symbols indicate the sample location. Significance of the SGS pattern is indicated by asterisks (see levels in Table 4). S. Calderona: Serra Calderona.

Chapter 4

Local-scale genetics — genetic structure in natural *Pinus pinaster* populations at short spatial scales

This chapter is based upon the manuscript:

Budde KB, Heuertz M, Gugerli F, Herández-Serrano A, Pausas JG, Verdú M, González-Martínez SC. Localscale genetics - genetic and phenotypic structure in natural *Pinus pinaster* populations at short spatial scales. *In preparation*.

Local-scale genetics — genetic structure in natural *Pinus pinaster* populations at short spatial scales

Katharina B. Budde[1], Myriam Heuertz[1], Ana Hernández-Serrano[2], Felix Gugerli[3], Miguel Verdú[2], Juli G. Pausas[2], Santiago C. González-Martínez[1]

[1]Department of Forest Ecology and Genetics, INIA Forest Research Centre, 28040 Madrid, Spain.

[2]Centro de Investigaciones sobre Desertificación (CIDE-CSIC/UV/GV), 46113 Moncada (Valencia), Spain.

[3]WSL Swiss Federal Research Institute, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland.

Abstract

Landscape genetic studies aim to disentangle which environmental factors, such as climate, topography, landscape elements or composition, shape the genetic structure of species or populations. Many recent studies have determined environmental factors affecting neutral and adaptive genetic variation in plant species and populations on a regional to continental scale. However, selection is expected to be hard to detect at short spatial scales, where extensive gene flow can exceed the migration–selection equilibrium. Nevertheless, genetic structure over short distances has often been described in herbaceous and woody plant species, but the effect of environmental factors has rarely been taken into account at the local scale.

Here we tested the effect of topography (altitude, slope and aspect) and biotic indicators (vegetation cover and herbaceous species richness), which characterize micro-environmental conditions, on the genetic structure at single-nucleotide polymorphisms (SNPs) and on phenotypes (such as water use efficiency, tackled by carbon isotope discrimination ($\delta^{13}C$), wood density and growth of tree rings) within natural populations of *Pinus pinaster* in eastern Spain. In one site (Eslida) an altitudinal gradient of 300 m was identified to significantly affect the genetic structure and the growth of tree rings, while no environmental factor was apparently strong enough to affect the genetic structure in other sites.

In Eslida, a number of environmental factors could vary along the altitudinal gradient, which could lead to divergent selection between upper and lower parts of slopes. Furthermore, adaptive and/or plastic responses in the flowering phenology could lead to assortative mating and shape the genetic structure. Our results showed that a genetic structuring, not due to isolation by distance, but in relation to environmental factors can be detected even at a very small spatial scale.

Keywords: *Pinus pinaster*, environmental correlations, genetic differentiation, gene flow, carbon isotope discrimination, wood density, growth rings

Introduction

Spatial genetic structure, that is, the nonrandom distribution of alleles or genotypes in space, is influenced by mutation, migration, selection and drift (e.g. Wright, 1949). Genetic structure is found at the level of species, among populations or within populations (Epperson 2003). Numerous studies have described genetic patterns, such as clines (Bergmann 1978), isolation by distance

(reviewed in Heywood 1991; Vekemans & Hardy 2004), isolation by adaptation (Nosil *et al.* 2007) or genetic discontinuities in gene flow (Jaramillo-Correa *et al.* 2010). The increasingly popular discipline called "landscape genetics" aims to explain these genetic patterns using ecological data (Manel *et al.* 2003; Manel & Holderegger 2013). This discipline integrates population genetics, landscape ecology and spatial statistics to investigate how and which landscape variables shape genetic diversity and genetic structure (Storfer *et al.* 2007). Several recent studies have described landscape features which limit gene flow such as rivers, roads or anthropogenic habitat fragmentation in distinct species (reviewed in Storfer *et al.* 2010). Also the topographic relief of mountains is a typical factor influencing gene flow in various species at least at broad spatial scales (Storfer *et al.* 2010), including plants (Trénel *et al.* 2008). However, depending on the species, the effects of these barriers can be reversed, as e.g. rivers can have an enhancing effect on gene flow for invasive species (Walker *et al.* 2009).

Furthermore, divergent selection of different habitats leads to local adaptation. Loci under selection show stronger divergence (outlier loci) than do neutral loci or loci under balancing selection. However, adaptive divergence can also promote genome-wide differentiation due to general barriers to gene flow which enhance genetic drift and lead to isolation by adaptation (IBA, Nosil *et al.* 2007; Andrew *et al.* 2012). This can cause ecotype divergence with phenotypic and neutral genetic differentiation, as it was described for walking-sticks (*Timema cristinae*) showing preference for distinct host plants (Nosil *et al.* 2007) and for sunflowers (*Helianthus petiolaris*) growing in distinct habitats (sand dune versus non-dune; Andrew *et al.* 2012). These kinds of patterns can be difficult to distinguish from neutral processes such as isolation by distance (IBD). Depending on the scale, IBD and IBA can cause population or fine-scale spatial genetic structure (SGS). Landscape genetic studies help to disentangle the underlying processes by identifying landscape features related to the genetic structure. Under IBD, a spatial autocorrelation pattern is predicted, that is distant populations or individuals are expected to be genetically more differentiated than closer ones. In contrast, under IBA a correlation of environmental factors causing the divergence and genetic distance is expected, which should still be apparent when controlling for the effect of geographic distance. Furthermore, IBA is expected to cause phenotypic divergence of adaptive plant traits due to the same

environmental factors which shape the genetic structure (Nosil *et al.* 2007).

So far, most landscape genetic studies have been conducted on regional to continental scales because gene flow is expected to overwhelm selection at small spatial scales. A recent study suggested that any signature of landscape effects will be masked by immigrating gene flow if the study area is too small relative to the scale of gene flow (Cushman *et al.* 2013). In fact, when gene flow exceeds the migration–selection equilibrium the probability of local adaptation and its perpetuation decrease (Bridle & Vines 2007; Yeaman & Otto 2011). However, plants can show selective divergence at very short spatial scales, for instance at scales of tens of meters in herbaceous species and of hundreds of meters in woody species (reviewed in Linhart & Grant 1996). Moderate gene flow is expected to increase adaptation rates due to the emergence of new alleles exposed to heterogeneous environments, which are then subjected to selection (Kremer & Le Corre 2012). Furthermore, SGS due to isolation by distance can facilitate local adaptation caused by locally structured selection (Audigeos *et al.* 2013). Additionally, the high number of genetic markers available nowadays enhances the statistical power to detect genetic structure (Kohn *et al.* 2006). In effect, recent studies revealed genetic structure at fine spatial scales despite the presence of gene flow (e.g. Van Heerwaarden *et al.* 2010, Brousseau 2013; Audigeos *et al.* 2013). For example, Brousseau (2013) and Audigeos *et al.* (2013) found divergent selection acting between patches of *Eperua falcata* (Fabaceae) growing in seasonally flooded bottomlands and seasonally dry *terra firme* soils. Manel *et al.* (2010) studied genetic associations at local to regional scales in *Arabis alpina* and found different environmental factors associated to putatively adaptive loci at distinct geographic scales. In the present study, we aim to identify environmental factors shaping the genetic structure at short spatial scales within three natural populations of maritime pine (*Pinus pinaster*) from eastern Spain. A previous study revealed a significant fine-scale spatial genetic structure (SGS) in two of the three stands, which was hypothesized to be related to micro-environmental selection (Chapter 3). In the present study, we use approaches to detect environmental factors correlated with genetic variability which control for confounding effects, such as spatial autocorrelation (reviewed in Balkenhol *et al.* 2009).

Maritime pine is pollinated and dispersed by wind. Pollen flow is wide-ranging, fitting highly leptokurtic dispersal kernels with average dispersal distances between 78

and 174 m (de-Lucas *et al.* 2008). Gene flow via seeds is somewhat more restricted with nonleptokurtic dispersal kernels with average dispersal distance between 40 to 60 m (González-Martínez *et al.* 2002). Typically, selection pressures are strong in Mediterranean ecosystems and especially act on early life stages. We hypothesized that micro-environmental conditions probably related to light and water availability (with south facing, high-altitude sites at steep slopes as driest, but north facing, low-altitude sites with gentle slopes as most humid) might shape the genetic and phenotypic structure within populations. A combination of abiotic environmental factors, such as altitude, slope and aspect, were considered. Furthermore, biotic indicators such as herbaceous vegetation cover or herbaceous species richness might distinguish between patches with different water and nutrient availability, and these indicators were also included in the analyses. We aimed at testing whether it is possible to detect an effect of these environmental factors on SNP genotypes and adaptive phenotypes (such as water use efficiency as tackled by carbon isotope discrimination ($\delta^{13}C$), wood density and growth of tree rings) at small spatial scale. It was hypothesized, that the probability of detecting significant environmental effects increases with the degree of environmental heterogeneity in a natural stand. A number of different statistical methods were employed to address the following questions:

1. Can IBA be detected within populations where IBD is present by controlling for geographic distance?
2. Which environmental factors are correlated to the genetic structure of SNP markers?
3. Do the same environmental factors also shape the phenotypic variability?

Material and methods

Study populations

Three natural stands of *P. pinaster* in the East of the Iberian Peninsula were sampled. These stands have been previously studied and described in (Hernández-Serrano 2013; Budde *et al.* 2014, Chapter 3). The environmental variability in the study areas ranges from a plane site (Sinarcas) over a topographically intermediate site with a short altitude gradient but steep slopes (Serra Calderona) to a topographically more diverse site with an altitudinal gradient of 300 m and variable slopes (Eslida; Table 1). All stands grow on siliceous soils. Serra Calderona and Eslida are located close to the Mediterranean Sea and have a warm and

dry climate where crown fires are common. Sinarcas is located further inland above 800 m asl, with mountainous Mediterranean vegetation and sub-humid climate where crown fires are rare. The sites differ in their expression of fire-adaptive traits (Hernández-Serrano *et al.* 2013). However, all stands belong to the same gene pool (Budde *et al.* 2014, Chapter 3) and do not differ in terms of genetic diversity or demographic history; hence no fire related effect could be detected in the population genetics of *P. pinaster* in this region (Chapter 3). Fine scale spatial genetic structure (SGS) was insignificant as calculated on 11 nuclear microsatellites, but significant for 251 single-nucleotide polymorphisms (SNPs) in two sites (Sinarcas and Eslida). Therefore, enhanced SGS for functional markers due to micro-environmental selection was hypothesized in these sites (Chapter 3).

Environmental variables and phenotypic data

A total of 443 trees from the three stands was sampled (139 from Eslida, 155 from Sinarcas and 149 from Serra Calderona). At each tree, GPS coordinates and altitude were taken using a Garmin Oregon 550t (Garmin, Wichita, USA), the slope was measured using a Bitterlich-relascope (Relaskop-Technik, Salzburg, Austria), and the aspect in direction of the main slope was determined with a compass. Furthermore, plant cover was estimated in a 5 m radius around each tree and herbaceous and woody species were reported. From each tree a wood core was extracted at breast height using a Trephor (VITZANI, Belluno, Italy). The cores were stored in fungicide liquid (Fungel, Brimel, Musseros, Spain) in 2 ml SafeSeal Microcentrifuge tubes (Sorenson BioScience, Salt Lake City, USA). These wood samples were kept at constant humidity. The total length (L) and mean diameter (D; averaged over four diameter measurements along each sample) of each wood core were determined with a digital caliper, avoiding pressure of the caliper blades on the wood. Green volume was calculated by multiplying π with D and L. Oven-dry weight of each wood core was measured after drying in a well-ventilated oven at 100ºC during 48h. Subsequently, wood density was calculated as the ratio of the oven-dry weight of a wood core divided by its' green volume.

All dry wood cores were sanded to increase the visibility of the growth zones. High-resolution photos were taken with an incorporated camera in a binocular (Leica LED 2500 80x, Leica Microsystems, Wetzlar, Germany). The five most recent growth rings were measured on each

sample with the image analysis software ImageJ 1.47.

For carbon isotope discrimination ($\delta^{13}C$), last years' needles from all individuals from the south facing part of the tree crown were collected and oven dried at 80ºC. Oven-dried samples were ground to a fine powder using a ball mill. Sub-samples of about 2 mg were used for isotopic determinations by means of GC-combustion isotope ratio mass spectrometry (GC-C-IRMS). Specifically, analysis were performed using a Thermo GC/C-IRMS system composed of a Trace GC Ultra gas chromatograph (Thermo Electron Corp., Milan, Italy) coupled to a Delta V Advantage isotope ratio mass spectrometer through a GC/C-III interface (Thermo Electron Corp., Bremen, Germany).

384-plex SNP assay and genotyping

Needles were collected from each tree and desiccated using silica gel. Genomic DNA was isolated using the Invisorb® DNA Plant HTS 96 Kit/C kit (Invitek GmbH, Berlin, Germany).

Genotyping was performed with Illumina VeraCode® technology for a 384-plex SNP Oligo Pool Assay, OPA (design file see Budde *et al.* 2014). This OPA is based on a subsample of the 1536-plex SNP assay developed for *P. pinaster* by Chancerel *et al.* (2011). For more details on the loci included in this SNP assay see Budde *et al*. (2014). First the genotyping was conducted for samples from Serra Calderona and Sinarcas, yielding 240 and 251 SNPs, respectively. In a second genotyping run, 313 SNPs could be obtained for the individuals from Eslida.

Environmental determinants and genetic structure

The environmental variables in this study were only weakly correlated (all Spearman correlation coefficients < 0.6), therefore all variables were included in the analyses. The aspect was *cosine* transformed to indicate northness, and all environmental variables were standardized (mean=0, variance=1). All analyses were conducted in R version 2.15.0 (R Development Core Team 2012).

Mantel tests and multiple regression of distance matrices (MRDM)

First, partial Mantel tests (Smouse *et al.* 1986) were conducted as implemented in the R-package VEGAN 2.0-7 (Oksanen *et al.* 2013). SNP genotype data was read as population object using the GSTUDIO 0.8 package (Dyer 2012) and pairwise genetic distances were calculated with the AMOVA inter-individual option of this package. Euclidean distances between pairs of samples were calculated for each

environmental variable. A partial Mantel test was conducted correlating the Euclidean distance matrix of each environmental factor and the genetic distance matrix controlling for geographic distance, using the Spearman correlation coefficient. Significance was assessed with 10,000 permutations, and to adjust *P*-values due to multiple tests, q-values were calculated using the Benjamini & Hochberg (1995) procedure.

Mantel tests are commonly used in landscape genetics (see e.g. Balkenhol *et al.* 2009; Andrew *et al.* 2012). Although, we are aware that they have been discussed controversially (see e.g. Legendre & Fortin 2010; Guillot & Rousset 2013). Here we combine this approach with other methods to avoid method-dependent results, as it was suggested in a review on suitable methods to identify genetic structure in a landscape context (Balkenhol et al. 2009).

Subsequently, a multiple regression of distance matrices (MRDM, Legendre *et al.* 1994) was carried out using the vectors of the previously mentioned distance matrices as input data. The "*mrm"* function implemented in the R- package ECODIST (Goslee & Urban 2007) was used, and the backward and forward selection procedures described by Legendre *et al.* (1994) were employed to choose the best set of predictor variables. Only variables identified by both selection procedures are reported. Levels of significance were assessed by 10,000 permutations. To adjust *P*-values of the standardized regression coefficients of the environmental variables due to multiple tests, q-values were calculated using the Benjamini & Hochberg (1995) procedure.

Redundancy Analysis (RDA) and variance partitioning

Furthermore, a constrained ordination method was conducted. The RDA is a multivariate analogue of regression and assumes a linear relationship when describing gradients in the first (dependent) set of variables (SNP genotypes) in terms of the second (explanatory) variable set (environmental data; Legendre & Legendre 1998). This analysis was conducted using the R-package VEGAN 2.0-7. First, an RDA was performed including all environmental variables, and the best set of explanatory variables was selected using the "ordistep" function and the backward selection procedure implemented in VEGAN. Only environmental variables that were identified as significant with both procedures are reported. Significance was assessed using 10,000 permutations. Variance partitioning can be performed for different factors or groups of environmental factors (Leyer & Wesche 2007). It was conducted for the RDA

following recognized procedures (Borcard *et al.* 1992; Liu 1997; Økland 1999). The explainable genetic variance (i.e. inertia, or in this case mean squared contingency coefficient) was partitioned into percentages of variance explained by the best set of environmental variables (altitude in Eslida) on the one hand and the geographic coordinates on the other hand.

Phenotypic correlations

To assess the effect of landscape features on phenotypes (water use efficiency as tackled by carbon isotope discrimination ($\delta^{13}C$), wood density and growth), multiple linear regressions were performed. Using a backward selection procedure by stepwise removing the variable with the highest *P*-value until all variables in the model had *P*-values below 0.05. At each step, the significance of the removal effect was tested by an ANOVA comparing the previous and the new model. Finally, only significant regressions after multiple test correction following Bonferroni (Rice 1989) are reported.

Results

Partial Mantel tests revealed differences in altitude in Eslida to influence the genetic distance between trees. This was the only environmental variable significantly related to genetic distance (*rho*=0.0489, *P*=0.0219, *Q*=0.1095; Table 2). In the MRDM, by using the backward and forward selection procedure to choose the best model, again altitude in Eslida was identified as significantly related to genetic distance (Table 3), while in Serra Calderona slope and northness (cosine[aspect]) yielded significant correlations, so as herbaceous plant cover in Sinarcas. Finally, the combination of the backward selection procedure and the "ordistep" function applied on the RDA confirmed that altitude had a significant effect in Eslida. In Serra Calderona and Sinarcas, no environmental variable was significantly associated with the genetic structure. In summary, the only environmental variable identified consistently by all methods was altitude in Eslida. Although the partial RDA of SNP genotypes and altitude in Eslida conditioned on the geographic coordinates was only marginally significant (*P*=0.0619), variance partitioning revealed ~ 30% of the explainable variance to be attributed to altitude, ~ 66% to geographic position and ~3 % to both due to colineartity. In a partial RDA using altitude as independent variable to explain the variability in the SNP data after controlling for the geographic position, the first RDA axis reflects the part of the explainable variance associated to altitude and is

visualized in a map (Fig. 1). However, the first three axis of an unconstrained PCA were not significantly related to any environmental variable in this population. Altitude also significantly influenced the growth of *P. pinaster* tree rings in Eslida. For water use efficiency, as tackled by carbon isotope discrimination ($\delta^{13}C$), slope was more important, indicating a significant positive relationship. In Serra Calderona a positive relationship between slope and wood density was found (Table 4).

Discussion

We hypothesized that the probability to detect a significant environmental effect might increase with increasing environmental heterogeneity. In fact, the steepest environmental cline, a 300 m altitudinal gradient was consistently associated with the genotypic pattern. However, it was not any combination of the environmental variables measured as indicators characterizing micro-environmental patches that shaped the genetic structure, but the altitudinal cline in Eslida. Altitudinal gradients are known to shape the population genetic structure of plants (reviewed in Ohsawa & Ide 2008), however, a structuring, although weak, at such a small spatial scale and in a species with high gene flow, is remarkable.

Altitude could reflect gradients related to a number of environmental factors such as a decrease in total atmospheric pressure, temperature reduction, increasing solar radiation, and increasing UV-B radiation with altitude (Körner 2007). Assuming a lapse rate of -0.6 °C per 100 m altitude denotes a temperature difference of ~1.8°C between low and high altitudes in Eslida. Furthermore, soil depth and associated water and nutrient availability probably increase with distance to hilltop in the study area. Additionally, abiotic conditions shape biotic responses. Especially temperature is an important factor triggering the flowering phenology in pine species (Jackson & Sweet 1972) which influences the mating system and the spatial genetic structure. Gauzere *et al.* (2013) found differences in the mating system within and among populations along a ~300 m altitudinal gradient of *Fagus sylvatica,* another wind pollinated species, with seeds at high altitudes being more often the product of immigrating pollen than at low altitudes. Furthermore, they found lags in the flowering phenology of several days between trees at different altitudes. Lags in the flowering phenology might lead to temporally conditioned assortative mating (Gérard *et al.* 2006), which could shape the genetic structure along environmental gradients also in the absence of divergent selection (Soularue & Kremer 2012). However, also selection

at genes related to flowering phenology has been described along altitudinal gradients (Alberto *et al.* 2013).

In Eslida, where all methods consistently indicated altitude to have a significant effect on the genetic structure of *P. pinaster*, latent factor mixed models (LFMM, Frichot *et al.* 2013) were used to identify SNP loci potentially associated with altitude. The number of components K was set to 1 as all individuals belong to the same population. After adjusting *P*-values from LFMM for multiple testing, no SNP was significantly associated with altitude in Eslida (see Supplementary Information Table S1). Therefore, we did not detect any loci under selection, but a general signal of limited gene flow along the altitudinal gradient.

Furthermore, *P. pinaster* in Eslida grows in a region with a high frequency of stand-replacing fires and exhibits serotinous cones (Hernández-Serrano *et al.* 2013; Budde *et al.* 2014, Chapter 3). This cone type remains closed and forms a canopy seed bank until high temperatures (e.g. during a fire event) trigger the opening and seed release (Lamont *et al.* 1991). Synchronized seed release after fire events might facilitate selection acting on a high diversity of genotypes covering the ground in each micro-environmental patch. Selection is strong on early live stages in these populations (Vizcaíno-Palomar *et al.* submitted). Soil depth, water and nutrient availability and solar radiation are important factors in early life stages and determine the successful establishment of seedlings in micro-sites (Gómez-Aparicio 2008). Hence, we would expect stronger selection at high altitudes and stony slopes due to harsher conditions for young stages.

No correlation between genetic structure and environment was found in Serra Calderona and Sinarcas. We assume that environmental gradients in these populations were too weak to generate a detectable pattern. Hence, gene flow might exceed the migration–selection equilibrium in these stands, or the relevant environmental variables have not been measured in this study.

Altitude also significantly influenced the growth of tree rings in *P. pinaster* in Eslida, while $\delta^{13}C$ was significantly related to slope in this stand. Growth of tree rings was positively affected by altitude and might reflect less stressful conditions, at least for adult life stages, in upper elevations. Typically, altitude has a negative effect on carbon isotope discrimination and growth in other conifer species (Warren *et al.* 2001; Liang *et al.* 2010), at least along higher and longer altitudinal gradients on broader scales. Water use efficiency (as evaluated by $\delta^{13}C$) was positively correlated with slope which indicates a more efficient use of water at steeper slopes. In Serra

Calderona, slope also significantly affected wood density. As it had to be expected, trees at shallow slopes exhibited significantly less wood density while trees at steeper slopes had higher wood density probably due, at least in part, to slower growth. The phenotypic realtionship to environmental factors is probably a plastic response to different environmental conditions, although given our results based on potentially functional markers, genetic adaptation cannot be fully excluded.

Future studies should aim to characterize environmental conditions around each sampled tree in more detail. Factors that might be important and should be measured are soil depth, soil pH, soil type, nutrient availability, carbon content, grain size fractions, soil water storage capacity or microbial activity. Also a paired sampling in patches of contrasting environments, high vs. low altitudes, dry vs. humid patches etc., might be an elegant approach to characterize environmental contrasts (Fournier *et al.* 2006; Audigeos *et al.* 2013). In any case, pronounced environmental gradients are more likely to produce an effect on the genetic structure of plant populations. Furthermore, phenotypes such as bud burst or flowering time could contribute to a better understanding of the processes contributing to shape the genetic structure.

Our results revealed a significant structure along an altitudinal gradient at short spatial scale in a species with wide ranging gene flow. Different processes, such as assortative mating or divergent selection can lead to such a pattern, and future studies should try to disentangle the underlying forces. These findings are in line with recent studies that revealed divergent selection between patches of contrasting environmental conditions in forest trees (Brousseau 2013; Audigeos *et al.* 2013) and herbaceous species (Manel *et al.* 2010; Fischer *et al.* 2013). The increasing number of available genetic markers, also in non-model species, will improve the statistical power to detect such patterns at local scales. The importance of genetic variability and the strength of selection at small spatial scales have probably been underestimated so far. Especially with respect to climate change, the knowledge about genetic variation and processes that shape the genetic structure at different geographic scales are of utmost importance to develop suitable conservation strategies.

Acknowledgements

We thank Fernando del Caño, Delphine Grivet, Mario Zabal, Natalia Vizcaíno Palomar, Enrique Sáez Laguna, Diana Turrión and Miguel Navarro for field

assistance and Carmen García-Barriga (INIA-CIFOR) for assistance in the laboratory. The present study was funded by the following projects (Spanish Ministry of Science and Innovation, MICINN): VAMPIRO (CGL2008-05289-C02-01/02), LinkTree (EUI2008-03713, under the ERAnet-BiodivERsA call), and TREVOL (CGL2012-39938-C02-00/01). KBB acknowledges a PhD scholarship (FPI: BES-2009-015443) of the MICINN and an exchange grant of the European Science Foundation (ESF) within the framework of the ESF activity entitled 'Conservation Genomics: Amalgamation of Conservation Genetics and Ecological and Evolutionary Genomics'. MH is a 'Ramón y Cajal' researcher of the MICINN and acknowledges funding from JAE-Doc program and the Spanish Ministry for Economy and Competitiveness (CGL2012-40129-C02-02/ AFFLORA).

References

Alberto FJ, Derory J, Boury C *et al.* (2013) Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics*, **195**, 495–512.

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, **21**, 2078–2091.

Audigeos D, Brousseau L, Traissac S, Scotti-Saintagne C, Scotti I (2013) Molecular divergence in tropical tree populations occupying environmental mosaics. *Journal of Evolutionary Biology*, **26**, 529–544.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, **32**, 818–830.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, **57**, 289–300.

Bergmann F (1978) The allelic distribution at an acid phosphatase locus in Norway spruce (*Picea abies*). *Theoretical and Applied Genetics*, **52**, 57–64.

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

Bridle JR, Vines TH (2007) Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, **22**, 140–147.

Brousseau L (2013) Diversity and evolution in tropical rainforest trees - Example of *Eperua falcata* in French Guiana. University of Lorraine.

Budde KB, Heuertz M, Hernández-Serrano A *et al.* (2014) *In situ* genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). *New Phytologist*, **201**, 230–241.

Chancerel E, Lepoittevin C, Le Provost G *et al.* (2011) Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine and

comparative mapping with loblolly pine. *BMC Genomics*, **12**, 368.

Cushman SA, Shirk AJ, Landguth EL (2013) Landscape genetics and limiting factors. *Conservation Genetics*, **14**, 263–274.

de-Lucas A, Robledo-Arnuncio JJ, Hidalgo E, González-Martínez SC (2008) Mating system and pollen gene flow in Mediterranean maritime pine. *Heredity*, **100**, 390–399.

Dyer RJ (2012) gstudio: GeneticStudio packages for R. R package version 0.8., http://CRAN.R–project.org/package=gstudio.

Epperson BK (2003) *Geographical Genetics*. Princeton University Press, Princeton, NJ.

Fischer MC, Rellstab C, Tedder A *et al.* (2013) Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, **22**, 5594–5607.

Fournier N, Rigling A, Dobbertin M, Gugerli F (2006) Faible différenciation génétique , à partir d ' amplification aléatoire d ' ADN polymorphe (RAPD), entre les types de pin sylvestre (*Pinus sylvestris* L .) d ' altitude et de plaine dans les Alpes à climat continental. *Annals of Forest Science*, **63**, 431–439.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using Latent Factor Mixed Models. *Molecular Biology and Evolution*, **30**, 1687–1699.

Gauzere J, Klein EK, Oddou-Muratorio S (2013) Ecological determinants of mating system within and between three *Fagus sylvatica* populations along an elevational gradient. *Molecular Ecology*, **22**, 5001–5015.

Gérard PR, Klein EK, Austerlitz F, Fernández-Manjarrés JF, Frascaria-Lacoste N (2006) Assortative mating and differential male mating success in an ash hybrid zone population. *BMC Evolutionary Biology*, **6**, 96.

Gómez-Aparicio L (2008) Spatial patterns of recruitment in Mediterranean plant species: linking the fate of seeds, seedlings and saplings in heterogeneous landscapes at different scales. *Journal of Ecology*, **96**, 1128–1140.

González-Martínez SC, Gerber S, Cervera MT *et al.* (2002) Seed gene flow and fine-scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. *Theoretical and Applied Genetics*, **104**, 1290–1297.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods in Ecology and Evolution*, **4**, 336–344.

Van Heerwaarden J, Ross-Ibarra J, Doebley JF *et al.* (2010) Fine scale genetic structure in the wild ancestor of maize (*Zea mays ssp. parviglumis*). *Molecular Ecology*, **19**, 1162–1173.

Hernández-Serrano A (2013) Fire structures pine serotiny at different scales. *American Journal of Botany*, **100**, 2349–2356.

Hernández-Serrano A, Verdú M, González-Martínez SC, Pausas JG (2013) Fire structures pine serotiny at different scales. *American Journal of Botany*, **100**, 1–8.

Heywood J (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics*, **22**, 335–355.

Jackson D, Sweet G (1972) Flower initiation in temperate woody plants. *Horticultural Abstracts*, **42**, 9–24.

Jaramillo-Correa JP, Grivet D, Terrab A *et al.* (2010) The Strait of Gibraltar as a major biogeographic barrier in Mediterranean conifers: a comparative phylogeographic survey. *Molecular Ecology*, **19**, 5452–5468.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution*, **21**, 629–637.

Körner C (2007) The use of “altitude” in ecological research. *Trends in Ecology & Evolution*, **22**, 569–574.

Kremer A, Le Corre V (2012) Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity*, **108**, 375–385.

Lamont B, Maitre D Le, Cowling RM, Enright NJ (1991) Canopy seed storage in woody plants. *The Botanical Review*, **5**, 277–317.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Legendre P, Lapointe F, Casgrain P (1994) Modeling brain evolution from behavior: a permutational regression approach. *Evolution*, **48**, 1487–1499.

Legendre P, Legendre L (1998) *Numerical Ecology*. Elsevier, Amsterdam, Netherlands.

Leyer I, Wesche K (2007) *Multivariate Statistik in der Ökologie*. Springer-Verlag GmbH, Berlin, Germany.

Liang E, Wang Y, Xu Y, Liu B, Shao X (2010) Growth variation in *Abies georgei* var. *smithii* along altitudinal gradients in the Sygera Mountains, southeastern Tibetan Plateau. *Trees*, **24**, 363–373.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, **27**, 237–277.

Liu Q (1997) Variation Partitioning by Partial Redundancy Analysis (RDA). *Environmetrics*, **8**, 75–85.

Loveless M, Hamrick J (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, **15**, 65–95.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*, **28**, 614–621.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Molecular Ecology*, **19**, 3824–35.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Nosil P, Egan SP, Funk DJ (2007) Heterogeneous genomic differentiation between walking-stick

ecotypes: “isolation by adaptation” and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Ohsawa T, Ide Y (2008) Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography*, **17**, 152–163.

Økland RH (1999) On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science*, **10**, 131–136.

Oksanen J, Blanchet FG, Kindt R *et al.* (2013) vegan: Community Ecology Package. R package version 2.0-7., http://CRAN.R–project.org/package=vegan.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Smouse P, Long J, Sokal R (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

Soularue J-P, Kremer A (2012) Assortative mating and gene flow generate clinal phenological variation in trees. *BMC evolutionary biology*, **12**, 79.

Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the “landscape” in landscape genetics. *Heredity*, **98**, 128–42.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–514.

Trénel P, Hansen MM, Normand S, Borchsenius F (2008) Landscape genetics, historical isolation and cross-Andean gene flow in the wax palm, *Ceroxylon echinulatum* (Arecaceae). *Molecular Ecology*, **17**, 3528–3540.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Walker NF, Hulme PE, Hoelzel AR (2009) Population genetics of an invasive riparian species, *Impatiens glandulifera*. *Plant Ecology*, **203**, 243–252.

Warren CR, McGrath JF, Adams MA (2001) Water availability and carbon isotope discrimination in conifers. *Oecologia*, **127**, 476–486.

Wright S (1949) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*, **65**, 2123–2129.

Author contributions

KBB performed the study, did most data analyses and wrote the first draft of the manuscript; MH, JGP, MV and SGCM designed the study; FG contributed to data analyses; AHS produced the phenotypic data, SGCM contributed to manuscript writing. All authors critically read and revised different versions of the manuscript.

Tables

Table 1 Characterization of the study stands of *Pinus pinaster* in terms of environmental variables and phenotypes. Values in brackets indicate the standard deviation of the variable in the respective column. *N*, number of samples; DBH, diameter at breast height, $\delta^{13}C$, stable carbon isotope ratio. NA, data not available because it is a plain site.

Stand	Latitude, Longitude [°]	*N*	Altitude [m]	Aspect [°]	Slope [%]	Plant cover [%]	Species richness	DBH [cm]	Wood density [g/mm^3]	$\delta^{13}C$	Growth [mm/year]
Sinarcas	39.79, -1.20	157	860-897	NA	NA	70 (18.63)	4-12	32.18 (4.17)	0.00044 (0.00004)	-26.74029 (0.61236)	1.17706 (0.3976)
Serra Calderona	39.75, -0.50	149	710-864	60 (93.15)	60.70 (15.22)	45.38 (15.22)	0-11	26.55 (4.53)	0.00043 (0.00005)	-26.37851 (0.77154)	1.2117 (0.54684)
Eslida	39.88, -0.30	139	392-734	261 (95.15)	52.41 (17.74)	30.99 (59)	0-8	30.04 (6.32)	0.00047 (0.00005)	-27.10577 (0.85017)	1.30352 (0.51691)

Table 2 Results of the Mantel and partial Mantel tests (controlling for geographic distance) of distance matrices for the environmental variables and SNP data in the three natural *Pinus pinaster* stands. Spearman correlation coefficient (*rho*), *P*-values from permutation tests and, for Eslida, significance levels after multiple test correction (*Q*-values) are reported.

	Eslida			Calderona		Sinarcas	
	rho	*P*	*Q*	*rho*	*P*	*rho*	*P*
Mantel tests							
gen. dist. ~ geo. dist.	0.1023	0.0008	0.0042	-0.0074	0.6344	0.0044	0.4001
gen. dist. ~ cos(aspect)	0.0012	0.4793	0.4793	0.0324	0.1523		
gen. dist. ~ altitude	0.1112	0.0014	0.0042	-0.0003	0.5023	0.0110	0.3161
gen. dist. ~ slope	0.0286	0.2167	0.3716	-0.0445	0.9030		
gen. dist. ~ plant cov.	0.0172	0.2956	0.3716	-0.0123	0.7087	-0.0549	0.973
gen. dist. ~ spec. rich	0.0168	0.3097	0.3716	-0.0159	0.6775	-0.0214	0.7364
Partial Mantel tests							
gen. dist. ~ cos(aspect) | geo. dist.	0.0011	0.477	0.5963	0.0332	0.1439		
gen. dist. ~ altitude | geo. dist.	0.0489	0.0219	0.1095	0.0006	0.4903	0.0102	0.3510
gen. dist. ~ slope | geo. dist.	0.0149	0.3335	0.5558	-0.0439	0.8992		
gen. dist. ~ plant cov. | geo. dist.	-0.0124	0.6545	0.6545	-0.0123	0.7126	-0.0553	0.9756
gen. dist. ~ spec. rich | geo. dist.	0.019	0.2796	0.5558	-0.0159	0.6833	-0.0218	0.7417

Table 3 Results of the multiple regressions of distance matrices to reveal environmental variables associated with genetic distance in *Pinus pinaster*. Genetic distance was the dependent variable and the distance matrices of the environmental variables were the independent variables. Results accounting for effects of geographic distance are shown on the right side. Standardized regression coefficients, *b*; coefficients of determination, R^2 and *F*-values are reported. To account for multiple test correction, the *P*-values of the standardized regression coefficients of the respective environmental variables were used to calculate *Q*-values. Environmental factors significantly associated with genetic distance identified by backward and forward selection procedures are marked in bold.

	gen. dist. ~ environment					gen. dist. ~ environment + geo. dist.							
						geo.dist.		environment					
Variable	*b*	*P*	R^2	*F*	*P*	*b*	*P*	*b*	*P*	*Q*	R^2	*F*	*P*
Eslida													
geo. dist.	0.1023	0.0001	0.0105	101.45	0.0001								
cos(aspect)	0.0012	0.9182	0.0000	0.0148	0.9182	0.1023	0.0001	0.0011	0.9254	0.9254	0.0105	50.727	0.0001
altitude	**0.1111**	**0.0001**	**0.0124**	**119.93**	**0.0001**	**0.0367**	**0.1645**	**0.0816**	**0.0038**	**0.0190**	**0.0128**	**62.305**	**0.0001**
slope	0.0287	0.0321	0.0008	7.8445	0.0321	0.1003	0.0001	0.0150	0.2774	0.4623	0.0107	51.796	0.0001
plant cov.	0.0172	0.3290	0.0003	2.8460	0.3290	0.106	0.0001	-0.0128	0.4480	0.5600	0.0106	51.462	0.0001
spec rich.	0.0172	0.1747	0.0003	2.7219	0.1747	0.1027	0.0001	0.0194	0.1287	0.3218	0.0108	52.476	0.0001
Serra Calderona													
geo. dist.	-0.0074	0.507	0.0001	0.5959	0.507								
cos(aspect)	**0.0324**	**0.0033**	**0.0011**	**11.607**	**0.0033**	**-0.0103**	**0.3482**	**0.0334**	**0.0023**	**0.0058**	**0.0012**	**6.3873**	**0.0091**
altitude	-0.0003	0.9812	<0.0001	0.0007	0.9812	-0.0074	0.5044	0.0006	0.9584	0.9584	0.0001	0.3000	0.7930
slope	**-0.0447**	**0.0003**	**0.0020**	**21.898**	**0.0003**	**-0.0015**	**0.8917**	**-0.0445**	**0.0004**	**0.0020**	**0.0020**	**10.9600**	**0.0007**
plant cov.	-0.0123	0.2314	0.0002	1.6597	0.2314	-0.0074	0.5026	-0.0123	0.2254	0.2818	0.0002	1.1278	0.3979
spec rich.	-0.0161	0.1616	0.0003	2.7943	0.1616	-0.0074	0.5071	-0.0161	0.156	0.2600	0.0003	1.6994	0.2874
Sinarcas													
geo. dist.	0.0057	0.6589	<0.0001	0.4023	0.6589								
altitude	0.0144	0.2305	0.0002	2.5404	0.2305	-0.0027	0.8588	0.0159	0.2454	0.2454	0.0002	1.3031	0.4965
plant cov.	**-0.0553**	**0.0001**	**0.0030**	**37.345**	**0.0001**	**0.0086**	**0.497**	**-0.0557**	**0.0001**	**0.0003**	**0.0031**	**19.122**	**0.0001**
spec rich.	-0.0462	0.0001	0.0021	25.249	0.0001	0.0060	0.622	-0.0464	0.0002	0.0003	0.0021	12.841	0.0007

Table 4 Linear regressions of environmental factors and phenotypes of *Pinus pinaster*. Only environmental variables identified by backward selection as significantly related to phenotypes and with *P*-values below Bonferroni corrected *P*-value ($P_{Bonferroni}$= 0.0167) due to multiple test correction are shown. Standardized regression coefficients, *b*; coefficients of determination, R^2 and *F*-values are reported. S. Calderona = Serra Calderona

	Env. variable	*b*	adj. R^2	DF	*P*
$\delta^{13}C$					
Eslida	Slope	0.3375	0.1028	137	0.0001
Wood density					
S. Calderona	Slope	0.2544	0.0559	144	0.0024
Growth					
Eslida	Altitude	0.4404	0.1834	81	<0.0001

Figures

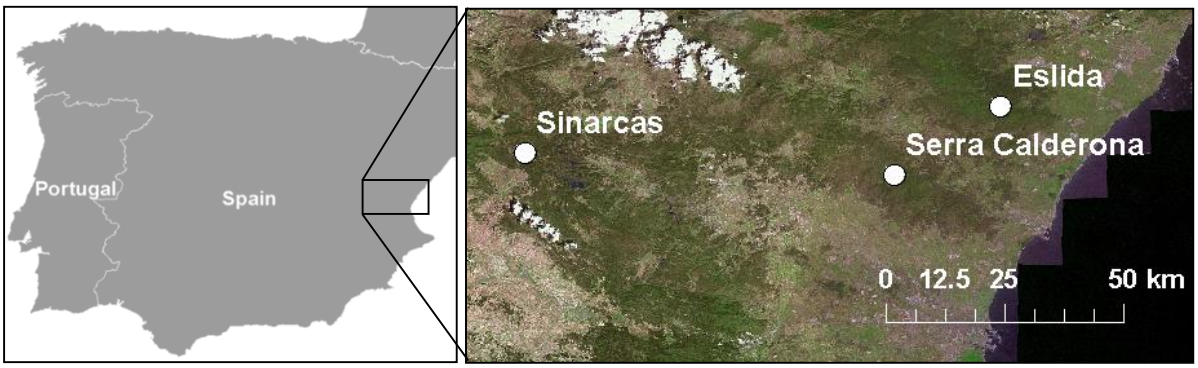

Figure 1 Location of the sample stands of *Pinus pinaster* in the eastern part of the Iberian Peninsula.

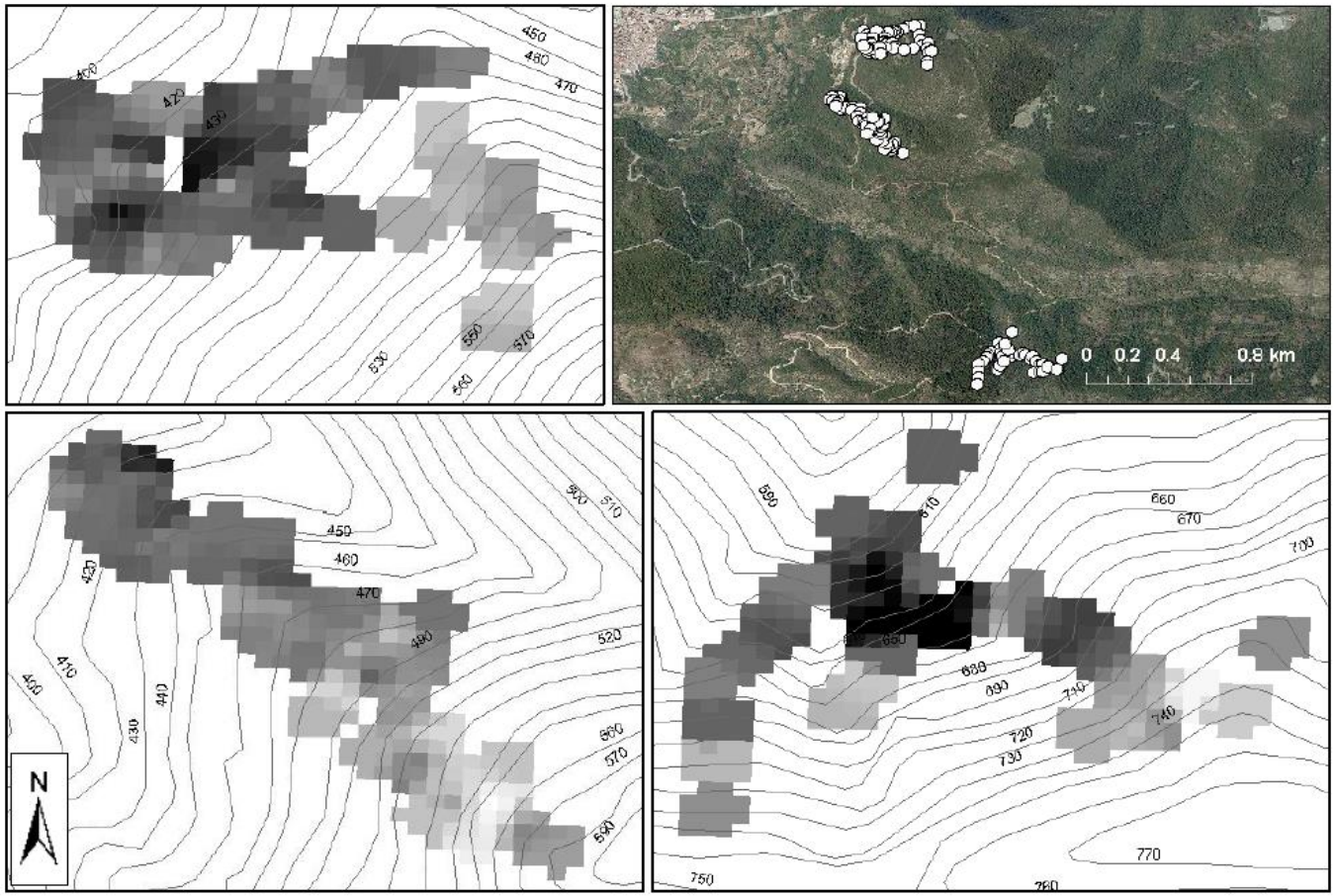

Figure 2 Map of three sample patches in a natural stand of *Pinus pinaster* in Eslida (above right) and detailed maps of the three patches with interpolation of the first axes of a partial redundancy analysis (pRDA) with altitude as independent variable, SNP genotypes as dependent variable conditioned by the UTM coordinates. A trend from lower to higher altitude is visible in all three sample patches. The maximum distance between sample points is ~1.8 km.

SECTION III: Molecular basis of local adaptation in nature

Chapter 5

In situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*)

This chapter reproduces explicitly the following publication:

Budde KB, Heuertz M, Hernández-Serrano A, Pausas J, Vendramin GG, Verdú M, González-Martínez SC (2014)*In-situ* genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster* Aiton). *New Phytologist*, **201**, 230-241.

New Phytologist

Research

In situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*)

Katharina B. Budde[1], Myriam Heuertz[1], Ana Hernández-Serrano[2], Juli G. Pausas[2], Giovanni G. Vendramin[3], Miguel Verdú[2] and Santiago C. González-Martínez[1]

[1]Department of Forest Ecology and Genetics, INIA Forest Research Centre, 28040, Madrid, Spain; [2]Centro de Investigaciones sobre Desertificación (CIDE-CSIC/UV/GV), 46113, Moncada, Valencia, Spain; [3]Plant Genetics Institute, National Research Council, 50019, Sesto Fiorentino, Florence, Italy

Author for correspondence:
Santiago C. González-Martínez
Tel: +34 913471499
Email: santiago@inia.es

Received: *11 June 2013*
Accepted: *28 July 2013*

New Phytologist (2013)
doi: 10.1111/nph.12483

Key words: association genetics, ecologically relevant traits, forest fires, Mediterranean pines, serotiny, single nucleotide polymorphisms (SNPs).

Summary

• Wildfire is a major ecological driver of plant evolution. Understanding the genetic basis of plant adaptation to wildfire is crucial, because impending climate change will involve fire regime changes worldwide. We studied the molecular genetic basis of serotiny, a fire-related trait, in Mediterranean maritime pine using association genetics.

• A single nucleotide polymorphism (SNP) set was used to identify genotype : phenotype associations *in situ* in an unstructured natural population of maritime pine (eastern Iberian Peninsula) under a mixed-effects model framework. RR-BLUP was used to build predictive models for serotiny in this region. Model prediction power outside the focal region was tested using independent range-wide serotiny data.

• Seventeen SNPs were potentially associated with serotiny, explaining approximately 29% of the trait phenotypic variation in the eastern Iberian Peninsula. Similar prediction power was found for nearby geographical regions from the same maternal lineage, but not for other genetic lineages.

• Association genetics for ecologically relevant traits evaluated *in situ* is an attractive approach for forest trees provided that traits are under strong genetic control and populations are unstructured, with large phenotypic variability. This will help to extend the research focus to ecological keystone non-model species in their natural environments, where polymorphisms acquired their adaptive value.

Introduction

Wildfires have a long history in shaping natural ecosystems (Pausas & Keeley, 2009), and are a characteristic feature in many regions of the Mediterranean Basin (as reviewed by Pausas *et al.*, 2008). Understanding the genetic basis of plant adaptation to wildfire is especially important, because impending climate change will involve fire regime changes world-wide (Mouillot & Field, 2005; Krawchuk *et al.*, 2009). In the Mediterranean Basin, forest fire frequency and size increased significantly during the last century (Piñol *et al.*, 1998; Pausas, 2004; Pausas & Fernández-Muñoz, 2012), and are expected to increase further in the near future (Mouillot *et al.*, 2002). These new fire regimes may fall outside of the historic variability, creating new selective pressures on plant populations.

Plant populations confronted with new environments, including new fire regimes, will migrate, adapt locally or go extinct (Aitken *et al.*, 2008). As migration rates for most plant species appear to be insufficient to track the rapid environmental shifts predicted from climate change models (Malcolm *et al.*, 2002; McLachlan *et al.*, 2005), long-term persistence will mostly rely on *in situ* adaptation (Hoffmann & Sgrò, 2011). Typically, tree populations harbor high genetic diversity on which selection can act. They also have, generally, little genetic structure because of outcrossed mating systems, high gene flow and large population sizes (Petit & Hampe, 2006). High fecundity and strong selection in early life stages (Le Corre & Kremer, 2003) enable pronounced local adaptation despite strong gene flow (Kawecki & Ebert, 2004). This process can result in divergent natural phenotypes within populations (e.g. Pausas *et al.*, 2012 for flammability). Genetic differentiation at loci underlying these traits is also expected (Howe *et al.*, 2003; Kremer & Le Corre, 2012), which can be explored using association genetic approaches (Neale & Savolainen, 2004).

Fire is a strong selective driver, and there is an emerging view that fire shapes the intraspecific variability of multiple traits (e.g. bark thickness, mature height, self-pruning, age to maturity, serotiny, longevity, flammability) and generates phenotypic variability among plant populations (Keeley *et al.*, 2011; Moreira *et al.*, 2012; Pausas & Schwilk, 2012; Pausas *et al.*, 2012). In particular, multiple phenotypic traits are selected for by stand-replacing crown fire regimes in pines, such as thin bark, absence of self-pruning, early maturity and the presence of serotinous cones (Keeley & Zedler, 1998; Tapias *et al.*, 2001; Keeley *et al.*, 2011;

Keeley, 2012), which results in correlated evolution of fire-related life-history traits in these taxa (Schwilk & Ackerly, 2001; He *et al.*, 2012). In this study, we used serotiny, estimated as the proportion of serotinous cones, as an indicator for multi-trait fire phenotypes in maritime pine (*Pinus pinaster* Aiton), a species living in Mediterranean fire-prone regions. Serotiny refers to the persistence of closed mature cones in the tree canopy until seed release is triggered by high temperatures, such as those that accompany crown fires (Lamont *et al.*, 1991). A simple genetic control (one locus with two alleles) was proposed for serotiny because a fast selection-driven response to fire was detected after only one generation in different pine species (Teich, 1970; Perry & Lotan, 1979). However, more recent evidence points towards a polygenic quantitative trait (e.g. Pike *et al.*, 2010; Parchman *et al.*, 2012). Serotiny is a highly heritable trait in pine species (Perry & Lotan, 1979; Wymore *et al.*, 2011), with narrow-sense heritability (h^2) of 0.20–0.67 (Pike *et al.*, 2010; J. Climent, pers. comm.). Serotiny also shows large phenotypic variation within and among species and populations of Mediterranean pines (e.g. Tapias *et al.*, 2004; He *et al.*, 2012).

Fire-related traits, such as bark thickness, self-pruning and serotiny, are difficult to evaluate in common gardens because they are expressed late in long-lived species. Moreover, natural selection is relaxed when germinating and growing seedlings in optimal glasshouse conditions before trial establishment. An alternative approach, given the high heritability of traits such as serotiny, is to phenotype directly in unstructured natural populations for association studies (e.g. Parchman *et al.*, 2012). Allelic effects are then assessed in exactly the same environment in which they give a selective advantage, without any bias caused by plant manipulation.

Association genetics has been successful in detecting allele effects for adaptive traits in plants (Atwell *et al.*, 2010; Strange *et al.*, 2011), including in some conifers (e.g. González-Martínez *et al.*, 2007; Eckert *et al.*, 2009; Holliday *et al.*, 2010; Cumbie *et al.*, 2011; Westbrook *et al.*, 2013). Recent reports of strong linkage disequilibrium (LD) in non-coding regions of conifer genomes (e.g. in *Cryptomeria japonica*; Moritsuka *et al.*, 2012) have improved previous expectations (see Neale & Savolainen, 2004) of identifying regions of the genome associated with traits of ecological interest using candidate gene approaches in this group of species, despite their large genome sizes (25 227 Mbp in pine; Plant DNA C-values Database, release 5.0, March 2012, http://data.kew.org/cvalues). As a drawback, higher LD could make it more difficult to identify the actual causal polymorphisms within these regions. The use of candidate genes for adaptive traits should also allow the construction of predictive models for adaptive phenotypes with lower genotyping effort. Indeed, marker densities from 2–3 to 10–20 markers/cM, depending on training population size, are considered to be necessary to achieve reasonable accuracy in phenotypic predictive models based solely on random genome sampling and background LD with phenotypic traits (Grattapaglia & Resende, 2011; Resende *et al.*, 2012). However, models based on candidate genes were able to achieve similar prediction power with a much smaller number of molecular markers. For instance, Holliday *et al.* (2012) were able to explain *c.* 28–34% of the phenotypic variance in predictive models for bud set and cold injury based on a set of only 20 loci, albeit carefully selected from expressional candidate genes (Holliday *et al.*, 2010).

Maritime pine (*P. pinaster*) is an iconic Mediterranean conifer that forms large forests in the western Mediterranean Basin. Forest fires appear to be one of the main drivers that have shaped its life history (see Tapias *et al.*, 2004; Keeley, 2012; and references therein) and large differences in fire-related traits are found among populations, such as for serotiny (from zero in Portugal to 73% in Algeria; Tapias *et al.*, 2004; see also Fig. 1). Molecular marker studies have found three completely isolated maternal lineages (based on mitochondrial DNA, mtDNA) in this species: a western lineage (most of the Iberian Peninsula, Atlantic France and Punta Cires in northern Morocco), an eastern lineage (Catalonia in the northeastern Iberian Peninsula, southeastern France, Corsica, Italy, Tunisia and Algeria) and one endemic to Morocco (Burban & Petit, 2003; see also Bucci *et al.*, 2007), as well as several distinct gene pools (based on nuclear markers) within these broad geographical regions (Salvador *et al.*, 2000; Eveno *et al.*, 2008; Santos-del-Blanco *et al.*, 2012). A long history of population isolation, in particular among populations from the different (non-overlapping) maternal lineages, could have resulted in both parallel and lineage-specific adaptations in fire-related traits, as has been shown for other main drivers of tree adaptation (e.g. Prunier *et al.*, 2012 for climate adaptation in *Picea mariana*).

In this study, we used a 384-plex single nucleotide polymorphism (SNP) array (251 successfully scored and polymorphic SNPs) enriched for well-known candidate genes for adaptive traits in forest trees to identify markers potentially associated with serotiny in maritime pine. The study was conducted in an unstructured natural population (Supporting Information Fig. S1; see also Bucci *et al.*, 2007; Santos-del-Blanco *et al.*, 2012) with high phenotypic variability for serotiny (0–100% serotinous cones per tree, average proportion of serotinous cones of 36.29% and standard deviation of 23.36%; Fig. S2; see also Tapias *et al.*, 2004), which is representative of the eastern Iberian maritime pine range. We then used the subset of SNPs potentially associated with serotiny to build a predictive model for fire phenotypes in the sampled region and tested the model accuracy by cross-validation. Finally, to evaluate the utility of the model outside of the geographical range for which it was constructed, we genotyped the same SNPs and tested the model prediction power across range-wide populations, including the three distinct maternal (mtDNA) lineages recognized in maritime pine.

The focus on well-known candidate genes for adaptive traits in pine, the high phenotypic variability in fire phenotypes (including serotiny) in eastern Iberian maritime pine stands, the fact that serotiny is heritable and gauges a multi-trait fire syndrome and the complete lack of population structure in the study region enabled us to successfully assess *in situ* phenotype : genotype associations. These potentially associated SNPs provide insights into a variety of candidate genes that could underlie fire phenotypes in Mediterranean pines and constitute the basis to construct predictive models for fire-related traits of major ecological importance.

New Phytologist

Research 3

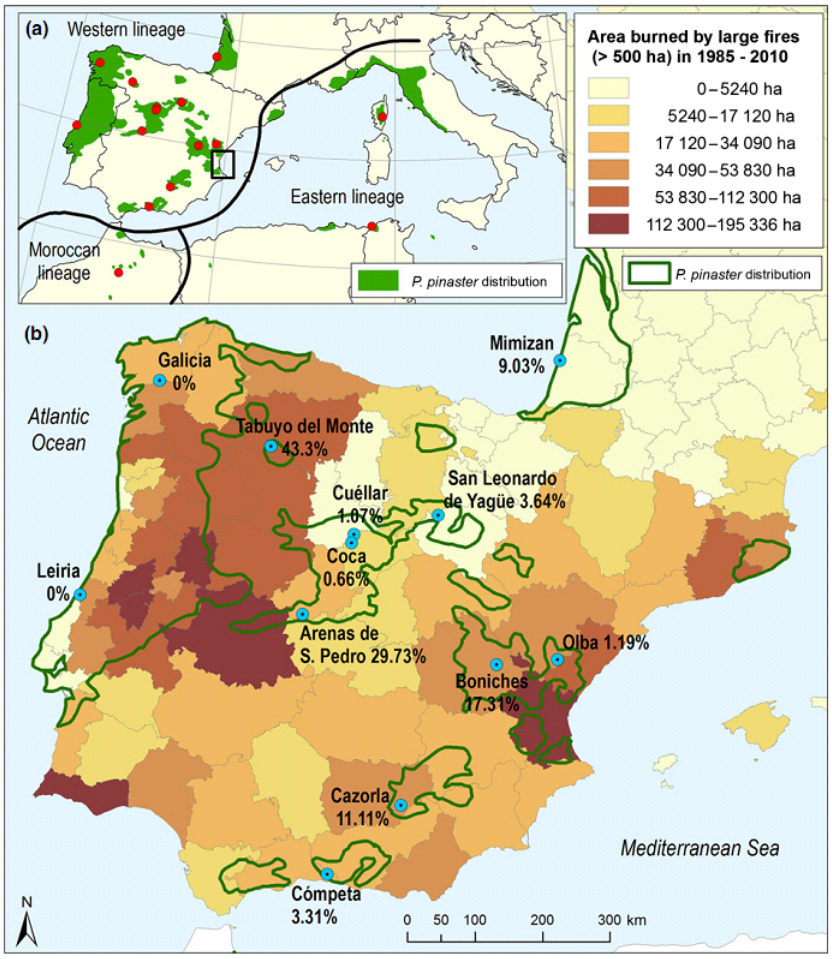

Fig. 1 (a) Focal region in eastern Spain (black box) and range-wide sampling, showing the distribution of the three maternal (mitochondrial DNA, mtDNA) lineages (black lines) known in *Pinus pinaster* (Burban & Petit, 2003). (b) Serotiny estimates for populations within the western maternal lineage (data retrieved from Tapias *et al.*, 2004) superimposed on provincial data of burned area by large fires (> 500 ha, from 1985 to 2010; European Commission, 2010). Maritime pine range is shown in green in the two maps.

Materials and Methods

Plant material

The sampling included a total of 509 maritime pine trees (Fig. 1): (1) 199 individuals collected from three stands in eastern Spain (coordinates: Calderona, 39.75, −0.50; Sinarcas, 39.79, −1.20; Eslida, 39.88, −0.30; each represented by 66–67 trees) which show high within-stand variability for fire phenotypes (0–100% serotinous cones per tree, average proportion of serotinous cones of 36.29% and standard deviation of 23.36%; Fig. S2); this dataset was used to identify loci potentially associated with serotiny and to construct SNP-based predictive models for fire phenotypes in this region; (2) 310 individuals from 15 range-wide populations (average of *c.* 21 trees per population) including the three distinct maternal (mtDNA) lineages known in maritime pine (see Table S1); this dataset was used to test for model prediction power outside the focal region. Needles were collected from the 509 (199 + 310) individuals and desiccated using silica gel. Genomic DNA was isolated using the Invisorb® DNA Plant HTS 96 Kit/C kit (Invitek GmbH, Berlin, Germany).

Nuclear microsatellites (simple sequence repeats, SSRs)

Eleven nuclear microsatellites (SSRs) were used to test for population genetic structure (which may bias genetic association approaches) within the sampling region in eastern Spain ($n = 199$; see also Bucci *et al.*, 2007; Santos-del-Blanco *et al.*, 2012; microsatellite data were deposited in the Dryad repository with doi: 10.5061/dryad.1p2s5/3). Primer sequences were obtained from Mariette *et al.* (2001) (*ITPH4516* and *FRPP94*), Chagné *et al.* (2004) (*rpTest11*, *Ctg4363*, *NZPR1078* and *NZPR413*), Guevara *et al.* (2005) (*A6F03*), Steinitz *et al.* (2011) (*pEST2669*) and F. Sebastiani and G. G. Vendramin (pers. comm., June 2011) (*epi3*, *epi5* and *gPp14*). Forward primers were 5′ end-labeled with fluorochromes (HEX, FAM, VIC or PET) and amplified using the Qiagen Multiplex PCR Kit (Qiagen, Venlo, the Netherlands) following the manufacturer's instructions. Amplified allele fragments were separated using an ABI 3730 genetic analyzer (Applied Biosystems, Carlsbad, CA, USA) and their sizes were determined with reference to the GeneScan −500 LIZ® Size Standard (Applied Biosystems) using GeneMapper software version 4.0 (Applied Biosystems).

384-plex SNP assay and genotyping

Genotyping ($n = 509$) was performed with Illumina VeraCode® technology for a 384-plex SNP Oligo Pool Assay (OPA) (design file provided as Table S2; SNP data deposited in the Dryad repository with doi: 10.5061/dryad.1p2s5/2). This OPA is based on a subsample of the 1536-plex SNP assay developed for *P. pinaster* by Chancerel *et al.* (2011), including polymorphisms from drought stress response candidate genes associated with climate variables in Mediterranean (*P. pinaster* and *P. halepensis*) and American (*P. taeda*) pines, as reported in Grivet *et al.* (2011) and Eckert *et al.* (2010a,b), respectively. The OPA also included 50 and 68 expressional candidate genes for stress response in maritime pine (Perdiguero *et al.*, 2013) and loblolly pine (Lorenz *et al.*, 2011), as well as a wide representation of functional candidate genes for biotic and abiotic stress responses, physical and chemical wood properties, phenology and growth in maritime pine (e.g. Pot *et al.*, 2005; Eveno *et al.*, 2008; Grivet *et al.*, 2011; Lepoittevin *et al.*, 2012; J. P. Jaramillo-Correa, pers. comm.) and other conifers (e.g. González-Martínez *et al.*, 2007, 2008; Eckert *et al.*, 2010a,b; Mosca *et al.*, 2012a,b). For example, Lepoittevin *et al.* (2012) found that SNPs *hdz31-2268* and *m1027* in this OPA were strongly associated with variation in growth and wood cellulose content, respectively, and Grivet *et al.* (2011) and J. P. Jaramillo-Correa (pers. comm.) found the allele frequency of several SNPs in this OPA (including *m705*, *m1196* and *m1211*, see the Results and Discussion sections) to be significantly associated with climatic variables (particularly with temperature variables) at regional and range-wide spatial scales.

Phenotypic measurements

Serotiny was estimated for 199 individuals in three natural stands in eastern Spain (as described previously; phenotypic data deposited in the Dryad repository with doi: 10.5061/dryad.1p2s5/1). Dominated trees and trees with a diameter of < 10 cm were avoided. For each sampled tree, serotinous (closed) and non-serotinous (open or partially open) cones were counted using binoculars on two pairs of opposite branches belonging to the upper and second third of the canopy, respectively. Cones in the main trunk were also counted, as they are abundant in this species. The serotiny level for each tree was then estimated as the number of closed cones (those remaining closed after maturation) with respect to the total number of cones (open and closed). Because changes in humidity can close open cones, serotiny was assessed during hot spring and summer days (i.e. the dry season).

Range-wide serotiny data (population means for the 15 SNP-genotyped populations; as described previously) were retrieved from Tapias *et al.* (2004). The serotiny level was estimated here by counting all closed and open cones. Population means were based on 32 individuals per population (480 observations).

Population genetic structure and kinship

Population genetic structure within the sampling region in eastern Spain ($n = 199$) was assessed using 11 nuclear microsatellites and the Bayesian clustering method implemented in STRUCTURE 2.2 (Pritchard *et al.*, 2000). Ten runs were performed for each number of clusters, $K = 1$ to $K = 5$, with a burn-in length of 50 000 and a run length of 500 000 iterations, and using an admixture model with correlated allele frequencies.

Pairwise kinship was estimated using all available markers (i.e. 11 SSRs and 251 successfully genotyped and polymorphic SNPs) to approximate the covariance matrix among the individuals used in mixed-effects linear models (MLMs, to be described), following Yu *et al.* (2006). The kinship estimator of Loiselle *et al.* (1995), as implemented in SPAGeDi 1.3 (Hardy & Vekemans, 2002), and the skewness of the pairwise kinship distribution were computed. Deviation of this distribution from normal expectations centered on mean zero, as evaluated by D'Agostino's skewness test, implies significant family structure within the population.

Identification of marker loci potentially associated with fire phenotypes

Single-locus approach The identification of SNPs with significant single-locus allelic effects on fire phenotypes followed a two-step approach. First, a preliminary selection of SNPs was based on MLMs (see Yu *et al.*, 2006), fitted independently for each SNP marker, as implemented in Tassel 3.0 (Bradbury *et al.*, 2007). The covariance matrix among individuals for the MLMs was approximated using all available markers (SSRs and SNPs), as explained previously. Negative kinship values were set to zero following Yu *et al.* (2006). We considered three alternative genetic models accounting for additive allele effects, over-dominance and allele dominance. A false discovery rate (FDR) approach (Storey, 2002; Storey & Tibshirani, 2003) was used to estimate the proportion of true null hypotheses, π_0 (i.e. $1 - \pi_0$ indicates the expected proportion of significant associations), among all tests for each genetic model. Second, for marker loci potentially associated with fire phenotypes and minor allele frequency (MAF) > 0.10, a Bayesian mixed-effect association approach (Bayesian Association with Missing Data, BAMD; Gopal *et al.*, 2009; Quesada *et al.*, 2010; Li *et al.*, 2012) in R v.2.13.1 (R Development Core Team, 2008) was used to estimate single-locus allelic effects under the three genetic models. Mean allelic effects (γ) and 95% confidence intervals were obtained from the distribution of the last 20 000 iterations (50 000 in total). Only those SNPs with confidence intervals not overlapping zero were considered to be potentially associated with the trait.

Multi-locus approach A stepwise mixed model strategy was used to further identify SNPs potentially associated with fire phenotypes. This multi-locus approach combined a multiple regression selection strategy, together with the mixed model, following Segura *et al.* (2012). Briefly, the most significant SNPs, based on genetic variance estimates from a mixed model using restricted maximum likelihood (REML), were included, one by one, as cofactor in the mixed model at each step. Then, the Bayesian Information Criterion (BIC) was used to select the best model. Only common SNPs (i.e. with MAF > 0.10) were considered in these analyses.

New Phytologist

Research 5

SNP-based predictive model for serotiny at a local scale

Common (MAF > 0.10) SNPs potentially associated (either alone or in combination with other SNPs, see the Results section) with fire phenotypes in the eastern Iberian range of maritime pine (17 loci) were employed to construct predictive models for serotiny in this region using ridge regression in a mixed-effects modeling framework (RR-BLUP). SNP-based breeding values for serotiny (i.e. numerical predictions of the relative genetic merit of each tree based exclusively on its genotype) were obtained using the rrBLUP R-software package (Endelman, 2011; Endelman & Jannink, 2012). This 'penalized' regression technique is commonly used to circumvent the classical 'large *p*, small *n* problem' (Johnstone & Titterington, 2009), although it was originally designed to deal with predictors' collinearity (Hoerl & Kennard, 1970). Despite the typically high effective population size in conifers, RR-BLUP has been shown to produce models that have moderate to high accuracy (0.17–0.51, as estimated by the Pearson correlation between de-regressed breeding values from quantitative genetic analyses and SNP-based breeding values) for a wide range of phenotypic traits (h^2 of 0.07–0.45) in loblolly pine (Resende *et al.*, 2012). In our study case, the use of preselected SNPs potentially linked to the trait (or to causal SNPs, if they are not the causal SNPs themselves) is expected to improve the predictive value of the models (see Westbrook *et al.*, 2013). Six-fold cross-validation was used to evaluate model accuracy in the sampled region (eastern Spain). Model accuracy was estimated by the Pearson's correlation coefficient r between observed serotiny and SNP-based breeding values, and the corresponding adjusted R^2.

Model prediction power at wide geographical scales

The recent literature has shown that SNP loci can have either local or wide-range adaptive value (e.g. Hancock *et al.*, 2011; Prunier *et al.*, 2012). If the same polymorphisms underlie fire-related traits in different regions, predictive models for serotiny based on candidate genes could still be valid outside the focal region. Conversely, different genetic lineages may have undergone independent (i.e. lineage-specific) adaptive processes, restricting the predictive value of the model to the local scale. Moreover, differences in LD, for example those caused by contrasted demographic history (e.g. Heuertz *et al.*, 2006), or in the strength of selection, could also affect predictive model accuracy in distinct geographical regions. To test these alternative hypotheses, the RR-BLUP model was used to predict fire phenotypes for 310 individuals from 15 range-wide populations representing the three distinct maternal (mtDNA) lineages recognized in maritime pine (western, eastern and Moroccan lineages; Burban & Petit, 2003). Serotiny phenotypes (averages by population) were retrieved from Tapias *et al.* (2004) for the same populations. It should be noted that this is a fully independent dataset to that used for model construction. Significant correlations between population means of estimated breeding values based on SNPs and observed serotiny for distinct maritime pine gene pools were tested using Kendall's τ non-parametric rank correlation coefficient.

Results

The 384-plex SNP OPA conversion rate was relatively high (*c.* 66%; Gentrain, GC50 and GC10 scores of 0.822, 0.819 and 0.801, respectively), and we were able to obtain high-quality genotypes for 251 polymorphic SNP loci in all 509 trees. Successfully genotyped SNPs included numerous loci with suggested adaptive value in previous pine studies (see the Materials and Methods section).

Population genetic structure and kinship

No genetic structure was detected in any of the STRUCTURE runs (Fig. S1). Average pairwise kinship in the sample was close to zero (−2.858E-05), as expected. However, the pairwise kinship distribution was skewed towards positive kinship values (D'Agostino's skewness test: skew = 0.208, z = 7.777, P = 7.387E-15; Fig. S3), thus indicating certain pairs of related trees in the population, which supports the use of mixed-model approaches for genetic association.

Identification of marker loci potentially associated with fire phenotypes

Single-locus MLMs identified 26 SNPs under the additive and allele dominance genetic models that were potentially associated ($P < 0.05$) with serotiny (see Table S3; no significant SNPs were found for the over-dominance model), including one locus with $Q < 0.10$ (*m692*; see Fig. 2). The inflation factor λ was 1.11–1.29, depending on the genetic model, which indicates moderate inflation of P values. Nevertheless, the FDR approach resulted in $\pi_0 = 0.7568$ (additive model) and $\pi_0 = 0.7532$ (allele dominance model), which suggests that at least some associations are not false positives. Twelve SNPs with MAF > 0.10 had a significant (at $\alpha = 0.05$) allelic effect on serotiny phenotypes, as estimated by BAMD, under the different genetic models tested (two-tailed test; Table 1 and Fig. 2). Two of these loci, *m15* and *m816*, best fitted the over-dominance model, whereas the other ten (*m594*, *m692*, *m696*, *m698*, *m705*, *m912*, *m955*, *m974*, *m1194* and *m1196*) best fitted the additive or allele dominance models, in agreement with the MLM results. The best stepwise mixed model combined the effects of 11 SNPs (Fig. 3). This set included six SNPs (*m594*, *m692*, *m698*, *m974*, *m1194* and *m1196*) of the 12 with significant single-locus allelic effects reported above, and five additional SNPs (*m289*, *m817*, *m959*, *m1211* and *m1414*; see Table S4).

SNP-based predictive model for serotiny at a local scale

Seventeen common SNPs (MAF > 0.10) potentially associated with fire phenotypes in the eastern Iberian Peninsula (12 SNPs from single-locus MLM/BAMD analysis and five additional SNPs detected in the stepwise mixed model) were used to construct a phenotypic predictive model based on RR-BLUP. The model explained 29.15% (Pearson's product-moment correlation coefficient r of 0.556) of the phenotypic variation in serotiny in

New Phytologist

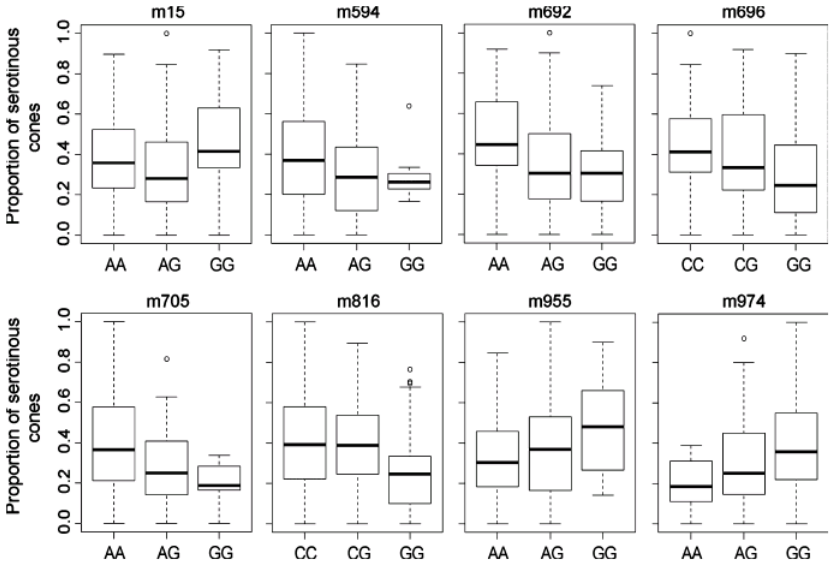

Fig. 2 Genotypic effects (box plots) of eight exemplary common single nucleotide polymorphisms (SNPs) (minor allele frequency (MAF) > 0.10) showing significant association with observed serotiny in *Pinus pinaster*, as identified by a two-step approach based on mixed-effects linear models (MLMs) and Bayesian association analysis (Bayesian Association with Missing Data, BAMD). Marker codes follow Chancerel *et al.* (2011). For the box plots, the box indicates the interquartile range and the band inside it the median; the whiskers extend to the furthest data point within 1.5 times the length of the box; outliers are depicted with circles.

Table 1 Significant marker effects of common single nucleotide polymorphisms (SNPs) (minor allele frequency (MAF) > 0.10) on serotiny in *Pinus pinaster*, as identified by a two-step approach based on mixed-effects linear models (MLMs) and Bayesian genetic association (Bayesian Association with Missing Data, BAMD)

								Marker effects			
								Mixed linear model			
SNP	Annotation	SNP motif	Site annotation[a]	LG	MAF	*N*	Genetic model	*F*	*P*	R^2	Bayesian model Mean γ (95% CIs)
m15	Defectively organized tributaries 2 (DOT2)	T/C	nc		0.3706	197	A	6.052	0.003	0.062	−0.2467 (−0.4653, −0.0282)
m594	Pyrophosphate-energized vacuolar membrane proton (AVP)	T/C	syn	8[b,c]	0.1878	197	D	6.915	0.009	0.035	−0.3678 (−0.6264, −0.1105)
m692	Unknown	A/G	unk	3[b]	0.4133	196	D	12.932	4.0E-04	0.067	−0.4696 (−0.7964, −0.1426)
m696	Arabinogalactan-like protein (AGP)	C/G	nc	3[b,c]	0.4031	196	D	5.722	0.018	0.029	−0.3206 (−0.5773, −0.0658)
m698	Nascent polypeptide-associated complex subunit alpha-like protein (NAC-alpha)	T/C	syn		0.2864	199	A	ns	ns	ns	0.2938 (0.1130, 0.4751)
m705	Carotenoid cleavage dioxygenase (CCD)	A/G	nc		0.1231	195	D	6.737	0.010	0.034	−0.3083 (−0.5993, −0.0197)
m816	Receptor protein kinase clavata1 (CLV1)	C/G	syn	1[b,c]	0.4924	197	O	ns	ns	ns	0.3391 (0.1080, 0.5706)
m912	Peroxidase 72-like (PER72)	A/T	non-syn		0.3795	195	D	5.791	0.017	0.031	0.2601 (0.0205, 0.4992)
m955	Unknown	A/G	unk	3[b,c]	0.2475	198	A	3.049	0.050	0.031	0.2294 (0.0469, 0.4160)
m974	1-Aminocyclopropane-1-carboxylate synthase (ACC)	A/G	syn	11[c]	0.1231	199	D	5.387	0.021	0.027	0.2946 (0.0137, 0.5731)
m1194	Cell division-related protein	C/G	syn		0.1439	198	A	3.714	0.026	0.038	0.2501 (0.0275, 0.4744)
m1196	Peptidyl-prolyl cis–trans isomerase (PPI)	A/C	syn		0.3266	199	A	ns	ns	ns	−0.1994 (−0.3871, −0.0086)

Bayesian mean allelic effects (γ) and 95% confidence intervals (CIs) were obtained from the distribution of the last 20 000 iterations in BAMD (for details, see Li *et al.*, 2012). Allelic effects are provided for the genetic model (A, additive; O, over-dominance; D, allele dominance) with higher effect on fire phenotype. Marker names and linkage groups (LG) as reported in Chancerel *et al.* (2011) and De Miguel *et al.* (2012); ns, not significant for that particular genetic model.

[a]Site annotation: nc, non-coding (untranslated regions or introns); non-syn, non-synonymous; syn, synonymous; unk, unknown.

[b]LG from Chancerel *et al.* (2011).

[c]LG from De Miguel *et al.* (2012).

New Phytologist

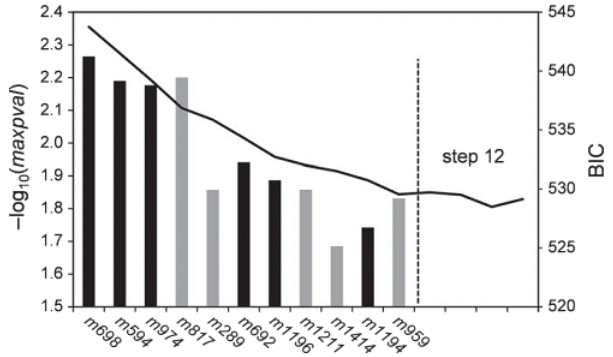

Fig. 3 Single nucleotide polymorphisms (SNPs) selected by the best stepwise mixed model in *Pinus pinaster* (step 12), as evaluated by the Bayesian Information Criterion (BIC); *maxpval*, *P* values for the best SNP introduced in each step. SNPs with black bars were also identified by the single-locus approach.

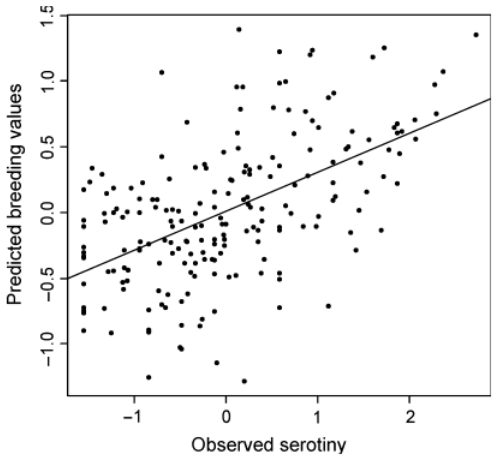

Fig. 4 Correlation of predicted breeding values based on ridge regression in a mixed-effects modeling framework (RR-BLUP) and observed levels of serotiny (standardized) in the eastern Iberian *Pinus pinaster* range. A linear trend is also shown.

this region (Fig. 4). Cross-validation showed a highly significant correlation between the SNP-based breeding values obtained from RR-BLUP and observed serotiny (Pearson's r of 0.429–0.632; P values of 0.00008–0.01270), with adjusted R^2 of 0.158–0.378. SNPs were not in LD, and even those obtained from the same gene (*m816* and *m817*) showed low allelic correlation values ($r^2 = 0.262$; see also LD plot in Fig. S4). Ten mapped SNPs belonged to five different linkage groups (four SNPs to LG1, three SNPs to LG3 and one SNP each to LGs 5, 8 and 11; Chancerel *et al.*, 2011; De Miguel *et al.*, 2012; see also Tables 1, S4). Thus, most significant associations represent (or are linked to) distinct causal SNPs.

Model prediction power at wide geographical scales

The predictive model for serotiny developed for the eastern Iberian Peninsula (as shown previously) had variable success outside the focal population depending on the geographical region. At range-wide scale, correlations between observed and predicted serotiny at the population level were still significant within the western maternal lineage of maritime pine (Kendall's $\tau = 0.44$, $z = 1.9933$, $P = 0.046$), but not when the eastern (Corsica and Tunisia) and Moroccan lineages were also considered (Kendall's $\tau = 0.13$, $z = 0.6937$, $P = 0.488$). This result is not surprising, considering that maternal lineages are completely isolated in maritime pine (Burban & Petit, 2003), which would facilitate lineage-specific adaptations, and that remarkable differences across geographical regions have been described for phenotypic and morphological traits in this species (Scott, 1962; Resch, 1974). Within the western maternal lineage, correlation between predicted and observed serotiny was higher in geographical regions closer to the focal population, from *c.* 30% in nearby populations from eastern and southern Spain to *c.* 20% in further away central Spain and Atlantic regions of maritime pine. In these models, low levels of serotiny were associated with negative breeding values, with the Atlantic and central Spain regions having more negative values (average of −0.228 and −0.036, respectively), and eastern and southern Spain (which are known for the higher incidence of forest fires; see Fig. 1) having more positive ones (average of 0.354 and 0.063, respectively).

Discussion

In this article, we have provided a case study for genetic association of an ecologically relevant trait (serotiny) evaluated *in situ* in natural forest tree populations. As demonstrated here, *in situ* genetic association can be achieved when the trait under study: (1) has large phenotypic variability within a region that lacks population genetic structure; (2) is under strong genetic control (i.e., heritability is high); and (3) can be accurately quantified in a large number of individuals. This approach is especially suitable for forest trees that generally form large random-mating unstructured natural populations with relatively high nucleotide diversity (Neale & Savolainen, 2004; Neale & Ingvarsson, 2008) and for which several heritable adaptive traits (e.g. female reproduction, Santos-del-Blanco, 2013; wood density, Cornelius, 1994; cold hardiness, Howe *et al.*, 2003 and references therein) can be readily evaluated in a large number of individuals.

Targeting traits that represent 'ecological syndrome' phenotypes (i.e. involving several correlated traits) that have evolved as a response to the same selective pressure (Reich *et al.*, 2003), such as serotiny for fire phenotypes, increases the chances of finding associated marker variation, even with relatively low genotyping effort. The interpretation of genetic associations for these complex phenotypes, however, can be obscured by genetic correlations among traits. In these cases, functional annotation of potentially associated loci can help to elucidate the specific traits involved and their underlying genetic architecture (to be described for serotiny). Genetic dissection of 'ecological syndrome' phenotypes, although challenging, is a promising field, as many of the most relevant ecological adaptations involve syndromes rather than single traits (e.g. pollination syndromes, Fenster *et al.*, 2004; plant defense

syndromes, Agrawal & Fishbein, 2006; Mediterranean plant syndromes, Verdú & Pausas, 2013).

In maritime pine, we identified 17 loci potentially associated with serotiny which together explained *c.* 29% of the phenotypic variation found in natural eastern Iberian populations of the species. Similar levels of explained variance have been reported in association studies (based on clonal banks or common gardens) in other conifers (e.g. *c.* 20% for wood properties in *Pinus taeda*, González-Martínez *et al.*, 2007; 17% for cold hardiness in *Pseudotsuga menziesii*, Eckert *et al.* 2009; *c.* 34% for bud set and *c.* 28% for cold hardiness in *Picea sitchensis*, Holliday *et al.*, 2010). Model accuracy within the target region (estimated using Pearson's correlation coefficient) was also similar to that found in a predictive model for oleoresin flow in loblolly pine (0.51–0.62 vs 0.43–0.63 in our study). In addition, Parchman *et al.* (2012) identified 11 SNPs that explained 50% of the variance for serotiny using a genotyping-by-sequencing (GBS) approach in *Pinus contorta*. One advantage of GBS approaches is the high marker density covering the whole genome, albeit site annotations and gene functions are normally unknown (in non-model species). Genome-wide approaches, such as GBS, are also able to reveal unexpected functional associations that would normally not have been considered in more targeted candidate gene studies. Nevertheless, Westbrook *et al.* (2013) found that 20–30 significantly associated SNPs had the same predictive value as the full dataset altogether (4854 SNPs) for oleoresin flow in loblolly pine, supporting the idea that a few, well-selected loci could have as much predictive power as genome-wide datasets.

The 'fire syndrome' arose in pines at the same time as the genus split into its two main lineages (i.e. subgenus *Pinus* and subgenus *Strobus*; He *et al.*, 2012). Species of the subgenus *Pinus*, such as *P. pinaster*, are typical of fire-prone ecosystems, whereas those in subgenus *Strobus* are mainly adapted to low productivity sites with either low soil nutrient conditions or hot and cold climatic extremes (Millar, 1998; Keeley, 2012). Interestingly, six of the loci potentially associated with serotiny in *P. pinaster* (see annotation in Table S4) are found in genes involved in the water stress response (*m289*, *m696*, *m698*, *m705*, *m912* and *m974*; Schwanz *et al.*, 1996; Zhu, 2002; Jenks & Wood, 2009), including one non-synonymous mutation in a gene coding for a peroxidase (*m912*). Another locus (*m1211*) is associated with winter temperatures at range-wide scales (J. P. Jaramillo-Correa, pers. comm.). Together, they may reflect a correlation of drought with fire phenotypes (as gauged by serotiny), a finding that has also been reported in the fire ecology literature (Pausas & Fernández-Muñoz, 2012). Three other loci (*m15*, *m816* and *m817*) are found in genes involved in cell differentiation and root, shoot and flower development (Ávila *et al.*, 2006; Casson *et al.*, 2009), and could be related to distinct growth habits and maturity age in highly serotinous trees adapted to crown fires (e.g. short trees without self-pruning and with early maturity age). Further loci are found in genes involved in more general functions (e.g. cell division, membrane transport or protein folding) or for which the function is still unknown.

The predictive value for serotiny of SNP-based models constructed in the eastern Iberian Peninsula (i.e. the linkage of the marker loci with the trait) was higher, as expected, in regions geographically closer to the focal population, and lost any predictive power outside the maternal genetic lineage within which the models were constructed. This is in line with recent findings in natural populations of *Arabidopsis thaliana*, the selfing model plant species. For example, Fournier-Level *et al.* (2011) found that alleles associated with higher survival in different sites were locally more abundant than genomic controls, and that different loci underlie the same trait in sites with contrasting environments. Our results are also in agreement with previous research in conifers, which reported lineage-specific adaptations at the molecular level (e.g. Prunier *et al.*, 2012 for climate adaptation in *Picea mariana*). Alternatively, reduction in model prediction power could reflect different levels of LD across regions (for instance, as a result of regional demographic events such as bottlenecks) or variable strength of selection. However, we did not find any evidence of lower LD in populations from the eastern (average r^2 of 0.034 vs. 0.018 in eastern Spain) or Moroccan (average r^2 of 0.052) maternal lineages of maritime pine (see also Fig. S5). Moreover, fire regime (assumed to be related to the strength of selection for serotiny; Gauthier *et al.*, 1996; Keeley & Zedler, 1998; Tapias *et al.*, 2004), although variable, is similar in regions with high and low model prediction power (see Tapias *et al.*, 2004 and Fig. 1). Finally, it is noteworthy to point out that the predictive value of the serotiny model, although limited, still extended to populations located hundreds of kilometers away from the focal population for which it was constructed. Westbrook *et al.* (2013) showed that significantly associated SNPs can be used to construct predictive models that are robust to environmental variation. Thus, a few well-constructed models covering the main gene pools of the species may be sufficient for accurate phenotypic prediction of serotiny in maritime pine.

Comparative studies of closely related species can shed light on adaptive evolutionary processes at higher phylogenetic scales than can within-species population genetics studies. Within the genus *Pinus*, serotiny evolved several times independently (Grotkopp *et al.*, 2004; He *et al.*, 2012), which provides a rich source of comparative data. Although evidence across species is still scarce, the only two available genetic association studies for serotiny in pines were able to explain substantial amounts of the phenotypic variance for this trait, 50% in lodgepole pine, a North American species, and *c.* 29% in the Mediterranean maritime pine (Parchman *et al.*, 2012 and this study, respectively), which makes comparative approaches promising. Furthermore, these two studies provide strong support to reject the previously proposed simple genetic control (one locus with two alleles) for serotiny in pines (Teich, 1970; Perry & Lotan, 1979), as distinct unlinked SNPs were potentially associated with serotiny in both species and, in the case of maritime pine (no mapping data are available for lodgepole pine), they mapped to five different linkage groups (see linkage maps in Chancerel *et al.*, 2011; De Miguel *et al.*, 2012).

Newly available genomic tools and analytical methods, such as association genetics, provide opportunities for a better understanding of the molecular basis of ecological adaptations in non-model species (Stinchcombe & Hoekstra, 2008; Stapley *et al.*, 2010), particularly with regard to climate change (Feder &

Mitchell-Olds, 2003; González-Martínez *et al.*, 2006; Hoffmann & Sgrò, 2011). Fire and climate are closely linked, and adaptive responses to forest fires will acquire increasing importance as climate changes (Pausas, 2004). Thus, current predictive models of range shifts under climate change would benefit from genetic knowledge, such as the spatial distribution of genetic variation for fire-related traits, including serotiny. In addition, predictive models for serotiny, such as that developed here, can help to identify populations and individuals with an expected good response to increased fire recurrence and intensity. Ideally, genetic effects should be studied in the natural environment where they confer adaptive value. So far, genetic association studies have been mostly focused on model organisms and species with economic value, such as major crops and some forest trees, and under controlled environments (e.g. González-Martínez *et al.*, 2007; Holliday *et al.*, 2010). However, it is important to extent this field of research to natural environments and to other ecological keystone species with distinct life-history traits and evolutionary history (Feder & Mitchell-Olds, 2003; González-Martínez *et al.*, 2006; Stinchcombe & Hoekstra, 2008; Stapley *et al.*, 2010). The probability of the detection of ecologically relevant functional markers increases with the strength of the selection drivers. Therefore, populations adapted to extreme environments (Feder & Mitchell-Olds, 2003) or that have undergone rapid environmental change (e.g. during invasion of new areas; Hoffmann & Sgrò, 2011) are ideal for the *in situ* identification of ecologically relevant genetic variation.

Acknowledgements

We thank M. Zabal-Aguirre, A. I. de-Lucas, C. Odoñez, Y. Kurt and J. Chofre for field assistance. Thanks are extended to C. Lepoittevin and P. Garnier-Géré for help during SNP array design, and to J. Fernández for preparing Fig. 1. Data for Fig. 1 were provided by the European Forest Fire Information System – EFFIS (http://effis.jrc.ec.europa.eu) of the European Commission Joint Research Centre. SNP genotyping services were provided by the Spanish 'Centro Nacional de Genotipado' (CEGEN-ISCIII, http://www.cegen.org). This study was financed by the following projects from the Spanish government: VAMPIRO (CGL2008-05289-C02-01/02) and TREVOL (CGL2012-39938-C02-00/01), as well as by the EU FP7 project 'Novel Tree Breeding Strategies' (NovelTree) and the Italian MIUR project 'Biodiversitaly' (RBAP10A2T4). This research was also funded by the ERA-Net BiodivERsA (LinkTree project, EUI2008-03713), with the Spanish Ministry of Economy and Competitiveness as national funder, part of the 2008 BiodivERsA call for research proposals.

References

Agrawal AA, Fishbein M. 2006. Plant defense syndromes. *Ecology* **87**: 132–149.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**: 95–111.

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT *et al.* 2010. Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627–631.

Ávila C, Pérez-Rodríguez J, Cánovas FM. 2006. Molecular characterization of a receptor-like protein kinase gene from pine (*Pinus sylvestris* L.). *Planta* **224**: 12–19.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Bucci G, González-Martínez SC, Le Provost G, Plomion C, Ribeiro MM, Sebstiani F, Alia R, Vendramin GG. 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology* **16**: 2137–2153.

Burban C, Petit RJ. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* **12**: 1487–1495.

Casson SA, Topping JF, Lindsey K. 2009. MERISTEM-DEFECTIVE, an RS domain protein, is required for the correct meristem patterning and function in *Arabidopsis*. *Plant Journal* **57**: 857–869.

Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JM, Echt C *et al.* 2004. Cross species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics* **109**: 1204–1214.

Chancerel E, Lepoittevin C, Le Provost G, Lin YC, Jaramillo-Correa JP, Eckert AJ, Wegrzyn JL, Zelenika D, Boland A, Frigerio JM *et al.* 2011. Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine. *BMC Genomics* **12**: 368.

Cornelius J. 1994. Heritabilities and additive genetic coefficients of variation in forest trees. *Canadian Journal of Forest Research* **24**: 372–379.

Cumbie WP, Eckert AJ, Wegrzyn J, Whetten R, Neale DB, Goldfarb B. 2011. Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L. *Heredity* **107**: 105–114.

De Miguel M, de María N, Guevara MA, Díaz L, Sáez-Laguna E, Sáchnez-Gómez D, Chancerel E, Aranda I, Collada C, Plomion C *et al.* 2012. Annotated genetic linkage maps of *Pinus pinaster* Ait. from a Central Spain population using microsatellite and gene based markers. *BMC Genomics* **13**: 527.

Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB. 2010a. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology* **19**: 3789–3805.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutowsky KV, St. Clair JB,Neale DB. 2009. Association genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* **182**: 1289–1302.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra R, González-Martínez SC, Neale DB. 2010b. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**: 969–982.

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* **4**: 250–255.

Endelman JB, Jannink JL. 2012. Shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genetics* **2**: 1405–1413.

European Commission. 2010 *Forest Fires in Europe 2009*. Luxembourg, Germany: Office for Official Publications of the European Communities.

Eveno E, Collada C, Guevara MA, Léger V, Soto A, Diaz L, Léger P, González-Martínez SC, Cervera MT, Plomion C *et al.* 2008. Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* **25**: 417–437.

Feder ME, Mitchell-Olds T. 2003. Evolutionary and ecological functional genomics. *Nature Reviews Genetics* **4**: 649–655.

Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD. 2004. Pollination syndromes and floral specialization. *Annual Review of Ecology, Evolution and Systematics* **35**: 375–403.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**: 86–89.

Gauthier S, Bergeron Y, Simon JP. 1996. Effects of fire regime on the serotiny level of jack pine. *Journal of Ecology* **84**: 539–548.

González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB. 2008. Association genetics in *Pinus taeda* L. II. Water use efficiency. *Heredity* **101**: 19–26.

González-Martínez SC, Krutovsky KV, Neale DB. 2006. Forest-tree population genomics and adaptive evolution. *New Phytologist* 170: 227–238.
González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB. 2007. Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175: 399–409.
Gopal V, Li Z, Casella G. 2009. *BAMD: Bayesian association model for genomic data with missing covariates.* R package version 3.1.16 [WWW document] URL http://CRAN.R-project.org/package=BAMD [accessed 18 January 2013].
Grattapaglia D, Resende MD. 2011. Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7: 241–255.
Grivet D, Sebastiani F, Alía R, Bataillon T, Torre S, Zabal-Aguirre M, Vendramin GG, González-Martínez SC. 2011. Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology & Evolution* 28: 101–116.
Grotkopp E, Rejmánek M, Sanderson MJ, Rost TL. 2004. Evolution of genome size in pines (Pinus) and its life-history correlates: supertree analyses. *Evolution* 58: 1705–1729.
Guevara MA, Chagné D, Almeida MH, Byrne M, Collada C, Favre JM, Harvengt L, Jeandroz S, Orazio C, Plomion C *et al.* 2005. Isolation and characterization of nuclear microsatellite loci in *Pinus pinaster* Ait. *Molecular Ecology Notes* 5: 57–59.
Hancock AM, Brachi B, Faure N, Horton MW, Sperone FG, Jarymowycz LB, Toomajian C, Roux F, Bergelson JB. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334: 83–86.
Hardy OJ, Vekemans X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.
He T, Pausas JG, Belcher CM, Schwilk DW, Lamont BB. 2012. Fire-adapted traits of Pinus arose in the fiery Cretaceous. *New Phytologist* 194: 751–759.
Heuertz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M, Lascoux M *et al.* 2006. Multi-locus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174: 2095–2105.
Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation and applications for non-orthogonal problems. *Technometrics* 12: 55–82.
Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. *Nature* 470: 479–485.
Holliday JA, Ritland K, Aitken SN. 2010. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* 188: 501–514.
Holliday JA, Wang T, Aitken SN. 2012. Predicting adaptive phenotypes from multi-locus genotypes in Sitka Spruce (*Picea sitchensis*) using Random Forest. *G3: Genes, Genomes, Genetics* 2: 1085–1093.
Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany* 81: 1247–1266.
Jenks MA, Wood AJ. 2009. *Genes for plant abiotic stress.* Ames, IA, USA: Wiley-Blackwell.
Johnstone IM, Titterington DM. 2009. Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society A* 367: 4237–4253.
Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7: 1225–1241.
Keeley JE. 2012. Ecology and evolution of pine life histories. *Annals of Forest Science* 69: 445–453.
Keeley JE, Pausas JG, Rundel PW, Bond WJ, Bradstock RA. 2011. Fire as an evolutionary pressure shaping plant traits. *Trends in Plant Science* 16: 406–411.
Keeley JE, Zedler PH. 1998. Evolution of life histories in Pinus. In: Richardson DM, ed. *Ecology and biogeography of Pinus.* Cambridge, UK: Cambridge University Press, 219–250.
Krawchuk MA, Moritz MA, Parisien MA, Van Dorn J, Hayhoe K. 2009. Global pyrogeography: the current and future distribution of wildfire. *PLoS ONE* 4: e5102.
Kremer A, Le Corre V. 2012. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* 108: 375–385.
Lamont BB, Le Maitre DC, Cowling RM, Enright NJ. 1991. Canopy seed storage in woody plants. *Botanical Review* 57: 277–317.
Le Corre V, Kremer A. 2003. Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* 164: 1205–1219.
Lepoittevin C, Harvengt L, Plomion C, Garnier-Géré P. 2012. Association mapping for growth, straightness and wood chemistry traits in the *Pinus pinaster* Aquitaine breeding population. *Tree Genetics & Genomes* 8: 113–126.
Li Z, Vikneswaran G, Li X, Davis JM, Casella G. 2012. Simultaneous SNP identification in association studies with missing data. *Annals of Applied Statistics* 6: 432–456.
Loiselle BA, Sork VL, Nason J, Graham C. 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* 82: 1420–1425.
Lorenz WW, Alba R, Yu YS, Bordeaux JM, Simões M, Dean JFD. 2011. Microarray analysis and scale-free gene networks identify candidate regulators in drought-stressed roots of loblolly pine (*P. taeda* L.). *BMC Genomics* 12: 264.
Malcolm JR, Markham A, Neilson RP, Garcia M. 2002. Estimated migration rates under scenarios of global climate change. *Journal of Biogeography* 29: 835–849.
Mariette S, Chagné D, Decroocq S, Vendramin GG, Lalanne C, Madur D, Plomion C. 2001. Microsatellite markers for *Pinus pinaster* Ait. *Annals of Forest Science* 58: 201–206.
McLachlan JS, Clark JS, Manos PS. 2005. Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86: 2088–2098.
Millar CI. 1998. Early evolution of pines. In: Richardson DM, ed. *Ecology and biogeography of Pinus.* Cambridge, UK: Cambridge University Press, 69–91.
Moreira B, Tavsanoglu Ç, Pausas JG. 2012. Local versus regional intraspecific variability in regeneration traits. *Oecologia* 168: 671–677.
Moritsuka E, Hisataka Y, Tamura M, Uchiyama K, Watanabe A, Tsumura Y, Tachida H. 2012. Extended linkage disequilibrium in non-coding regions in a conifer, *Cryptomeria japonica*. *Genetics* 190: 1145–1148.
Mosca E, Eckert AJ, Di Pierro EA, Rocchini D, La Porta N, Belletti P, Neale DB. 2012a. The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular Ecology* 21: 5530–5545.
Mosca E, Eckert AJ, Liechty JD, Wegrzyn JL, La Porta N, Vendramin GG, Neale DB. 2012b. Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications* 5: 762–775.
Mouillot F, Field CB. 2005. Fire history and the global carbon budget: a 1 degrees × 1 degrees fire history reconstruction for the 20th century. *Global Change Biology* 11: 398–420.
Mouillot F, Rambal S, Joffre R. 2002. Simulating climate change impacts on fire frequency and vegetation dynamics in a Mediterranean-type ecosystem. *Global Change Biology* 8: 423–437.
Neale DB, Ingvarsson PK. 2008. Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology* 11: 149–155.
Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. *Trends in Plant Science* 9: 325–330.
Parchman TL, Gompertz Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* 21: 2991–3005.
Pausas JG. 2004. Changes in fire and climate in the Eastern Iberian Peninsula (Mediterranean basin). *Climatic Change* 63: 337–350.
Pausas JG, Alessio G, Moreira B, Corcobado G. 2012. Fires enhance flammability in *Ulex parviflorus*. *New Phytologist* 193: 18–23.
Pausas JG, Fernández-Muñoz S. 2012. Fire regime changes in the Western Mediterranean Basin: from fuel-limited to drought-driven fire regime. *Climatic Change* 110: 215–226.
Pausas JG, Keeley JE. 2009. A burning story: the role of fire in the history of life. *BioScience* 59: 593–601.
Pausas JG, Llovet J, Rodrigo A, Vallejo R. 2008. Are wildfires a disaster in the Mediterranean basin? – A review. *International Journal of Wildland Fire* 17: 713–723.
Pausas JG, Schwilk DW. 2012. Fire and plant evolution. *New Phytologist* 193: 301–303.

Perdiguero P, Barbero MC, Cervera MT, Collada C, Soto A. 2013. Molecular responses to water stress in two contrasting Mediterranean pines (*Pinus pinaster* and *Pinus pinea*). *Plant Physiology and Biochemistry* **67**: 199–208.

Perry DA, Lotan JE. 1979. A model of fire selection for serotiny in lodgepole pine. *Evolution* **33**: 958–968.

Petit JR, Hampe A. 2006. The evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics* **37**: 187–214.

Pike C, Warren J, David A. 2010. *Results update: 2nd generation jack pine population. Annual report Minnesota tree improvement cooperative.* Minneapolis, MN, USA: University of Minnesota.

Piñol J, Terradas J, Lloret F. 1998. Climate warming, wildfire hazard, and wildfire occurrence in coastal Eastern Spain. *Climatic Change* **38**: 345–357.

Pot D, Mcmillan L, Echt C, Le Provost G, Garnier-Géré P, Cato S, Plomion C. 2005. Nucleotide variation in genes involved in wood formation in two pine species. *New Phytologist* **167**: 101–112.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multi-locus genotype data. *Genetics* **155**: 945–959.

Prunier J, Gérardi S, Laroche J, Beaulieu J, Bousquet J. 2012. Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Molecular Ecology* **21**: 4270–4286.

Quesada T, Gopal V, Cumbie WP, Eckert AJ, Wegrzyn JL, Neale DB, Goldfarb B, Huber DA, Casella G, Davis JM. 2010. Association mapping of quantitative disease resistance in a natural population of loblolly pine (*Pinus taeda* L.). *Genetics* **186**: 677–686.

R Development Core Team. 2008. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reich PB, Wright IJ, Caveder-Bares J, Craine JM, Oleksyn J, Westoby M, Walters MB. 2003. The evolution of plant functional variation: traits, spectra, and strategies. *International Journal of Plant Sciences* **164**: 143–164.

Resch T. 1974. Essai de distinction des races majeurs de *Pinus pinaster*. *Annales de la Recherche Forestière au Maroc* **14**: 91–102.

Resende MFR Jr, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M. 2012. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* **190**: 1503–1510.

Salvador L, Alía R, Agúndez D, Gil L. 2000. Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait.) in the Iberian Peninsula. *Theoretical and Applied Genetics* **100**: 89–95.

Santos-del-Blanco L. 2013. *Evolutionary ecology of reproduction in two Mediterranean pine species (*Pinus pinaster *Ait. and* P. halepensis *Mil.).* PhD thesis, University of Valladolid, Spain.

Santos-del-Blanco L, Climent J, González-Martínez SC, Pannell JR. 2012. Genetic differentiation for size at first reproduction through male versus female functions in the widespread Mediterranean tree *Pinus pinaster*. *Annals of Botany* **110**: 1449–1460.

Schwanz P, Picon C, Vivin P, Dreyer E, Guehl JM, Polle A. 1996. Responses of antioxidative systems to drought stress in pedunculate oak and maritime pine as modulated by elevated CO_2. *Plant Physiology* **110**: 393–402.

Schwilk DW, Ackerly DD. 2001. Flammability and serotiny as strategies: correlated evolution in pines. *Oikos* **94**: 326–336.

Scott CW. 1962. A summary of information on *Pinus pinaster*. *Forestry Abstracts* **23**: i–xviii.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M. 2012. An efficient multi-locus mixed model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**: 825–832.

Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. 2010. Adaptation genomics: the next generation. *Trends in Ecology and Evolution* **25**: 705–712.

Steinitz O, Troupin D, Vendramin GG, Nathan R. 2011. Genetic evidence for a Janzen–Connell recruitment pattern in reproductive offspring of *Pinus halepensis* trees. *Molecular Ecology* **20**: 4152–4164.

Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158–170.

Storey JD. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B* **64**: 479–498.

Storey JD, Tibshirani R. 2003. Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences, USA* **100**: 9440–9445.

Strange A, Li P, Lister C, Anderson J, Warthmann N, Shindo C, Irwin J, Nordborg M, Dean C. 2011. Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis thaliana* accessions. *PLoS ONE* **6**: e19949.

Tapias R, Climent J, Pardos JA, Gil L. 2004. Life histories of Mediterranean pines. *Plant Ecology* **171**: 53–67.

Tapias R, Gil L, Fuentes-Utrilla P, Pardos JA. 2001. Canopy seed banks in Mediterranean pines of south eastern Spain: a comparison between *Pinus halepensis* Mill., *P. pinaster* Ait., *P. nigra* Arn. and *P. pinea* L. *Journal of Ecology* **89**: 629–638.

Teich AH. 1970. Cone serotiny and inbreeding in natural population of *Pinus banksiana* and *Pinus contorta*. *Canadian Journal of Botany* **48**: 1805–1809.

Verdú M, Pausas JG. 2013. Syndrome-driven diversification in a Mediterranean ecosystem. *Evolution* **67**: 1756–1766.

Westbrook JW, Resende MFR, Munoz P, Walker AR, Wegrzyn JL, Nelson CD, Neale DB *et al.* 2013. Association genetics of oleoresin flow in loblolly pine: discovering genes and predicting phenotype for improved resistance to bark beetles and bioenergy potential. *New Phytologist* **199**: 89–100.

Wymore AS, Keeley ATH, Yturralde KM, Schroer ML, Propper CR, Whitham TG. 2011. Genes to ecosystems: exploring the frontier of ecology with one of the smallest biological units. *New Phytologist* **191**: 19–36.

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al.* 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203–208.

Zhu JK. 2002. Salt and drought stress signal transduction in plants. *Annual Review of Plant Biology* **53**: 247–273.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 STRUCTURE software bar plots showing ancestry proportions for $K = 2$–5.

Fig. S2 Phenotypic variability (box plot) for serotiny in the eastern Iberian Peninsula.

Fig. S3 Density plot for pairwise kinship, as estimated by SPAGeDi 1.3.

Fig. S4 Linkage disequilibrium (LD) heatmap for single nucleotide polymorphisms (SNPs) associated with serotiny in focal study region.

Fig. S5 Linkage disequilibrium (LD) heatmaps for all polymorphic single nucleotide polymorphisms (SNPs) in four range-wide populations.

Table S1 Details on range-wide populations included in the single nucleotide polymorphism (SNP) genotyping

Table S2 Illumina Oligo Pool Assay (OPA) design file and designability scores provided as Excel file

New Phytologist

Table S3 Marker effects of 26 single nucleotide polymorphisms (SNPs) with $P < 0.05$ in the mixed-effects linear models

Table S4 Annotation and mapping information for single nucleotide polymorphisms (SNPs) associated with serotiny

Annex

Supplementary information Chapter 1

The ancient tropical rainforest tree *Symphonia globulifera* L.f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa

Supplementary Information S1. Nuclear microsatellites in *Symphonia globulifera*: PCR conditions and population genetics parameters.

Table S1.1. PCR conditions of the nSSR loci analysed in *Symphonia globulifera*. IRDye, emission wavelength (nm) of the fluorescent label attached to the forward primer of each locus for analysis on a Li-Cor 4300 DNA Analyzer.

Locus	IRDye	$MgCl_2$ [mM]	Annealing Temperature [°C]	Number of cycles
SG03	800	1.5	55	38
SGC4	700	1	63	25
SG10	700	0.5	58	31
SG18	700	1	55	33
SG19	800	0.5	61	31

Table S1.2. Characterization of nSSR loci for all *Symphonia globulifera* samples. H_T, gene diversity; F_{IS}, fixation index; F_{ST} and R_{ST}, differentiation indices based on unordered and ordered alleles, respectively. Populations were defined as 26 sampling locations as in Table 1 of the manuscript. Null allele frequencies were calculated using the Expectation Maximization (EM) algorithm (Dempster *et al.*, 1977) included in the software FreeNA (Chapuis and Estoup, 2007). Significance levels denote deviation from zero for F_{ST} and evidence of phylogeographic structure for R_{ST} (permutation test with alternative hypothesis: R_{ST} > permuted R_{ST}): *, P <0.05; **, P <0.001; ***, P<0.0001.

Locus	No. of missing genotypes	No. of alleles	Allele size range [bp]	H_T	F_{ST}	R_{ST}	F_{IS}	Null allele frequencies
SG03	21 (9.3%)	18	312-368	0.807	0.241***	0.577**	0.408***	0.234
SGC4	18 (8.0%)	21	152-228	0.896	0.092***	0.175*	0.169***	0.108
SG10	8 (3.6%)	43	134-256	0.952	0.077***	0.112*	0.096**	0.058
SG18	31 (13.8)	9	294-324	0.736	0.186***	0.296	0.397***	0.196
SG19	32 (14.2%)	20	158-270	0.908	0.081***	0.114	0.249***	0.143
All		111			0.125***	0.188**		

Table S1.3. Estimates of population genetics parameters from SSR data in *Symphonia globulifera*. Population numbers are the same as in Table 1 and Figure 1. A_E, effective allele number; R_S, allelic richness and its standard error (SE) using rarefaction to a sample of 7 individuals per population; H_E, gene diversity; F_{IS}, fixation index calculated in FSTAT or INEst (Chybicki and Burczyk, 2009); *s*, selfing rate estimated with RMES (David *et al.*, 2007). Significance levels after Bonferroni correction: *, P <0.05; **, P <0.001; ***, P <0.0001.

Population	*n*	A_E	R_S(SE)	H_E	F_{IS} (FSTAT)	F_{IS} (SE) (INest)	*s* (RMES)
Benin (1)	19	4.730	5.96 (0.68)	0.773	0.157**	0.009 (0.040)	0.000
Cameroon Korup (2)	39	5.655	6.04 (0.81)	0.807	0.348***	0.056 (0.069)	0.057 n.s.
Cameroon SW (8)	22	4.736	5.68 (1.41)	0.621	0.134*	0.029 (0.064)	0.000
Gabon coast (15)	14	5.789	6.73 (1.21)	0.802	0.264***	0.000 (0.000)	0.081 n.s.
Gabon CW (16)	11	4.931	5.63 (1.51)	0.697	0.170*	0.001 (0.008)	0.000
São Tomé (26)	39	5.138	5.94 (0.82)	0.774	0.239***	0.008 (0.033)	0.000

Supplementary information S2. Exploration of demographic history in geographic regions of *Symphonia globulifera* using the Bottleneck program.

To assess the demographic history of each geographic region, the 'bottleneck' statistic T_2 was computed using Bottleneck 1.2.02 (Cornuet and Luikart, 1996). T_2 represents the average over loci of the standardized deviation of the gene diversity, H_E, from the gene diversity expected for the number of alleles in the population, H_A. T_2 is expected to be zero under mutation-drift equilibrium in a constant-size population. Positive T_2 (H_E excess) indicates a recent bottleneck, whereas negative T_2 (H_E deficiency) is consistent with recent population expansion. T_2 was computed under the Infinite Allele Model (IAM) and under the Two Phase Model (TPM, with 70% of stepwise mutations) which is likely to be appropriate for the system, given that R_{ST} was larger than R_{ST}[permuted] for some pairs of gene pools. The magnitude of H_E excess or deficiency across loci was tested against equilibrium expectations using Wilcoxon signed rank tests (Piry *et al.*, 1999).

Table S2.1. Nuclear genetic diversity, allelic richness and demographic signatures in four geographic regions of *Symphonia globulifera*. R_S, allelic richness and its standard error (SE) for a sample of 16 individuals per region; H_E (SE), gene diversity; T_2, bottleneck statistic under the infinite alleles (*IAM*) and two-phase (*TPM*) models. Significant values are underlined.

			IAM			*TPM*		
	R_S	H_E	T_2	P (H_E excess)	P (H_E deficiency)	T_2	P (H_E excess)	P (H_E deficiency)
Benin	7.86 (0.91)	0.770 (0.061)	2.238	<u>0.016</u>	1.000	1.617	<u>0.016</u>	1.000
West Cameroon	8.33 (1.47)	0.803 (0.041)	2.095	<u>0.016</u>	1.000	1.241	<u>0.016</u>	1.000
South Cameroon, Gabon	11.36 (2.40)	0.769 (0.086)	0.697	0.313	0.891	-0.891	0.594	0.500
São Tomé	8.16 (1.27)	0.772 (0.057)	1.287	0.078	0.953	0.325	0.500	0.594

Supplementary information S3. Population size change scenarios compared with Approximate Bayesian Computation.

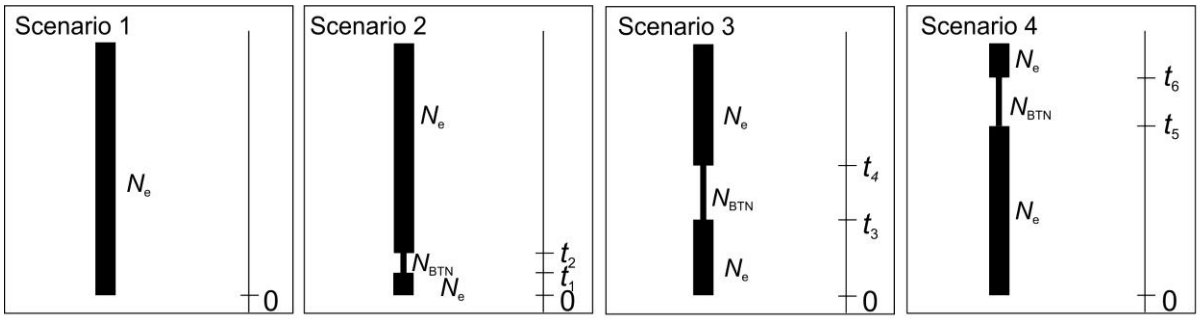

Figure S3.1. Graphical representation of four population size change scenarios simulated in DIYABC. Scenario 1 with constant population size, scenario 2 with a recent bottleneck event (approx. 1,000 - 3,000 BP), scenario 3 with an LGM bottleneck event (approx. 15,000 - 22,000 BP), and Scenario 4 with a bottleneck during the penultimate glacial. ”0” corresponds to present time.

Table S3.2. Prior distributions of demographic and historical parameters used for the population size change simulations in DIYABC (Figure S3.1). N_e, effective population size, N_{BTN}, effective population size during bottleneck event; time events for population size changes are considered from present (0) backwards in time: t_1, beginning of third millennium BP rainforest disturbance; t_2, end of third millennium BP rainforest disturbance; t_3, beginning of LGM bottleneck; t_4, end of LGM bottleneck; t_5, beginning of penultimate glacial bottleneck; t_6, end of penultimate glacial bottleneck. Mean μ, mean mutation rate of microsatellites; Mean P, mean parameter of geometric distribution (GSM, Generalized Stepwise Mutation Model). All time events are given in generations, assuming a generation time of 100 years. Conditions: $t_6 > t_5$, $t_4 > t_3$, $t_2 > t_1$.

Parameter	Prior distribution
N_e	Uniform (1,000-50,000) [a]
N_{BTN}	Uniform (10-500)
t_1	Uniform (10-30)
t_2	Uniform (20-40)
t_3	Uniform (100-180)
t_4	Uniform (150-220)
t_5	Uniform (1100-1300)
t_6	Uniform (1700-2000)
Mean μ	Log uniform (1E-06-1E-03)
Mean P	Uniform (0.1-0.5)

[a]The effective population size of *Pinus taeda* has been estimated to be ~100,000 (Willyard *et al*., 2007). This tree has a wide distribution range in the SE of the USA and grows at much higher density than *Symphonia globulifera*. Setting the upper bound for N_e to 200,000 did not change the results (not shown).

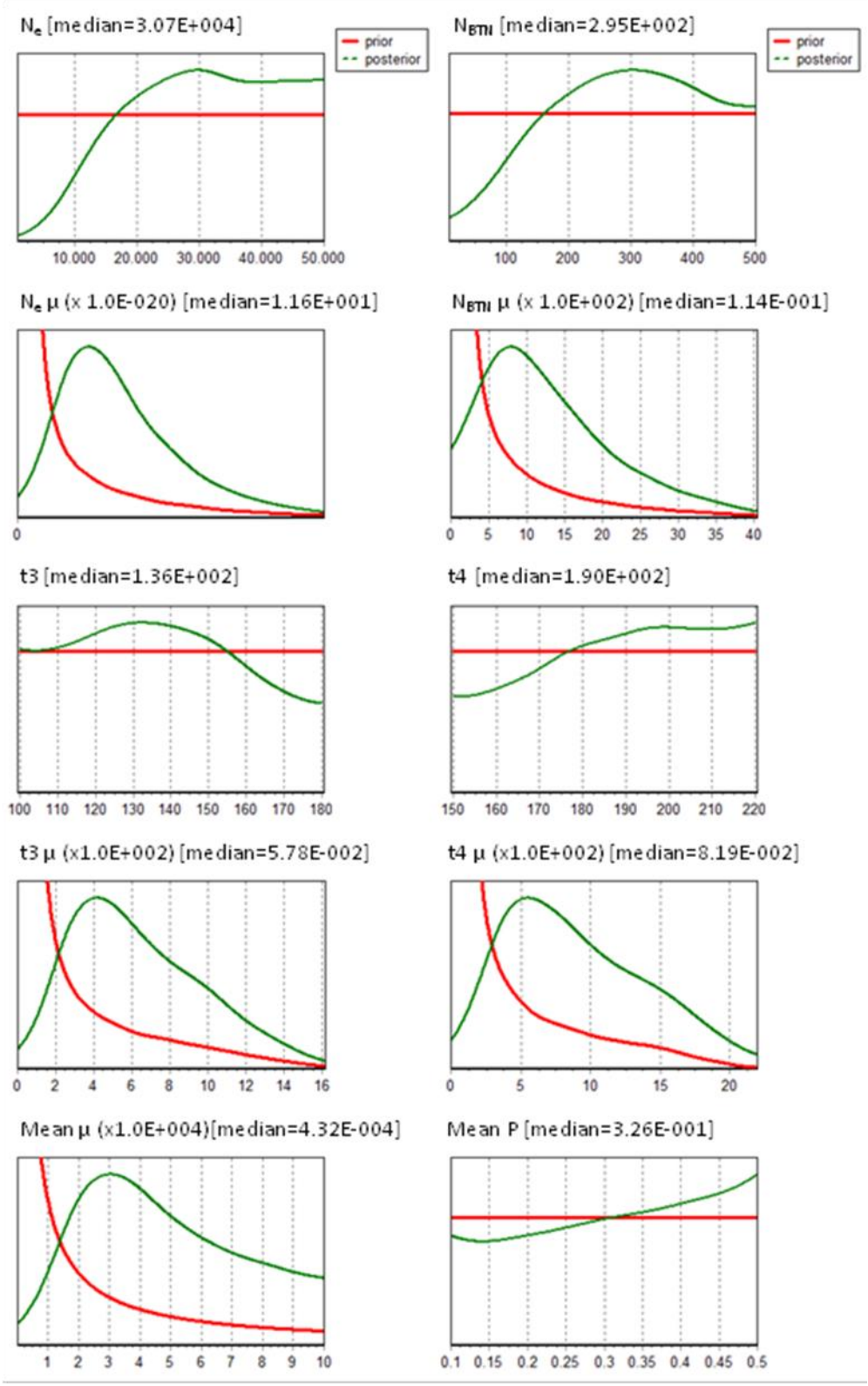

Figure S3.3. Posterior parameter estimates for the West Cameroon region under the best-fitting demographic scenario: scenario 3 (bottleneck coinciding approximately with the LGM). Note that there is a better fit for scaled parameters. The posterior parameter estimates had similar fit in the other regions (results not shown).

Supplementary information S4. Population divergence scenarios compared with Approximate Bayesian Computation.

Table S4.1. Distance matrix of R_{ST} estimates between representative populations of the four gene pools used to infer an UPGMA population divergence tree (see Scenario 1, Figure 3).

ALL LOCI	Benin	Cameroon Korup	Cameroon SW	São Tomé
Benin		0.410	0.218	0.415
Cameroon Korup	0.410		0.208	0.213
Cameroon SW	0.218	0.208		0.150
São Tomé	0.415	0.218	0.150	

Table S4.2. Distance matrix of $(\delta\mu)^2$ estimates between representative populations of the four gene pools used to infer an UPGMA population divergence tree (see Scenario 1, Figure 3).

ALL LOCI	Benin	Cameroon Korup	Cameroon SW	São Tomé
Benin		152.62	54.08	193.15
Cameroon Korup	152.62		62.59	77.05
Cameroon SW	54.08	62.59		52.72
São Tomé	193.15	77.05	52.72	

Table S4.3. Prior distributions of the demographic and historical parameters used for population divergence scenarios in DIYABC. N_e, effective population size, N_{BTN}, effective population size during bottleneck event; times for population size changes and divergence events are considered from present (0) backwards in time: t_1, beginning of bottleneck; t_2, end of bottleneck[a]; t_3, t_4, t_5 and t_6, time events when populations merged[b]; Mean μ, mean mutation rate of microsatellites; Mean P, mean parameter of geometric distribution (GSM, Generalized Stepwise Mutation Model). All time events are given in generations, assuming a generation time of 100 years.

Parameter	Prior distribution
N_e	Uniform (1000 - 50,000)
N_{BTN}	Uniform (10 - 500)
t_1	Uniform (150 - 1999)[a]
t_2	Uniform (150 - 1999)[a]
t_3	Uniform (2,000 - 50,000)
t_4	Uniform (2,000 - 50,000)
t_5	Uniform (2,000 - 50,000)
t_6	Uniform (2,000 - 50,000)
Mean μ	Log uniform (5E-06 - 5E-03)
Mean P	Uniform (0.1 - 0.5)

[a]These priors were set for a bottleneck to occur from the LGM to the penultimate glacial (marine isotope stages, MIS2 to MIS6, both included) to keep the model at the same time simple and realistic. This is justified as rainforest regression (interspersed with bouts of expansion) was observed in this period in Africa, but was not well in phase between sites (Dupont, 2011).

[b]The date of the emergence of São Tomé island (13 Ma) has been considered for prior definition.

Table S4.4. Posterior distributions of the demographic and historical parameters for population divergence scenarios obtained in DIYABC. N_e, effective population size, N_{BTN}, effective population size during bottleneck event; times for population size changes and divergence events are considered from present (0) backwards in time: t_1, beginning of bottleneck; t_2, end of bottleneck; t_3, t_4, t_5 and t_6, time events when populations merged; Mean μ, mean mutation rate of microsatellites; Mean P, mean parameter of geometric distribution (GSM, Generalized Stepwise Mutation Model).

Scenario 1: $(\delta\mu)^2$-based population tree							
Parameter	mean	median	mode	q025	q050	q950	q975
N_e	2.75E+04	2.72E+04	2.20E+04	6.13E+03	8.18E+03	4.74E+04	4.86E+04
t_1	1.20E+03	1.26E+03	1.71E+03	2.37E+02	3.17E+02	1.88E+03	1.92E+03
N_{BTN}	3.38E+02	3.59E+02	4.68E+02	8.41E+01	1.20E+02	4.88E+02	4.94E+02
t_2	1.29E+03	1.37E+03	1.93E+03	3.25E+02	4.12E+02	1.95E+03	1.97E+03
t_3	1.49E+04	1.33E+04	6.41E+03	2.72E+03	3.34E+03	3.21E+04	3.55E+04
t_4	2.29E+04	2.21E+04	1.93E+04	5.81E+03	7.50E+03	4.09E+04	4.36E+04
t_5	3.49E+04	3.64E+04	4.92E+04	1.25E+04	1.58E+04	4.88E+04	4.94E+04
Mean μ	1.98E-04	1.50E-04	8.60E-05	4.10E-05	5.00E-05	5.05E-04	6.37E-04
Mean P	1.96E-01	1.60E-01	1.00E-01	1.00E-01	1.02E-01	4.19E-01	4.65E-01
N_e μ	4.55E+00	3.61E+00	2.07E+00	1.13E+00	1.34E+00	1.09E+01	1.37E+01
t_1 μ	2.35E-01	1.72E-01	7.60E-02	2.40E-02	3.30E-02	6.61E-01	8.24E-01
N_{BTN} μ	6.64E-02	4.80E-02	2.20E-02	8.00E-03	1.10E-02	1.78E-01	2.33E-01
t_2 μ	2.60E-01	1.85E-01	7.60E-02	2.80E-02	3.90E-02	7.29E-01	9.21E-01
t_3 μ	2.57E+00	1.88E+00	8.30E-01	3.70E-01	4.70E-01	6.88E+00	8.93E+00
t_4 μ	4.07E+00	3.08E+00	1.97E+00	7.20E-01	9.20E-01	1.04E+01	1.31E+01
t_5 μ	6.46E+00	4.98E+00	3.13E+00	1.26E+00	1.61E+00	1.59E+01	2.02E+01
Scenario 2: simultaneous divergence							
N_e	2.33E+04	2.18E+04	1.34E+04	3.25E+03	4.57E+03	4.62E+04	4.81E+04
t_1	1.25E+03	1.33E+03	1.75E+03	2.52E+02	3.47E+02	1.89E+03	1.93E+03
N_{BTN}	3.51E+02	3.75E+02	4.98E+02	9.40E+01	1.30E+02	4.91E+02	4.96E+02
t_2	1.39E+03	1.49E+03	1.97E+03	3.67E+02	4.87E+02	1.96E+03	1.98E+03
t_6	1.69E+04	1.36E+04	3.17E+03	2.46E+03	2.87E+03	4.17E+04	4.58E+04
Mean μ	3.34E-04	2.29E-04	1.08E-04	5.50E-05	6.80E-05	9.51E-04	1.28E-03
Mean P	2.44E-01	2.17E-01	1.00E-01	1.00E-01	1.07E-01	4.61E-01	4.83E-01
N_e μ	5.70E+00	4.40E+00	2.00E+00	1.30E+00	1.50E+00	1.42E+01	1.85E+01
t_1 μ	4.10E-01	2.74E-01	1.18E-01	3.50E-02	5.20E-02	1.20E+00	1.60E+00
N_{BTN} μ	1.16E-01	7.60E-02	3.00E-02	1.30E-02	1.70E-02	3.37E-01	4.74E-01
t_2 μ	4.73E-01	3.07E-01	1.25E-01	4.40E-02	6.30E-02	1.41E+00	1.93E+00
t_6 μ	4.06E+00	3.01E+00	1.33E+00	6.10E-01	7.80E-01	1.08E+01	1.37E+01

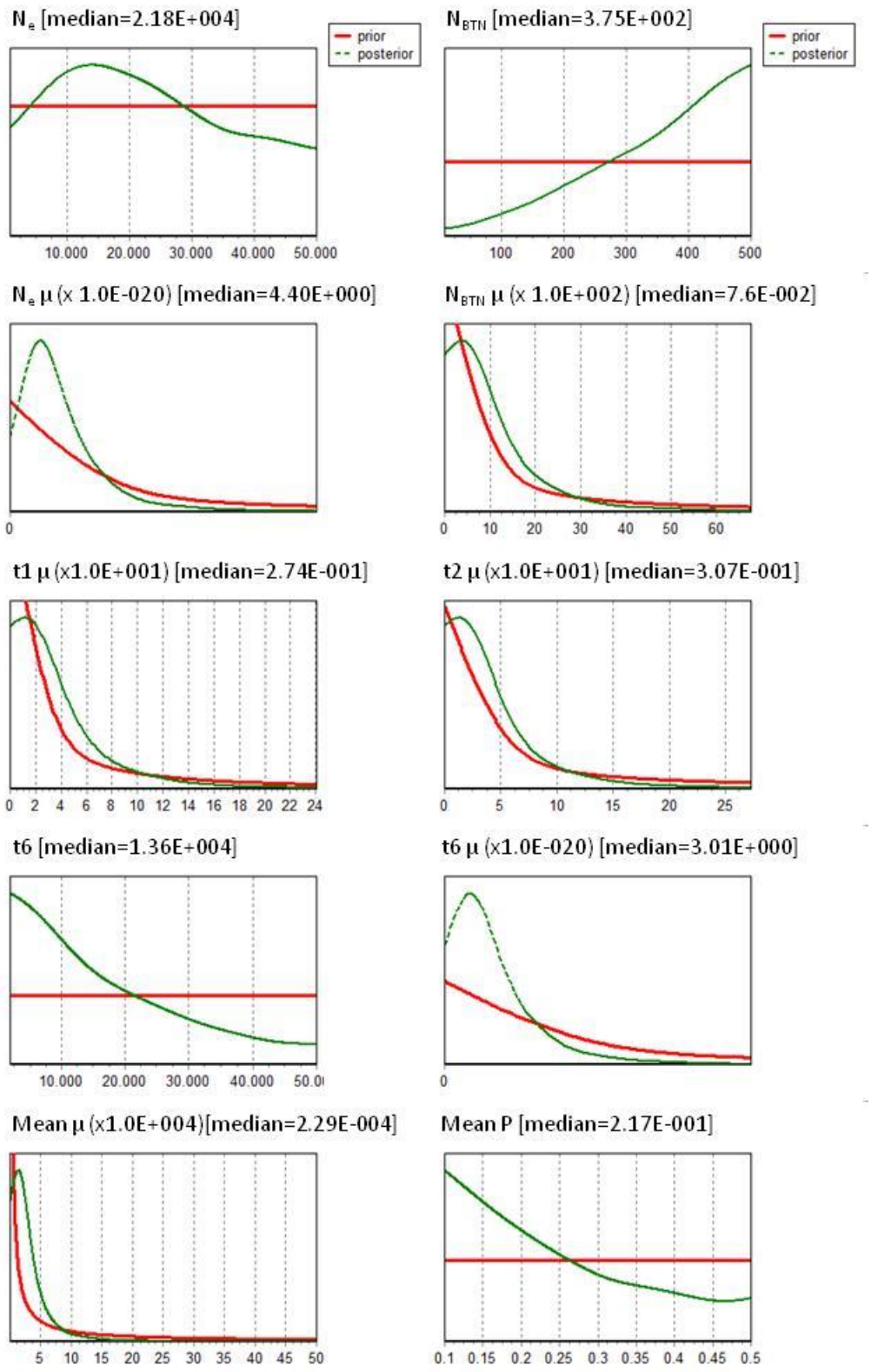

Figure S4.5. Prior and posterior distributions of some divergence parameters obtained for divergence scenario 2, based on the simultaneous divergence of representative populations.

Supplementary information S5. *PsbA-trnH* plastid DNA haplotype definition and haplotype distribution in *Symphonia globulifera*.

Table S5.1. Plastid DNA haplotype definition in *Symphonia globulifera* based on sequences of the *psbA-trnH* intergenic spacer region. *n*, sample size; *psbA-trnH* SNPs: positions of single nucleotide polymorphisms on the sequence.

Hap	n	SNP positions in *psbA-trnH*																						Genbank accession
		70	77	101	107	160	365	366	421	449	467	475	481	501	502	508	528	544	552	586	649	671	673	
H01	3	T	G	T	A	C	T	A	C	T	T	G	T	T	A	T	A	A	G	G	A	G	T	
H02	1	T	G	T	A	C	T	T	C	A	T	G	T	T	A	T	A	A	G	G	A	G	T	
H03	5	T	G	T	A	C	T	A	T	T	T	G	T	T	A	T	A	C	G	G	A	G	T	
H04	13	T	G	T	A	C	T	A	C	T	T	G	T	T	A	T	C	A	G	G	A	G	A	
H05	4	T	G	G	A	C	T	A	C	A	T	G	T	T	A	T	A	A	G	G	A	C	T	
H06	5	T	G	T	A	C	T	A	C	A	T	G	T	T	T	T	A	A	G	G	A	C	T	
H07	10	T	G	T	A	C	T	A	C	A	T	G	T	T	A	T	A	A	G	G	A	G	T	
H08	1	T	G	T	A	C	G	A	C	A	T	G	G	T	A	T	A	A	G	G	A	G	T	
H09	14	T	G	T	A	G	T	A	C	A	T	G	T	T	A	T	A	A	G	G	A	G	T	
H10	2	T	G	T	A	C	T	A	C	A	T	G	T	T	A	G	A	A	G	G	A	G	T	
H11	1	T	G	T	G	C	T	A	C	A	T	G	T	T	A	T	A	A	G	G	A	G	T	
H12	1	T	G	G	A	C	T	A	C	T	T	G	T	T	A	T	A	A	G	G	T	G	T	
H13	14	T	G	T	A	C	T	A	C	T	T	G	T	T	A	T	A	A	G	G	A	G	G	
H14	1	T	G	T	A	C	T	A	C	T	G	T	T	T	T	T	A	A	G	G	A	C	T	
H15	2	T	G	T	A	C	T	A	C	T	T	G	T	T	A	T	T	A	G	G	A	C	T	
H16	3	T	G	T	A	C	T	A	C	T	T	G	T	T	A	T	A	A	G	C	A	C	T	
H17	6	T	G	T	A	C	T	A	C	T	T	G	T	T	T	T	A	A	G	G	A	C	T	
H18	1	T	A	T	A	C	T	A	C	A	T	G	T	T	A	T	A	A	G	G	A	G	T	
H19	1	A	G	T	A	C	T	A	C	A	T	G	T	T	T	T	A	A	G	G	A	C	T	
H20	1	T	G	T	A	C	T	A	C	A	T	G	T	A	A	T	A	A	T	G	A	G	T	
H21	3	T	G	T	A	C	T	A	C	A	T	G	T	T	A	T	A	A	G	A	A	G	T	
H22	2	T	G	T	A	C	T	A	C	A	T	G	T	A	A	T	A	A	G	A	A	G	T	
Clusia grandiflora		T	C	T	A	C	T	A	C	A	T	C	A	-	-	-	-	-	-	-	-	-	-	

Table S5.2. Distribution of plastid DNA haplotypes (*psbA-trnH* intergenic spacer) in *Symphonia globulifera*. Latitude and longitude are given in decimal degrees; negative latitudes indicate southern hemisphere, all longitudes are eastern hemisphere.

	Location	Latitude	Longitude	H01	H02	H03	H04	H05	H06	H07	H08	H09	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22
1	Benin	6.4426	2.4051				13																		
2	Cameroon Korup	5.0615	8.8370													14									
5	Cameroon S1	2.7430	13.8142	2															3						
6	Cameroon S2	2.9666	12.9650										2				1								
7	Cameroon SE	2.7615	15.0497	1														2							
8	Cameroon SW	3.1230	10.5340						5	1										6		1			
9	Cameroon W	4.7843	9.6816																				1		
10	Cameroon W coast	4.0152	9.2132							1															
12	Gabon NE1	0.8101	13.0045							2															
14	Gabon NW	0.4726	10.2711					4																	
15	Gabon coast	0.4147	9.3349									14													
18	Gabon Waka	-1.2153	11.3054							1											1				
20	Gabon SE	-1.8629	13.8771												1										
21	Gabon SW	-2.6595	11.2501		1					1															
22	Gabon SW coast	-2.7791	10.0640							4	1			1											
23	DRC E1	-2.4666	16.5707																					2	2
24	DRC E2	-1.1186	17.9137																					1	
26	Sao Tome	0.2740	6.5524			5																			

Supplementary Information S6. Bayesian Clustering analyses on *Symphonia globulifera* SSRs.

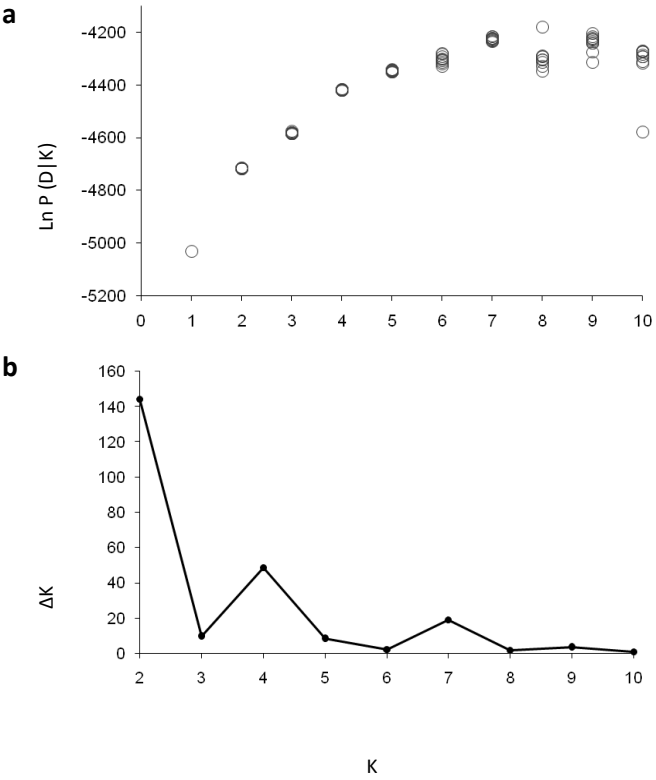

Figure S6.1. Characteristics of the STRUCTURE analysis (Pritchard *et al*., 2000). We used an admixture model with correlated allele frequencies with burn-in of 100,000 and run length of 200,000 iterations and carried out 10 repetitions for each *K*. a) Posterior log-likelihood of data for *K* clusters; b) Application of the *ad hoc* method by Evanno *et al.* (2005) to identify the best *K*.

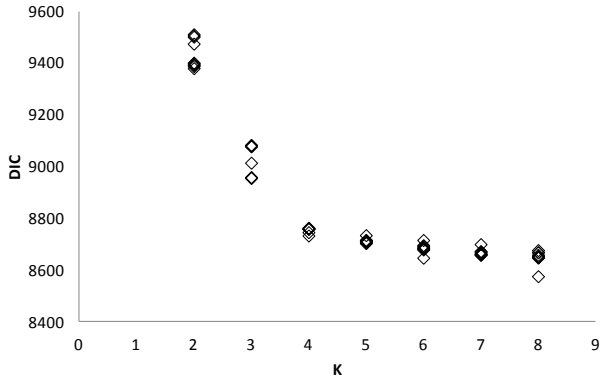

Figure S6.2. Deviance information criterion (DIC) for *K* clusters obtained from the TESS 2.3.1 analysis (Chen *et al*., 2007) on the *Symphonia globulifera* SSR data. We used a conditional autoregressive admixture model with burn-in of 10,000 and run length of 50,000 iterations, a spatial interaction parameter of 0.99 and a trend degree surface of 1. Ten repetitions were carried out for each *K*. A plateau beginning at *K*=4 indicates that *K*=4 is the simplest correct clustering solution for the data.

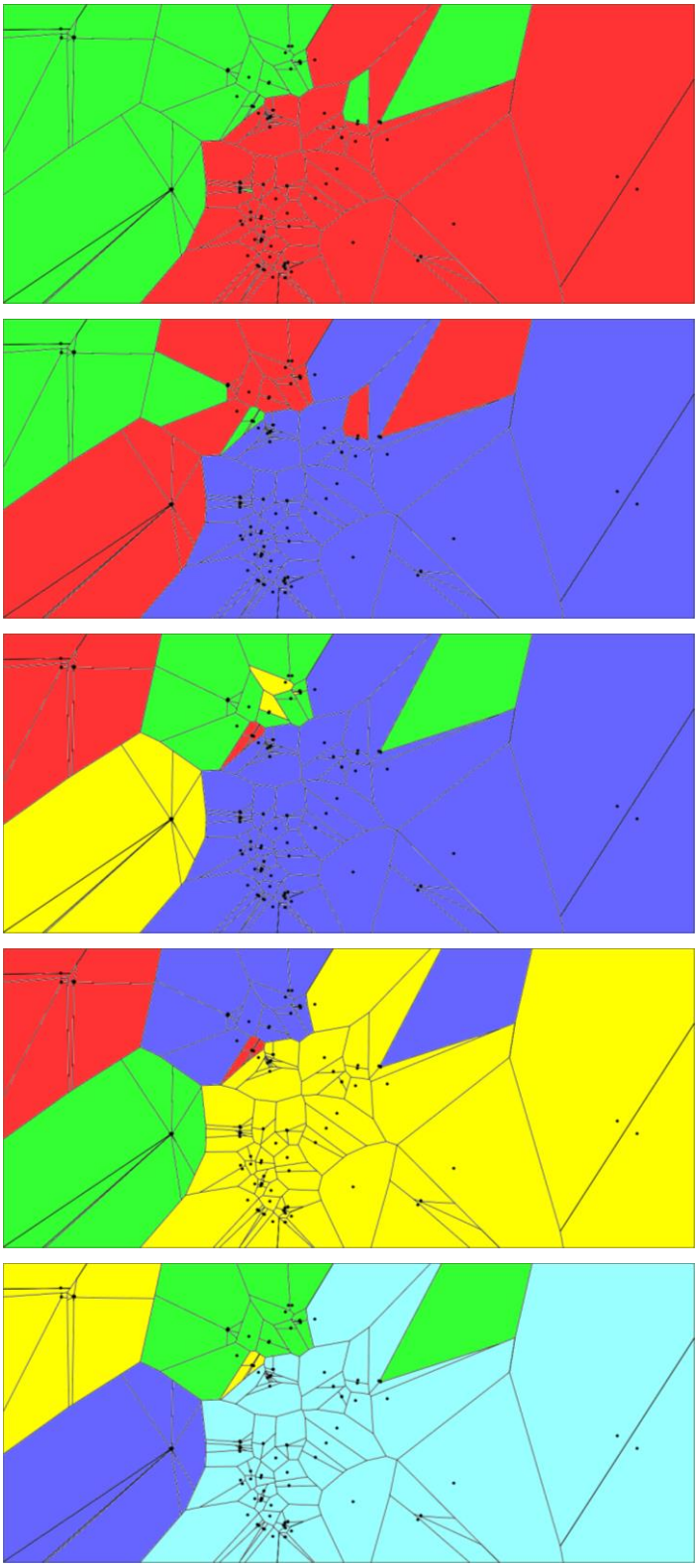

Figure S6.3. Hard clustering results for *K*=2 to *K*=6 clusters obtained from the TESS 2.3.1 analysis (Chen *et al*., 2007) on the *Symphonia globulifera* SSR data. The geographic assignment of individuals is shown for the run with the lowest deviance information criterion for each *K*, as obtained from a conditional autoregressive admixture model (see Figure S6.2).

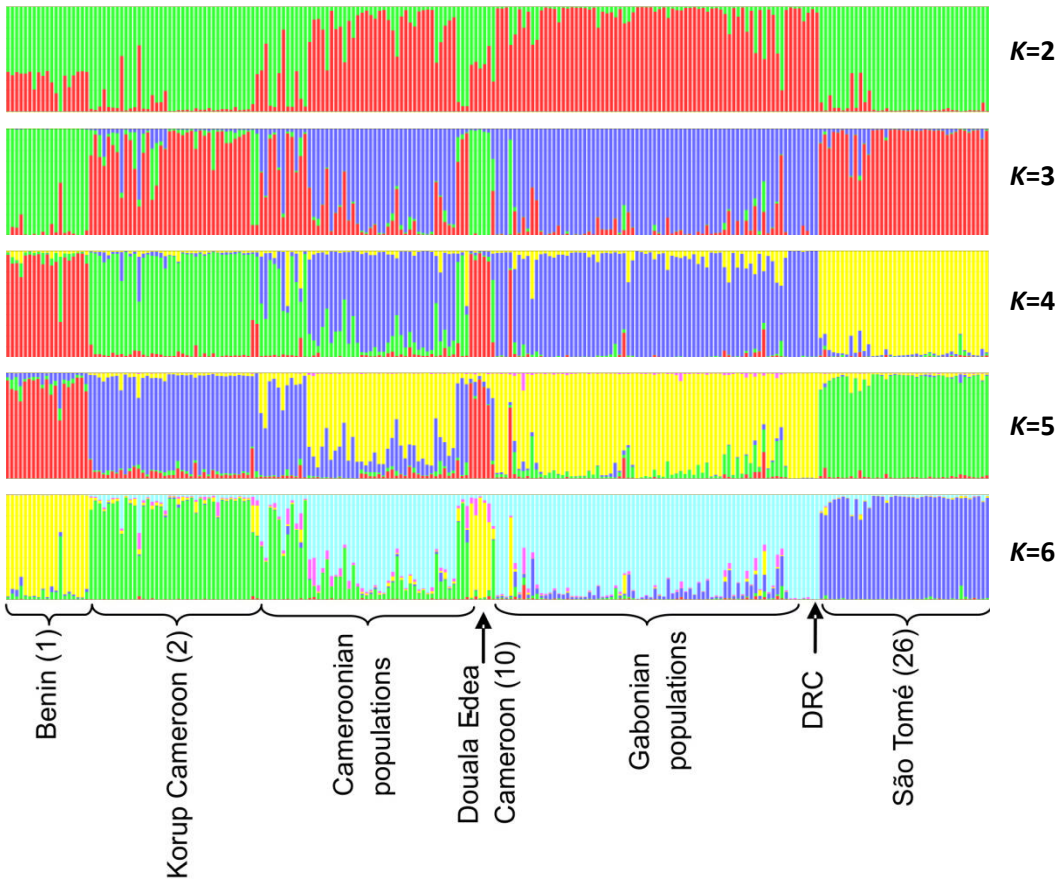

Figure S6.4. Ancestry proportions for *K*=2 to *K*=6 clusters obtained from the TESS 2.3.1 analysis (Chen *et al*., 2007) on the *Symphonia globulifera* SSR data. The assignment of individuals is shown for the run with the lowest deviance information criterion for each *K*, as obtained from a conditional autoregressive admixture model using a a spatial interaction parameter of 0.99 and a trend degree surface of 1 (see Figure S6.2).

References

Chapuis MP, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Mol Biol and Evol* **24**: 621-631.

Chen C, Durand E, Forbes F, François O (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Mol Ecol Notes* **7**: 747-756.

Chybicki IJ, Burczyk J (2009). Simultaneous estimation of null alleles and inbreeding coefficients. *J Hered* **100**: 106-113.

Cornuet JM, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001-2014.

David P, Pujol B, Viard F, Castella V, Goudet J (2007). Reliable selfing rate estimates from imperfect population genetic data. *Mol Ecol* **16**: 2474-2487.

Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc* **39**: 1-38.

Dupont L (2011). Orbital scale vegetation change in Africa. *Quaternary Sci Rev* **30**: 3589-3602.

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

Piry S, Luikart G, Cornuet JM (1999). BOTTLENECK: a computer program for detecting recent reduction in the effective population size using allele frequency data. *J Hered* **90**: 502-503.

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

Willyard A, Syring J,Gernandt DS, Liston A, Cronn R (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for Pinus. *Mol Biol Evol* **24**: 90-101.

Supplementary information Chapter 3

Effects of fire regime on the population genetics of natural pine stand

S1 Population genetic structure

An Bayesian clustering algorithm implemented in STRUCTURE 2.2 (Pritchard 2000) was used on the SSR data to identify population genetic structure in the study region. We ran an admixture model with correlated allele frequencies between clusters. Ten runs were performed for each number of clusters *K*=1 to *K*=5 with a burn-in length of 100,000 and a run length of 200,000 iterations.

The STRUCTURE analysis revealed a single gene pool in the study region for each species (Figures S2 and S3) as expected from the post-glacial history of these species (see Introduction).

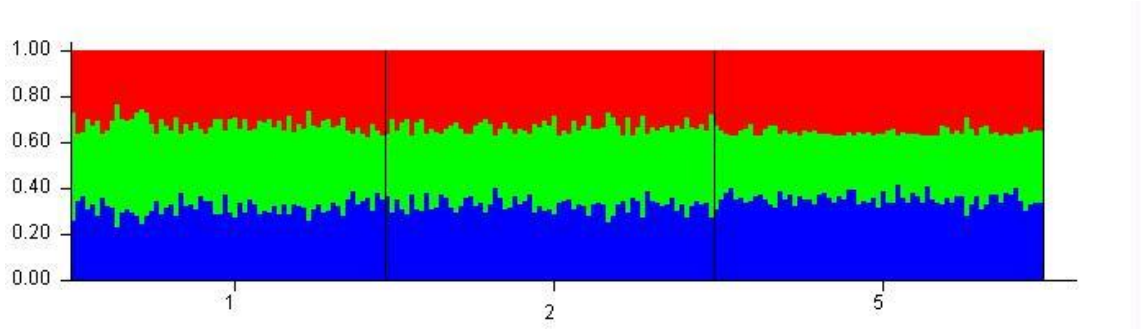

Fig. S1.1 Assignment proportions calculated in STRUCTURE for K=3 clusters for three stands of *Pinus halepensis* (1, Serra Calderona; 2, Sinarcas; 5= Eslida). Each individual is represented as a line segment which is vertically partitioned into *K* colored components representing the individual's estimated proportions of ancestry in the *K* clusters.

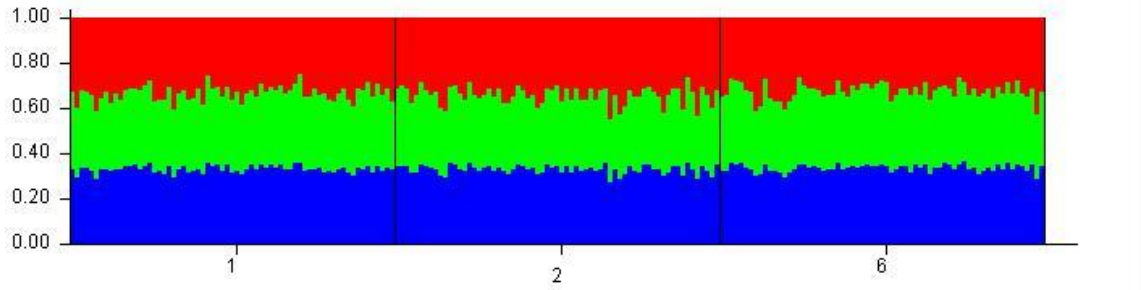

Fig. S1.2 Assignment proportions calculated in STRUCTURE for K=3 clusters for three stands of *Pinus pinaster* (1, Serra Calderona; 2, Sinarcas; 6, Eslida). Each individual is represented as a line segment which is vertically partitioned into *K* colored components representing the individual's estimated proportions of ancestry in the *K* clusters.

S2 Sample locations

The study stands where needle material was sampled from a total of 197 *Pinus halepensis* and 199 *P. pinaster* trees are labeled in red. Stands labeled in black were additionally included in the study for *P. halepensis* only. The stands closer to the Mediterranean coast, i.e., Alzira, Cabanes, Eslida, Serra Calderona and Serra d'Irta belong to a region with high frequency of crown fires (HiFi) whereas Sinarcas, Titaguas and Montán are located in a region with low frequency of crown fires (LoFi).

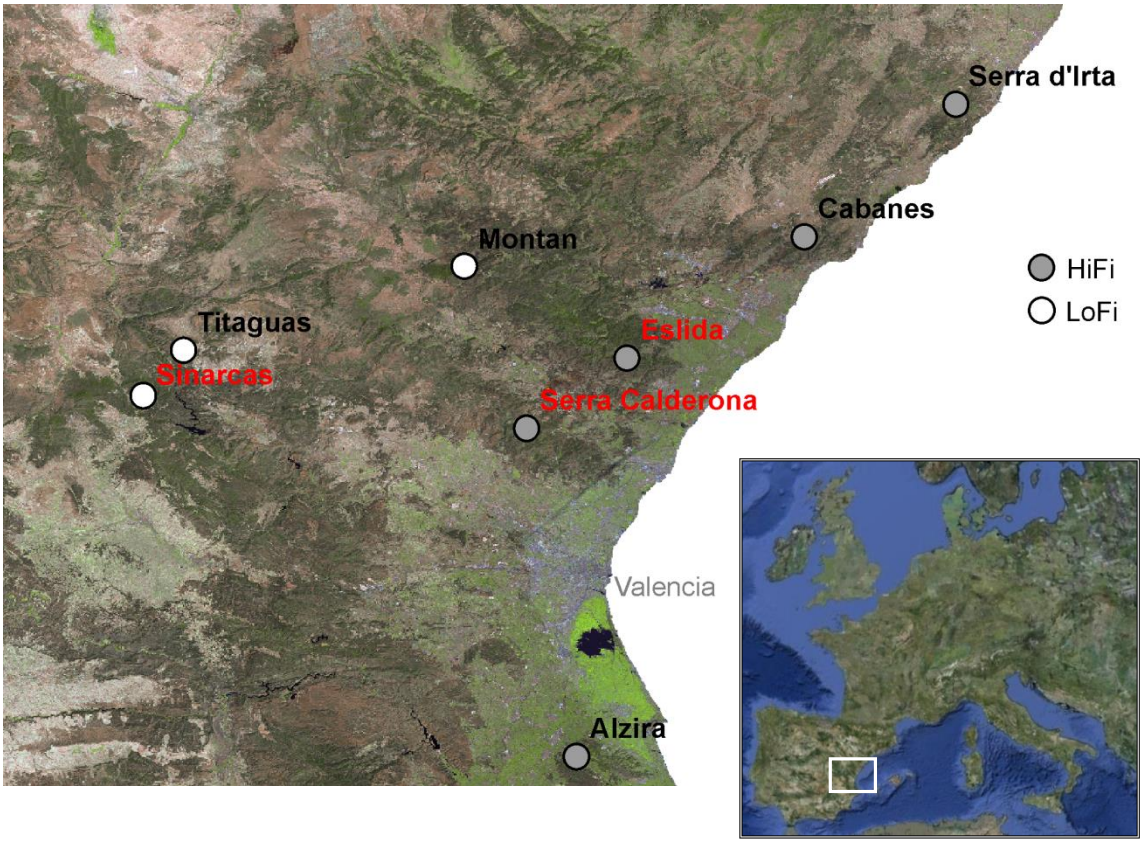

Fig. S2.1 Study stands *of Pinus pinaster* and *P. halepensis* in the Eastern Iberian Peninsula (stand names in red). HiFi stands are marked with a grey circle while LoFi stands are marked with a white circle. Sample stands with names in black letters indicate the position of additionally included *P. halepensis* stands.

Table S2.1 Central coordinates of the sampled stands in decimal degrees (WGS84).

		Pinus pinaster		*Pinus halepensis*	
Stand name		Latitude	Longitude	Latitude	Longitude
Sinarcas	LoFi	39.74769	-0.49667	39.74162	-0.48395
Serra Calderona	HiFi	39.78971	-1.20448	39.79760	-1.20414
Eslida	HiFi	39.87918	-0.29654	39.87070	-0.29247
Alzira	HiFi			39.12415	- 0.38785
Cabanes	HiFi			40.10132	0.04182
S. d'Irta	HiFi			40.35106	0.32312
Montan	LoFi			40.04873	-0.59526
Titaguas	LoFi			39.88786	-1.12823

Table S1.2 Pair-wise distance between sample stands in kilometers for *Pinus pinaster* (above diagonal) and *P. halepensis* (below diagonal).

	Sinarcas	Calderona	Eslida
Sinarcas		60.82	78.35
Calderona	62.02		22.52
Eslida	78.47	21.78	

S3 Details on genetic markers

Table S3.1 Details on microsatellite locus combinations in multiplex PCR mixes and labelling of forward primers.

Locus	Species	Multiplex-mix	Label	Fragment size range (bp)
NZPR5448	*P. halepensis*	1	FAM	243-251
Epi3	*P. halepensis*	1	NED	219-222
FRP94	*P. halepensis*	1	FAM	143-145
pEST2669	*P. halepensis*	1	NED	140-146
Itph4516	*P. halepensis*	1	PET	134-174
B4F08	*P. halepensis*	2	FAM	177-199
PtTX3030	*P. halepensis*	2	PET	348-357
pEST8	*P. halepensis*		FAM	146-150
PtTX3116	*P. halepensis*			131-138
ITPH4516_3	*P. pinaster*	1	PET	148-182
A6F03_4	*P. pinaster*	1	VIC	234-254
rptest11	*P. pinaster*	1	NED	204-213
Ctg4363_11	*P. pinaster*	2	VIC	184-200
NZPR1078_5	*P. pinaster*	2	PET	329-341
epi3	*P. pinaster*	2	NED	219-239
FRPP94	*P. pinaster*	2	FAM	132-160
NZPR413	*P. pinaster*	3	FAM	163-195
gPp14	*P. pinaster*	3	VIC	192-210
pEST2669	*P. pinaster*	3	NED	141-165
epi5	*P. pinaster*	3	PET	186-204

Table S3.2 Characterization of nine SSR loci for all *Pinus halepensis* samples: N_A, number of alleles and F_{IS}, fixation index. F_{IS} is followed by significance levels from 10,000 permutations and after Bonferroni multiple test correction: n.s., not significant.

Locus	*missing data [%]*	N_A	*Allele size range [bp]*	F_{IS}		
				Serra Calderona	Sinarcas	Eslida
ALL LOCI		44		0.044 n.s.	-0.073 n.s.	-0.006 n.s.
NZPR5448	1.5	3	244-254	-0.231 n.s.	-0.333 n.s.	-0.419 n.s.
Epi3	0.5	3	219-222	0.162 n.s.	-0.027 n.s.	-0.121 n.s.
FRP94	0.0	3	143-151	0.283 n.s.	-0.003 n.s.	0.309 n.s.
pEST2669	0.0	4	140-146	0.181 n.s.	-0.049 n.s.	-0.044 n.s.
Itph4516	1	16	134-174	0.075 n.s.	0.003 n.s.	-0.042 n.s.
pEST8	0	3	147-151	-0.098 n.s.	-0.171 n.s.	0.253 n.s.
PtTX3116	0	2	131-137	0.02 n.s.	-0.171 n.s.	0.288 n.s.
PtTX3030	1.5	5	348-357	0.048 n.s.	0.103 n.s.	0.022 n.s.
B4F08	1	5	177-199	0.027 n.s.	-0.051 n.s.	-0.103 n.s.

Table S3.3 Characterization of 11 SSR loci for all *Pinus pinaster* samples: N_A, number of alleles and F_{IS}, fixation index F_{IS} is followed by significance levels from 10,000 permutations and after Bonferroni multiple test correction: n.s., not significant.

Locus	*missing data [%]*	N_A	*Allele size range [bp]*	F_{IS}		
				Serra Calderona	*Sinarcas*	*Eslida*
ALL LOCI		86		0.056 n.s.	0.031 n.s.	0.016 n.s.
ITPH4516_3	3.5	14	150-182	0.03 n.s.	0.005 n.s.	-0.094 n.s.
A6F03_4	1.5	9	234-254	0.122 n.s.	0.013 n.s.	0.156 n.s.
rptest11	0.0	4	204-213	-0.018 n.s.	-0.006 n.s.	-0.085 n.s.
Ctg4363_11	0.0	8	84-100	-0.012 n.s.	0.062 n.s.	0.072 n.s.
NZPR1078_5	4.0	4	329-341	0.13 n.s.	0.055 n.s.	0.088 n.s.
epi3	3.0	10	219-239	0.027 n.s.	-0.062 n.s.	0.014 n.s.
FRPP94	10.9	11	132-160	0.18 n.s.	0.123 n.s.	0.036 n.s.
NZPR413	0.0	8	163-195	-0.106 n.s.	0.023 n.s.	0.009 n.s.
gPp14	2.5	5	192-210	-0.085 n.s.	-0.041 n.s.	-0.042 n.s.
pEST2669	2.5	6	141-165	0.166 n.s.	0.149 n.s.	-0.008 n.s.
epi5	2.5	7	186-204	0.092 n.s.	-0.071 n.s.	-0.054 n.s.

Table S3.4 Details on 251 polymorphic SNPs genotyped successfully in *Pinus pinaster* and *P. halepensis*.

	Pinus pinaster		***Pinus halepensis***	
Pop.	*# SNPs*	*# monomorphic SNPs*	*# SNPs*	*# monomorphic SNPs*
Calderona	251	11	251	10
Sinarcas	251	7	251	16
Eslida	251	19	251	31

S4 Summary statistics for skyline-plots

Table S4.1 Summary statistics calculated in arlsumstat for simulated and observed data.

	Computation of all statistics at the population level
K	Mean number of alleles over loci
Ksd	Standard deviation of the number of alleles over loci
H	Mean heterozygosity over loci
Hsd	Standard deviation of the heterozygosity over loci
GW	Mean Garza-Williamson statistic over loci
GWsd	Standard deviation of the Garza-Williamson index over loci
R	Mean allelic range over loci
Rsd	Standard deviation over loci of allelic range

S5 Details on spatial genetic structure analysis and spatially explicit simulations

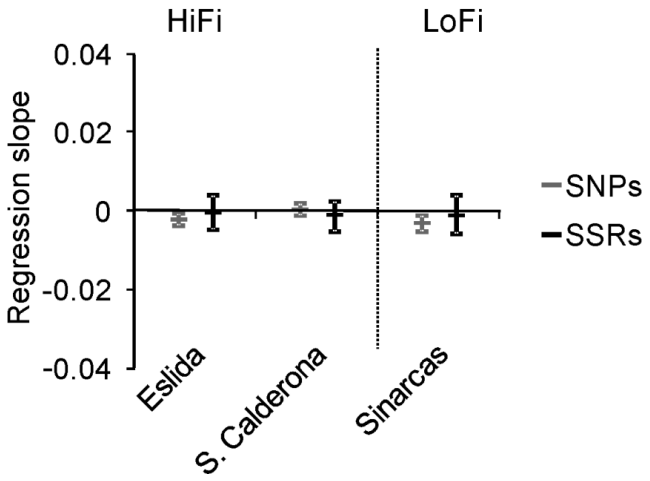

Fig. S5.1 Comparison of SGS between 11 SSR and 251 SNP markers in *Pinus pinaster*. Mean values represent the regression slopes of kinship – distance plots and error bars depict the 95% jackknife confidence interval. A negative regression slope indicates significant SGS and non-overlapping confidence intervals between markers would suggest that different evolutionary processes operate on SSRs vs. SNPs. S. Calderona: Serra Calderona.

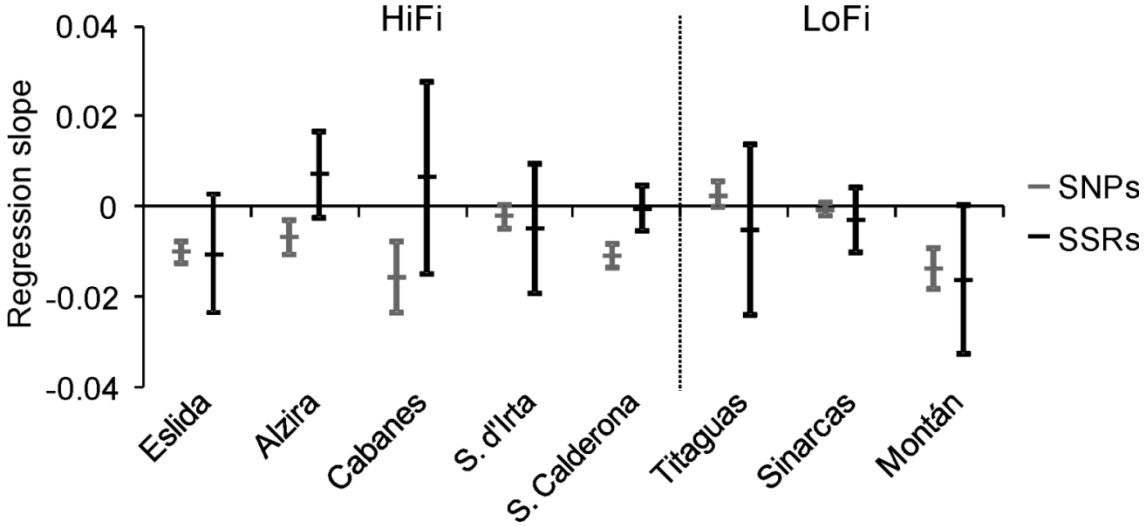

Figure S5.2 Comparison of SGS between nine SSR and 251 SNP markers in *Pinus halepensis*. Mean values represent the regression slopes of kinship – distance plots and whiskers depict the 95% jackknife confidence interval. A negative regression slope indicates significant SGS and non-overlapping confidence intervals between markers suggest that different evolutionary processes operate on SSRs vs. SNPs. S. d'Irta: Serra d'Irta; S. Calderona: Serra Calderona.

Spatially explicit simulations

To compare the power to detect and quantify SGS with both marker types (SSRs vs. SNPs), we conducted spatially explicit simulations of a reproducing and dispersing population comprising a grid of 100 x 100 individuals using the *simnew* program written by OJ Hardy (Heuertz *et al.* 2003; de-Lucas *et al.* 2009). Simulation parameters were chosen to be realistic for pines while using exaggeratedly restricted dispersal parameters to enhance SGS build-up, because realistic values normally result in low and non-significant SGS (de-Lucas *et al.* 2009). Simulations were conducted using respectively nine or eleven SSRs to mimic the observed *P. halepensis* and *P. pinaster* data sets and variable numbers of SNPs. Genotypes at generation 0 were assigned randomly to all individuals. At each generation, 50% of the simulated trees were replaced by offspring resulting from dispersal according to Gaussian dispersal curves with an axial standard deviation of dispersal for seeds σ_s = 3.77 m, for pollen σ_p = 22.72 m and for genes σ_g=16.5 m (see de-Lucas *et al.* 2009), allowing for 20% pollen immigration and 0% seed immigration per generation. SGS was allowed to build up for 64 generations, running 100 repetitions for each data type. At generations 2, 4, 8, 16, 32, 48 and 64, *b* and *Sp* were computed and the mean and standard deviation of *b* were compared across simulations for both marker types to determine the number of SNPs that yielded similar standard deviation of *b* as the 9 or 11 SSRs used in *P. halepensis* and *P. pinaster*, respectively. We then subsampled (10 independent subsamples) the SNPs to obtain data sets with equivalent power as the SSRs, and compared SGS between fire regimes using the 95% jackknife CIs.

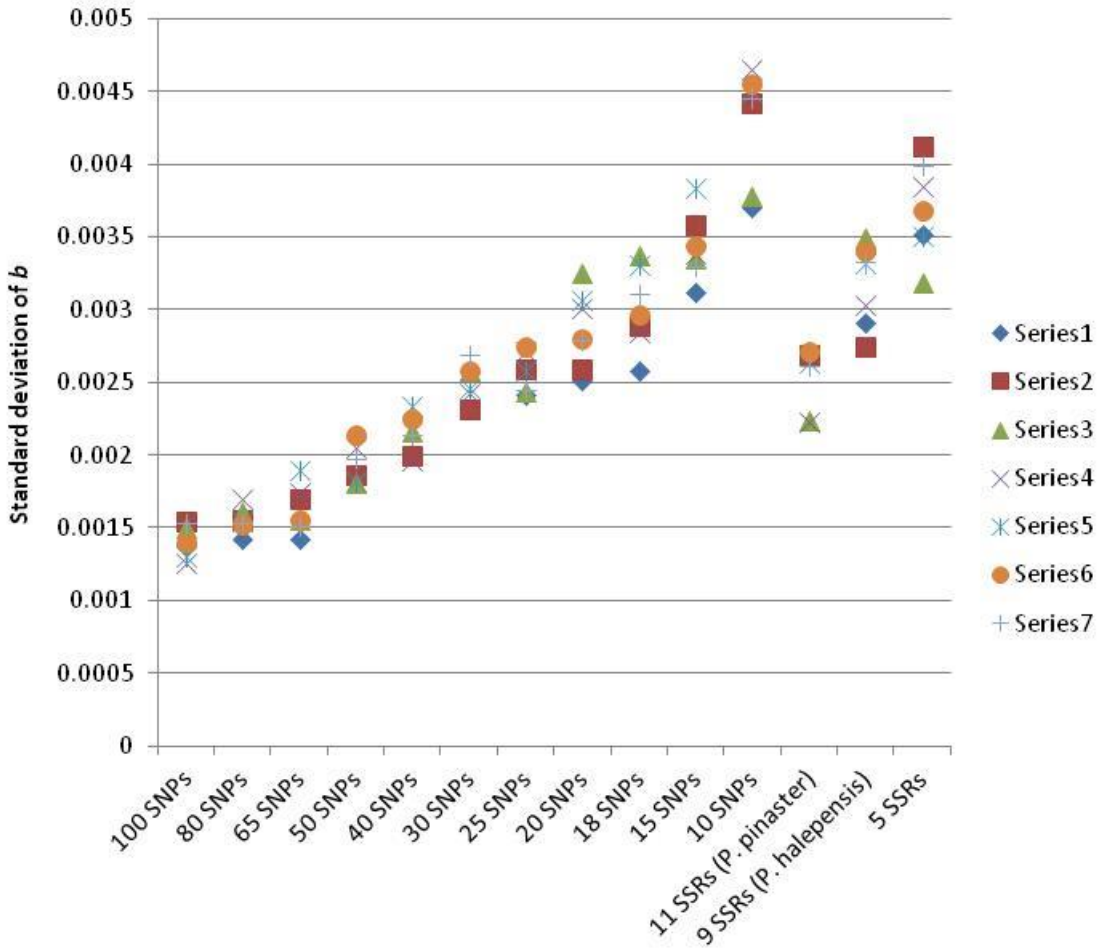

Fig. S5.3 Standard deviation of the regression slope *b* for simulated data sets with variable marker numbers and marker types across 64 generations. Simulated sets of 15 – 25 SNPs have similar power to detect and quantify SGS than the observed SSR data sets in *P. halepensis* and *P. pinaster*. Successive series correspond to simulations at generations 2, 4, 8, 16, 32, 48 and 64.

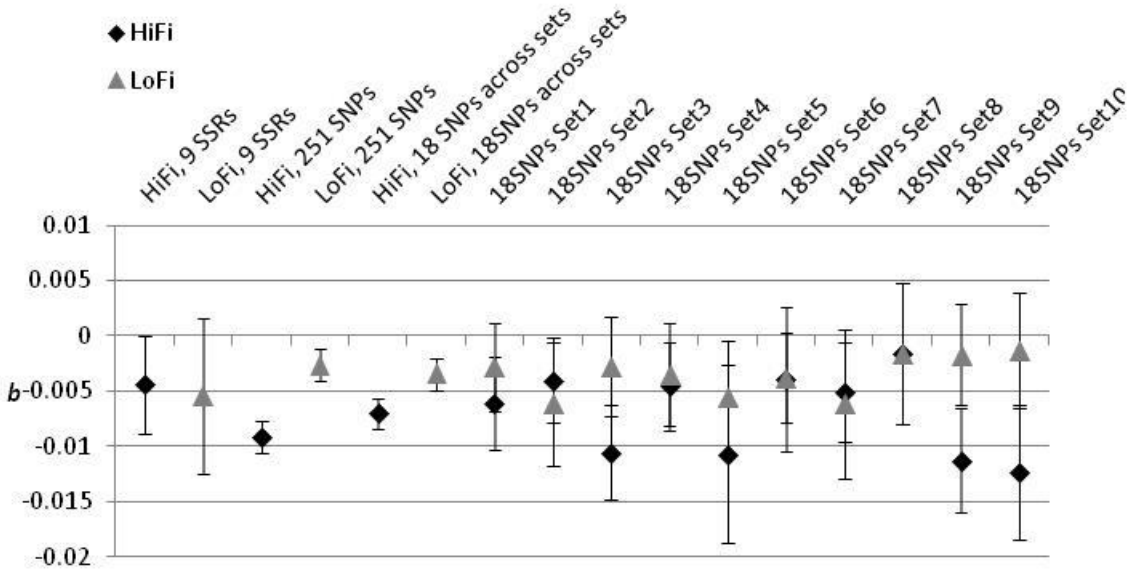

Fig. S5.4 Mean jackknife regression slopes and 95% confidence intervals of within stand SGS across *P. halepensis* HiFi (black diamonds) or LoFi (grey triangles) stands, respectively. HiFi stands had stronger SGS than LoFi stands at SNPs (non-overlapping 95% CIs) but not at SSRs markers, and this trend was confirmed for sub-samples of 18 SNPs that had equivalent power as 9 SSRs. In HiFi stands, SNPs showed a trend for a stronger structure than SSRs, suggesting fire-related mibroenvironmental selection.

Table 5.1 SGS results for 10 SNP subsets (18 SNPs for *P. halepensis* and 25 SNPs for *P. pinaster*) with equivalent statistical variance as the microsatellite markers calculated for each stand in comparison with the results of the whole SNP data set. *Sp,* intensity of SGS, *P*-value, significance values of regression slope, mean *Sp*, mean SGS, SGS intensity estimated over 10 SNP subsets, *SE,* standard error of *Sp* estimates, significant SGS (%), percentage of SNP subsets that yielded significant SGS tests.

		Pinus halepensis				
		251 SNPs		subsamples of 18 SNPs		
Location	Fire regime	*Sp*	*P*-value	*mean Sp*	*SE*	significant SGS (%)
Alzira	HiFi	0.0068	0.0012	0.003	0.0025	80
Cabanes	HiFi	0.016	0.0003	0.018	0.0062	60
S.Calderona	HiFi	0.0112	0	0.0109	0.0017	20
Eslida	HiFi	0.0104	0	0.0147	0.0022	10
Serrad'Irta	HiFi	0.002	0.0951	0.0014	0.0015	90
Montán	LoFi	0.0143	0	0.0118	0.0037	70
Sinarcas	LoFi	0.0005	n.s.	0.0016	0.0012	0
Titaguas	LoFi	-0.0027	n.s.	-0.0041	0.0011	0
		Pinus pinaster				
		251 SNPs		subsamples of 25 SNPs		
Location	Fire regime	*Sp*	*P*-value	*mean Sp*	*SE*	significant SGS (%)
S. Calderona	HiFi	-0.0012	0.5368	-0.0007	0.0004	0
Eslida	HiFi	0.0022	0.0171	0.0016	0.0007	10
Sinarcas	LoFi	0.0036	0.0014	<-0.0001	0.0006	0

References

de-Lucas A, González-Martínez SC, Vendramin GG, Hidalgo E, Heuertz M (2009) Spatial genetic structure in continuous and fragmented populations of Pinus pinaster Aiton. Molecular Ecology, 18, 4564–4576.

Heuertz M, Vekemans X, Hausman JF, Palada M, Hardy OJ (2003) Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. Molecular Ecology, 12, 2483–2495.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945–59.

Supplementary information Chapter 5

In-situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*)

Supporting Information

Table S1 Range-wide SNP genotyping to test for accuracy outside the focal population of the phenotypic model for serotiny constructed in the eastern Iberian Peninsula (see also Figure 1): populations sampled, geographical region, mtDNA lineage (i.e., maternal genetic lineage; Burban & Petit, 2003) and sample size (*N*); the mean observed percentage of serotiny is also provided (retrieved from Tapias *et al*., 2004; see *Material and Methods*).

Population	Geographical region	mtDNA lineage	*N*	Serotiny (%)
Mimizan	Atlantic France	western	19	9.03
Galicia	Atlantic Spain	western	21	0.00
Leiria	Atlantic Portugal	western	24	0.00
Tabuyo del Monte	Central Spain	western	20	43.30
Arenas de S. Pedro	Central Spain	western	27	29.73
Cuéllar	Central Spain	western	28	1.07
Coca	Central Spain	western	18	0.66
San Leonardo de Yagüe	Central Spain	western	20	3.64
Boniches	Eastern Spain	western	11	17.31
Olba	Eastern Spain	western	21	1.19
Cazorla	Southern Spain	western	15	11.11
Cómpeta	Southern Spain	western	4	3.31
Corsica	Corsica (France)	eastern	37	2.61
Tabarka	North Africa (Tunisia)	eastern	21	62.76
Tamrabta	North Africa (Morocco)	Moroccan	24	26.52

Table S2 Illumina VeraCode OPA design file and designability scores provided as Excel file ('GGT_Sequence_v1_Score251109.xls').
See: http://www.uv.es/jgpausas/papers/Budde-2014-NewPhytol_supp.pdf

Table S3 Marker effects of 26 SNPs with *P*<0.05 in the mixed-effects linear models (MLMs) for the additive and allele-dominance genetic models (no significant SNPs were found for the over-dominance model). SNP refers to SNP marker names reported by Chancerel *et al.* (2011) except for *myb2-337*; ns: not significant.

SNP	Genetic effects					
	Additive model			Allele dominance model		
	F	*P*	R^2	*F*	*P*	R^2
m15	6.0523	0.0028	0.0620	6.4522	0.0119	0.0330
m39	3.5752	0.0299	0.4529	ns	ns	ns
m79	5.8002	0.0036	0.0588	11.5088	0.0008	0.0583
m237	5.0627	0.0072	0.0542	9.9162	0.0019	0.0531
m493	7.2695	0.0076	0.0371	ns	ns	ns
m526	ns	ns	ns	3.9169	0.0492	0.3276
m594	3.4602	0.0334	0.0350	ns	ns	ns
m646	4.6751	0.0104	0.0489	9.3322	0.0026	0.0488
m663	4.3030	0.0148	0.0435	8.5825	0.0038	0.0434
m692	6.6319	0.0016	0.0686	ns	ns	ns
m696	ns	ns	ns	5.7223	0.0177	0.0294
m698	ns	ns	ns	4.9335	0.0275	0.0249
m705	3.3688	0.0365	0.0344	ns	ns	ns
m716	3.6902	0.0267	0.4529	ns	ns	ns
m757	3.5286	0.0313	0.0360	ns	ns	ns
m816	ns	ns	ns	5.3404	0.0219	0.0277
m817	3.2994	0.0390	0.4629	ns	ns	ns
m912	3.8011	0.0240	0.0409	5.7912	0.017	0.0312
m942	ns	ns	ns	4.6263	0.0327	0.0234
m955	3.0488	0.0497	0.0310	4.8997	0.028	0.0249
m965	3.5331	0.0311	0.0357	5.7636	0.0173	0.0292
m974	ns	ns	ns	5.3869	0.0213	0.0272
m983	ns	ns	ns	5.4546	0.0205	0.0278
m1194	3.7141	0.0261	0.0376	6.9170	0.0092	0.0350
m1196	ns	ns	ns	3.9690	0.0477	0.0200
myb2-337	3.1416	0.0454	0.0319	ns	ns	ns

Table S4 Gene function, site annotation, and mapping information for the 17 SNPs potentially involved in associations with serotiny.

SNP	Annotation	Putative function (UniProtKB/Swiss-Prot)	Site annotation[c]	Linkage Group
m15	defectively organized tributaries 2 (DOT2)	root, shoot, and flower development	nc	
m289	calcium-dependent protein kinase (CDPK)	signalling, stress response	syn	1[a,b]
m594	pyrophosphate-energized vacuolar membrane proton (AVP)	membrane transport	syn	8[a,b]
m692	unknown	unknown	unk	3[a]
m696	arabinogalactan-like protein (AGP)	cell walls, plant defense, stress response	nc	3[a,b]
m698	nascent polypeptide-associated complex subunit alpha-like protein (NAC-alpha)	protein transport, response to salt stress	syn	
m705	carotenoid cleavage dioxygenase (CCD)	stress response, signalling	nc	
m816	receptor protein kinase clavata1 (CLV1)	cell differentiation	syn	1[a,b]
m817	receptor protein kinase clavata1 (CLV1)	cell differentiation	nc	1[b]
m912	peroxidase 72-like (PER72)	stress response	non-syn	
m955	unknown	unknown	unk	3[a,b]
m959	unknown	unknown	unk	5[b]
m974	1-aminocyclopropane-1-carboxylate synthase (ACC)	ethylene metabolism, stress response	syn	11[b]
m1194	cell division related protein (DnaJ and myb-like DNA-binding domain-containing protein)	cell division	syn	
m1196	peptidyl-prolyl cis-trans isomerase (PPI)	protein folding	syn	
m1211	unknown	unknown	unk	1[b]
m1414	auxin response factor 4 (ARF4)	auxin metabolism, plant growth and development	nc	

[a]LG from Chancerel *et al*. (2011)
[b]LG from De Miguel *et al*. (2012)
[c]Site annotation: syn, synonymous; non-syn, non-synonymous; nc, non-coding (UTRs or introns); unk, unknown

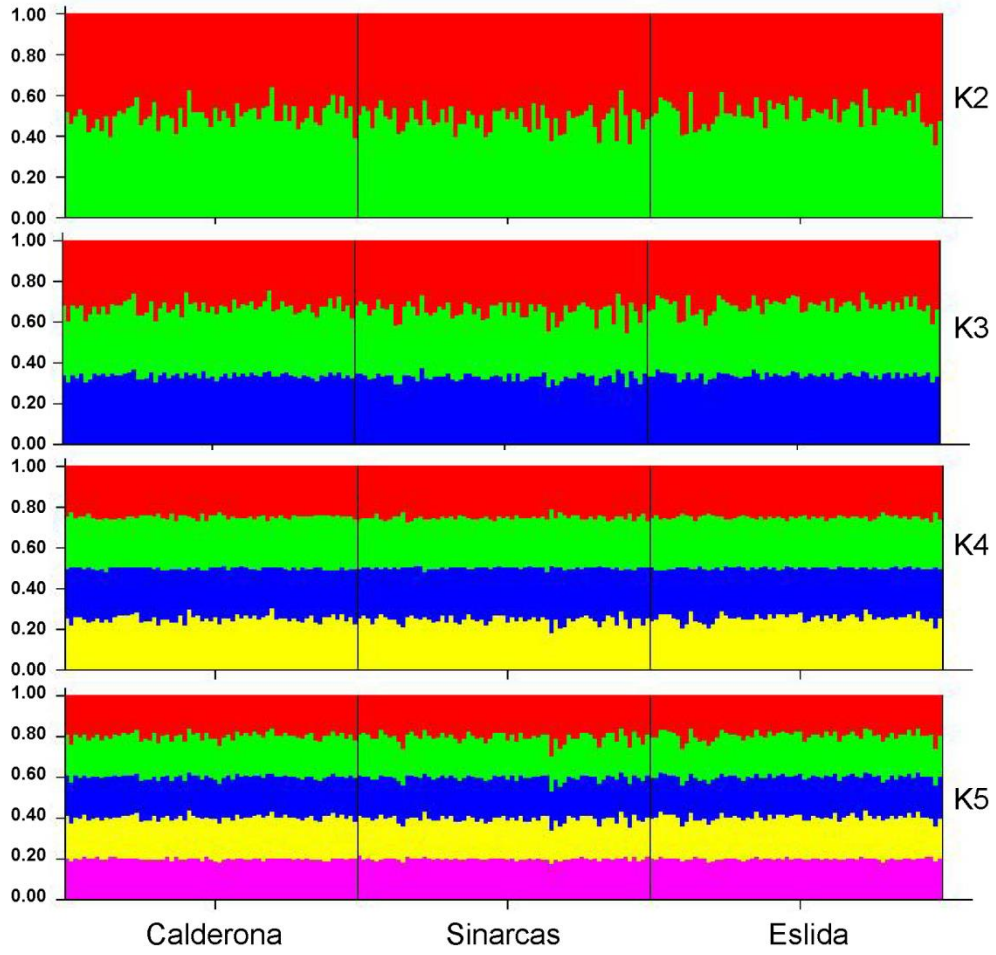

Figure S1 Bar plots showing ancestry proportions for *K*=2 to *K*=5 clusters (i.e., gene pools) as produced by STRUCTURE 2.2 software. Each individual is represented as a line segment which is vertically partitioned into *K* coloured components representing the individual's estimated proportions of ancestry in the *K* clusters. Approximately equal ancestry proportion for each gene pool in all individuals indicates lack of genetic structure.

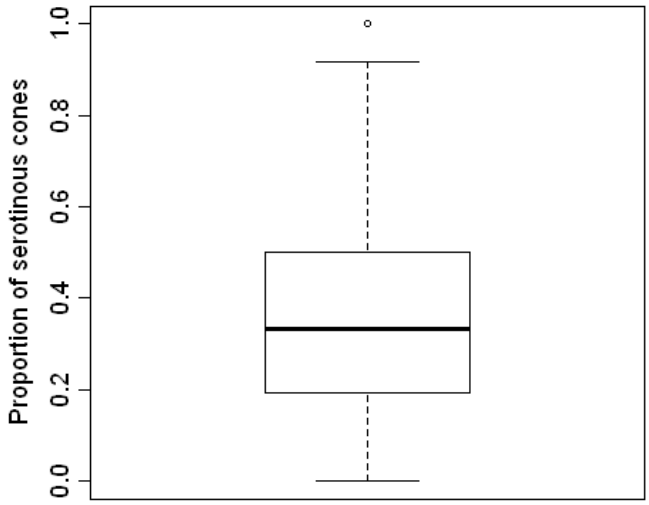

Figure S2 Phenotypic variability (box plot) for serotiny in eastern Iberian Peninsula (mean serotiny of 36.29% and standard deviation of 23.36%) based on 199 trees genotyped with SNP markers.

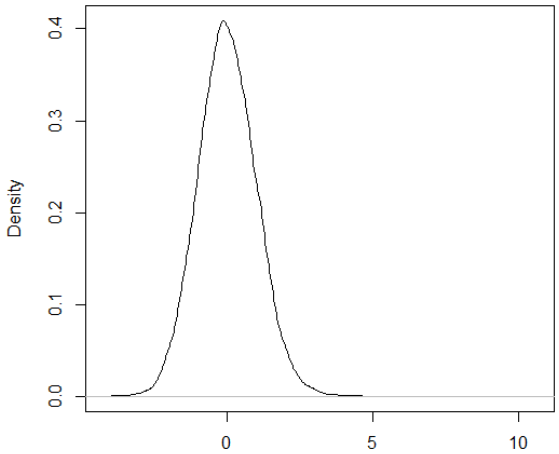

Figure S3 Density plot for pairwise kinship (as estimated by the Loiselle *et al*. [1995] kinship estimator implemented in SPAGeDi 1.3). Notice the positive skewness (D'Agostino's skewness test: skew = 0.208, z = 7.777, *P*-value = 7.387E-15) indicating some relatedness among trees within the studied stands.

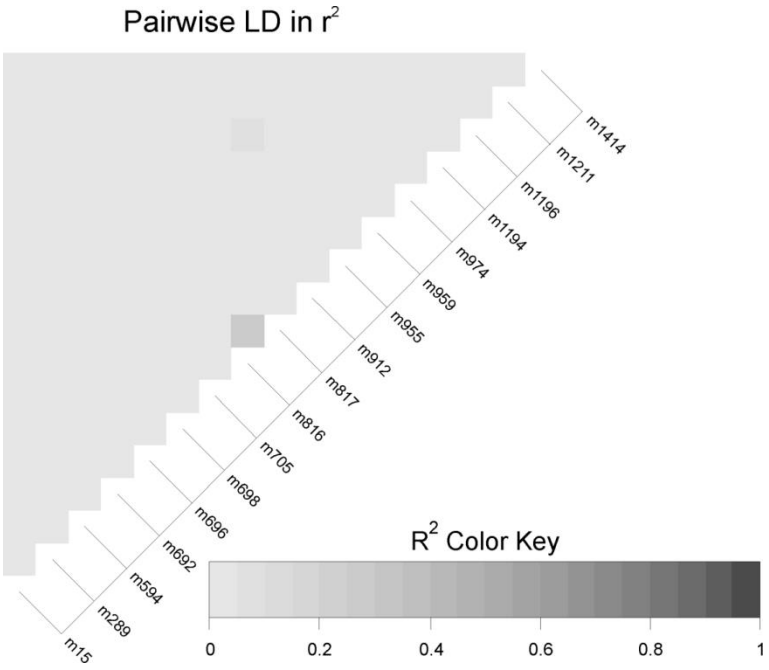

Figure S4 Linkage disequilibrium (LD) heatmap (LDheatmap, R software package) for the 17 SNPs potentially involved in associations with serotiny. Low, but significant, LD was detected between *m816* and *m817* (r^2=0.262), two SNPs from the same gene.

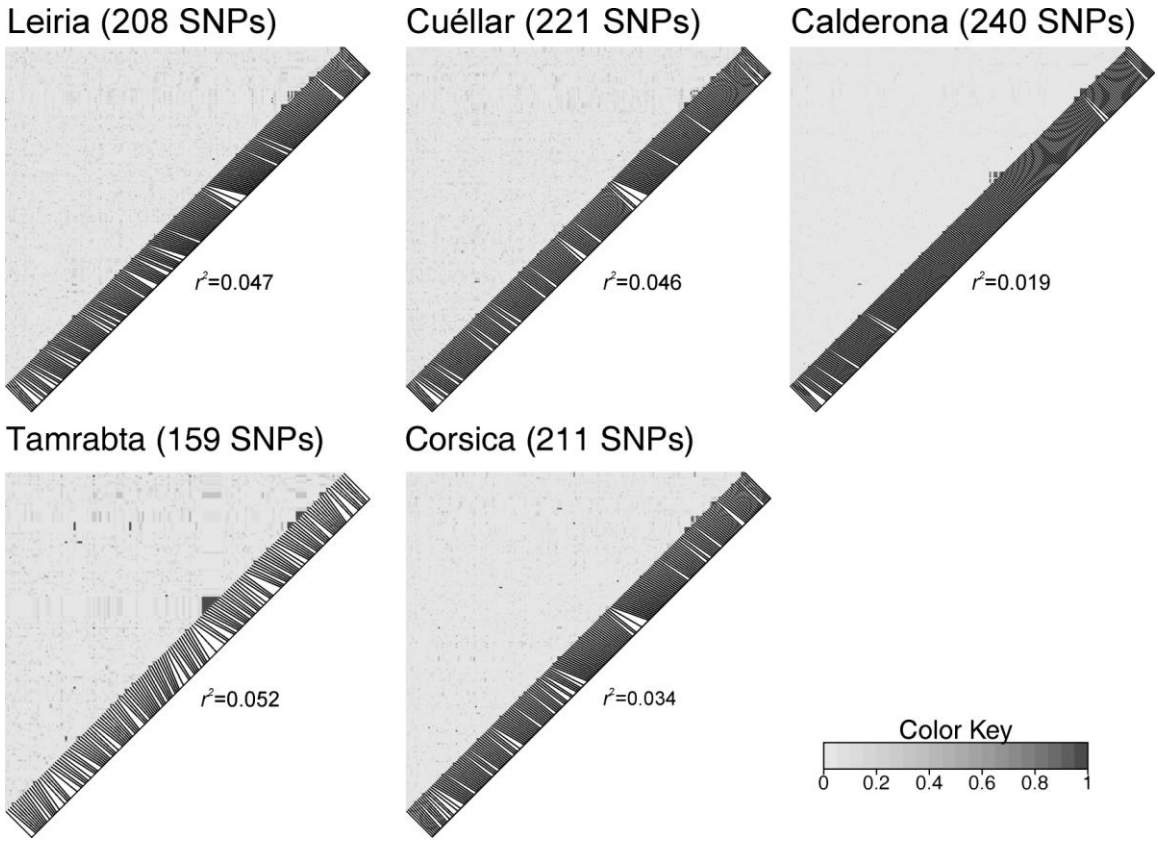

Figure S5 Linkage disequilibrium (LD) heatmaps (LDheatmap, R software package) for all polymorphic SNPs in four range-wide populations (Leiria: Portugal, Cuéllar: central Spain, Tamrabta: Morocco, and Corsica) and one stand (Calderona) from the focal population in the east of the Iberian Peninsula. For each population, the mean r^2 is given.

Acknowledgements

Acknowledgements

This thesis would not have been possible without the help and encouragement of many people to whom I am very grateful. I have learned so much during the last years and I have seen so many places. It was often challenging but it was absolutely a great time!

First of all I would like to thank Santi and Myriam for everything during the last years. I remember our first skype interview. Actually, I could see you but you could not see me, nevertheless you chose me as PhD student☺. Thank you for all your patience and help, I learned so much from you and although it was sometimes difficult to jump between *Pinus* and *Symglo* it was exactly what I wanted to do! But most importantly I had a great time with you in so many different situations: searching for *Symphonia (globulifera)* and other species in Madagascar and Cameroon, sharing a tent with Myriam in the middle of the tropical rain forest, on the top of the Roque de los muchachos in La Palma with both of you, sampling in Calderona with the whole family (including Helena), discussing science in almost every possible situation… Thank you for your advice and for sharing your house and family life with me several times during the last years!

Estoy muy agradecida a mi tutora de la Universidad de Complutense, Pilar Arana, por su ayuda y sus consejos durante los ultimos años.

También al grupo GPE del CIFOR, que ha invertido mucho trabajo y esfuerzo para que esta tesis fuera possible. Muchísimas gracias por todo el trabajo de campo duro y la compañía debajo del sol sin compasión! Especialmente agradecida estoy a Mario y a Fer, sin vosotros todo ésto no hubiera sido posible! Mario, que ha desarrollado mucha arte en el uso de la pértiga y Fer que sube a los árboles como un mono (no te olvides el arnés!). Muchas gracias a Ana de Lucas, Cristóbal y Yusuf por vuestra ayuda en los primeros meses! También a los refuerzos extra que habéis ayudado en algún momento en el trabajo de campo: Edu, Sergio, Diana, Paco, Quique, Natalia, Ana Hernández-Serrano y los chicos del CIDE! También a Jesús, David y Paco por su ayuda con los mapas de SIG.

Además, muchísimas gracias a Ricardo Alía, Delphine Grivet y Juanjo Robledo Arnuncio por las diferentes discusiones científicas, las comidas compartidas y por vuestra ayuda en diversos momentos. Muchas gracias también a Juan Pablo, Mr. Nickname;-) También estoy agradecida a Luis Santos y José Climent por discutir las diferentes adaptaciones de pinos al fuego y sus posibles consecuencias. Además les agradezco a Dori, Toñi, Marisa y Beatriz toda su ayuda con las tareas administrativas en el INIA y con el Ministerio!

En el laboratorio siempre he tenido a Carmen a mi lado y me echaba una mano si hacía falta. Muchas gracias por tu ayuda y tu compañía! Gregor y Ana de Lucas muchas gracias por echar una mano con el genotipado de algunos micros, y a Mario y Zaida por ayudas varias en el laboratorio! También muchas gracias a Luis Díaz, Mariló, María-Ángeles y Nuria por el tiempo compartido en el laboratorio.

Estoy muy agradecida por la colaboración con Miguel Verdú, Juli Pausas y Ana Hernández-Serrano del CIDE. Muchísimas gracias por vuestra ayuda, especialmente por aclararme los conceptos relativos a los regímenes de fuego y las múltiples discusiones científicas.

During the last years I have spent quite some time in other research groups. I am very happy to have had the possibility to interact with so many different people all around the world.

Je voudrais remercier Olivier Hardy pour son aide et son soutien pendant mon séjour à Bruxelles et sur la phylogéographie des espèces tropicales africaines. L'étude de phylogéographie n'etait pas possible sans l'aide de beaucoup de collaborateurs pour échantilloner en Afrique, merci à tous ceux qui ont aidé! Merci beaucoup aussi à tous les gens de l'équipe de recherche de l'Université libre de Bruxelles: Jerôme, Gilles, Guillaume Eben Ezer et Claire. Un merci spécial à Ingrid qui m'a acceuilli dans sa maison et à Alexandra pour sa compagnie et des discussions scientifiques.

Beppe Vendramin, thank you so much for all your help on our papers during the last years and the nice time I had in your lab in Florence, although you forgot to pick me up with you scooter;-). Special thanks also go to Francesca, Sara and Federico who helped me in the lab and in finding my way through the different coffee, pizza, pasta and cookie types in Italy!

Furthermore, I am very grateful to Sally Aitken for the many fruitful discussions and the great time in her group in Vancouver and to Tongli Wang for explaining me RandomForest. Also many thanks to Sam, Katie and Rose for the scientific discussions. Special thanks go to Susannah for her warm welcome in Vancouver and also many thanks to Kim and Jill. The time in Canada would not have been the same without Catherine, Lalu, Noemi and Victor. Thank you so much for your company, the great meals and the beautiful sailing trips.

I am also very grateful to the people from WSL in Switzerland where I spent the last months of my PhD. Felix Gugerli, herzlichen Dank für deine Hilfe bei der Beantragung des Stipendiums für die letzten Monate meiner Doktorarbeit und für deine Hilfe bei der Statistik und zum Konzept unseres Papers. Des Weiteren danke ich Christian Rellstab, Rolf Holderegger, Corine Schöbel und Pietro Milanesi für die verschiedenen wissenschaftlichen Diskussionen. The time in Switzerland was very special for me because of the company of Maje, Oli, Thomas, Andrea, Isa, Luis, Aitana, Sanjeev, Shiva, Jyoti, Santi, Myriam, Helena and Flori, thank you!

Concetta Burgarella y Miguel de Navascués muchísimas gracias por explicarme el mundo de los "skyline plots" y los métodos de ABC! Me lo he pasado genial en Montpellier con vosotros! Tener el despacho cerca de la playa no está nada mal;)

Además, estoy muy agradecida a Antón por ayudarme a entender el procedimiento para entregar la tesis. También, muchas gracias a Alex por compartir y resolver dudas juntos en las últimas semanas y por la compañía durante los últimos años.

Außerdem möchte ich Birgit Ziegenhagen danken! In deiner Arbeitsgruppe habe ich angefangen mit der wissenschaftlichen Arbeit und bin sehr froh, dass du mir weiterhin mit Rat und Tat (wenn auch meist aus der Ferne) zur Seite stehst. Auch Sascha möchte ich Danken für die interessanten Diskussionen über Koniferen-Transkriptomik und Ilona für ihre Hilfe bei der multivariaten Statistik. Katrin, herzlichen Dank für unsere Zusammenarbeit am *Symphonia* paper und ich freue mich auf weitere Kollaborationen.

Muchímas gracias a todos los compañeros del despacho de becarios del INIA (nuestro hábitat común☺) por todos los eventos y momentos que hemos vivido juntos! Gracias a vosotros, Gregor (unser kulinarischer Spezialist, der immer an alle denkt!), Marina (mi compi de patinaje y de la bicicrítica), Natalia (que me llevó al roco y me acogió en su casa), Isa y Luis (que me acogieron en Suiza y que siempre organizan planes), Maje (mi compi del café a las 15h en Birmensdorf), Quique (que siempre nos hace reir), Paloma Ruiz (que siempre está dispuesta a ayudar) y tambien a Laura Hernández (que me llevó a la montaña y a ver grullas, avutardas, flamencos..:), Laura Fernández (compi de la bici y de ir al cine), Javi (mi compi del camino en bici y con Tamara mis vecinos preferidos), Mario (mi link especial entre amigos;), mi tiempo en Madrid ha sido fenomenal! Además, muchísimas gracias también a David Macaya que me ha ayudado con la administración de las FPIs para pedir estancias. También muchas gracias a los compañeros nuevos Paloma, Rose y Ruth y los antiguos, Raúl, Javi, José Alberto e Iva por el tiempo y las discusiones compartidas, además a los que han pasado un tiempo con nosotros, especialmente a Quique Hernández y a Carlos por recorrer Madrid conmigo. También muchas gracias a Salva mi compi de la carrera de orientación.

A parte del INIA, mi tiempo en Madrid ha sido inolvidable particularmente gracias a Mati, con quien he compartido muchísimos momentos especiales durante los casi 4 años de convivencia, te echo mucho de menos! Muchísimas gracias también a Ángela y a Paloma Gallego por nuestra convivencia y a Paloma Gallego por involucrarme más en la cultura española;) También estoy muy agradecida a Lucie, por su compañía y por aguantarme en las últimas semanas de la recta final de la tesis y por revisar partes de la tesis. También las “Voces de ida y vuelta” han sido muy importantes para mí, muchas gracias a tod@s por la música y los momentos compartidos durante los últimos años!

Eike, vielen Dank für deine Freundschaft und dass wir es schaffen auch in der Ferne immer am Leben der anderen teilzuhaben! Du bist mein Anker, wenn gerade mal wieder alles dabei ist davon zu schwimmen!

Ganz besonders möchte ich mich bei meiner Familie, meinen Eltern und meinem Bruder, bedanken. Meine Eltern haben mir die Überzeugung mitgegeben, dass man fast alles erreichen kann, was einem wichtig ist. Zumindest sollte man es versuchen! Aber man braucht natürlich Hilfe und deswegen danke ich euch ganz herzlich für eure Unterstützung, obwohl ich mal wieder in weiter Ferne war.