UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA,
ALIMENTARIA Y DE BIOSISTEMAS

*Ecological and association genetics
in two Mediterranean pine species*

TESIS DOCTORAL

Isabel Rodríguez Quilón

Ingeniera Agrónoma

2017

CAMPUS
DE EXCELENCIA
INTERNACIONAL

Programa Oficial de Doctorado en Biotecnología y Recursos Genéticos de Plantas y Microorganismos Asociados.

DEPARTAMENTO DE BIOTECNOLOGÍA-BIOLOGÍA VEGETAL

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

UNIVERSIDAD POLITÉCNICA DE MADRID

Tesis doctoral:

Ecological and association genetics in two Mediterranean pine species

Autora:

Isabel Rodríguez Quilón

Ingeniera Agrónoma

Directores:

Ricardo Alía Miranda

Doctor Ingeniero de Montes

Santiago César González Martínez

Doctor Ingeniero de Montes

Año: 2017

UNIVERSIDAD POLITÉCNICA DE MADRID

Tribunal nombrado por el Magnífico y Excelentísimo Sr. Rector de la
Universidad Politécnica de Madrid, el día _____

Presidente:

Secretario:

Vocal:

Vocal:

Vocal:

Suplente:

Suplente:

A Luis y Sofía

*Nothing in life is to be feared,
it is only to be understood.
Now is the time to understand more,
so that we may fear less.*

Marie Curie

INDEX

RESUMEN.....	5
ABSTRACT.....	9
INTRODUCTION	13
Background.....	13
Case studies in Mediterranean forest trees: <i>P. pinaster</i> and <i>P. halepensis</i>	15
Genetic variation at the base of adaptation	18
Studying the relationship between genotype and phenotype	22
Polygenic adaptation	27
Environmental association: ‘reverse ecology’	28
Applications for conservation.....	30
Thesis overview and organization	33
OBJECTIVES	35
MATERIALS AND METHODS	37
Common garden experiments	37
Studied traits	41
Trait measurements	45
CHAPTER 1: HFCs IN MARITIME PINE.....	49
INTRODUCTION	50
MATERIAL AND METHODS.....	52
Study species.....	52
Plant material	53
Molecular markers.....	54
Fitness and heterozygosity estimates.....	54
Heterozygosity-Fitness Correlations (HFCs).....	56
RESULTS.....	60
Fitness and heterozygosity estimates.....	60
Heterozygosity-Fitness Correlations (HFCs).....	60

DISCUSSION	62
Lack of general fitness effects	63
Significant local fitness effects	64
CHAPTER 2: GENETICS OF INTEGRATED TRAITS IN ALEPPO PINE.....	69
INTRODUCTION	70
MATERIALS AND METHODS	73
Plant material and phenotypes	73
Quantitative genetics	75
Trait trade-offs	76
Trait-environment correlations	76
Molecular markers and population structure	77
Association tests	77
RESULTS	78
Quantitative genetics	78
Trait trade-offs	80
Trait-environment correlations	83
Molecular markers and population structure	84
Association tests	85
DISCUSSION	90
Phenotypic variation, environmental correlations and trade-offs	90
Association genetics	92
CHAPTER 3: POLYGENIC ADAPTATION IN MARITIME PINE	99
INTRODUCTION	100
MATERIALS AND METHODS	103
Plant material and phenotyping	103
Genotyping	104
Phenotypic trait analysis	104
Q_{ST} - F_{ST} comparisons	105
Identification of genotype-phenotype associations (GWAS)	105

Polygenic adaptation test	106
RESULTS	107
Phenotypic variation	107
Q_{ST} - F_{ST} comparisons	107
Association genetics and marker effects	109
Polygenic adaptation test	111
DISCUSSION	115
CHAPTER 4: MARITIME PINE CONSERVATION GENETICS	121
INTRODUCTION	122
MATERIALS AND METHODS	124
Plant material	124
Molecular markers	126
Quantitative traits	126
Identification of conservation groups	128
Characterization of conservation groups	129
Climatic zonation	129
RESULTS	130
Molecular markers	130
Quantitative traits	131
Identification and characterization of conservation groups	132
Usefulness of climatic zonation	136
DISCUSSION	136
Intraspecific conservation in forest trees	137
Usefulness of molecular markers and quantitative traits in conservation	137
Implications for maritime pine conservation	139
DISCUSSION CONCLUDING REMARKS	141
CONCLUSIONS	145
REFERENCES	147

SUPPLEMENTARY INFORMATION	169
Supplementary Information Chapter 1	169
Supplementary Information Chapter 2	175
Supplementary Information Chapter 3	181
Supplementary Information Chapter 4	185
ANNEX 1	197
ANNEX 2.....	213
AKNOWLEDGEMENTS.....	247

RESUMEN

Los bosques poseen un gran valor ecológico, económico y social, además de jugar un papel estratégico en el mantenimiento de la biodiversidad y como sumideros de dióxido de carbono. Sin embargo, durante las últimas décadas, el cambio climático está modificando la distribución geográfica de los bosques naturales y el funcionamiento de estos ecosistemas. Los bosques mediterráneos son especialmente sensibles al cambio climático, ya que se espera que su efecto sea más notable en regiones secas y templadas. En estos lugares muchas especies de árboles se encuentran además en su límite adaptativo. Como respuesta, los árboles tienen tres opciones: adaptarse, migrar o extinguirse. La capacidad de hacer frente a condiciones ambientales futuras en los rangos de distribución actuales podría lograrse mediante procesos de adaptación o a través de respuestas plásticas. La adaptación local, por su parte, implica una respuesta genética a las presiones de selección, mientras que la plasticidad podría permitir la aparición de fenotipos adaptados sin que exista un cambio genético subyacente. En general, las poblaciones de árboles forestales son genéticamente muy diversas, y se espera que la adaptación local surja de esta variabilidad genética dentro y entre poblaciones. Es fundamental comprender los procesos adaptativos actuales y pasados, y conocer los agentes ecológicos que han dado forma a la diversidad genética dentro de cada especie, para poder predecir el futuro de nuestros bosques e implementar estrategias efectivas de conservación. Sin embargo, seguimos sin conocer la base genética de la adaptación local, a pesar de su vital importancia en la persistencia de las especies y sus posibles cambios de distribución ante el cambio climático.

Esta tesis se enmarca dentro del campo de la genética ecológica. Está estructurada en cuatro capítulos, que corresponden a artículos científicos ya publicados o manuscritos en preparación. Incluye también tres anexos, con información adicional y artículos complementarios ya publicados. El objetivo principal de la tesis es evaluar los procesos evolutivos y ecológicos, pasados y actuales, que han dado lugar a la adaptación local de los árboles forestales. Se pretende desentrañar la compleja interacción entre genotipo, fenotipo y ambiente, avanzando en el conocimiento de las bases genéticas de los caracteres adaptativos. Estas preguntas de investigación se relacionan a continuación con cuestiones prácticas para la conservación dinámica de los recursos genéticos forestales.

Pinus pinaster Ait. (pino rodeno o negral) y *P. halepensis* Mill. (pino carrasco) son dos especies alógamas de pinos mediterráneos de gran importancia ecológica y económica, con diferentes historias demográficas y niveles de diversidad genética intraespecífica. Ofrecen así la oportunidad de integrar información respecto a la diferenciación poblacional, su

arquitectura genética, plasticidad y tendencias ecotípicas, proporcionando numerosas ventajas para el estudio de la adaptación local. Para ello se han utilizado datos genéticos de microsatélites nucleares (12 nuSSRs) y polimorfismos de un único nucleótido (384 y 6.100 SNPs), y datos fenotípicos, medidos en ensayos de ambiente común replicados en múltiples condiciones ambientales. Los datos procedentes de los ensayos de ambiente común permitieron estimar en qué medida la expresión fenotípica tiene base genética. Su replicación bajo ambientes contrastantes ayudó a estimar si esta expresión también está condicionada por una respuesta plástica. La genética cuantitativa, la genética de poblaciones y los estudios de asociación genotipo-fenotipo se utilizan de forma complementaria en los cuatro capítulos de la tesis para avanzar en la comprensión de la base genética de los caracteres adaptativos en los pinos rodeno y carrasco.

En general, a lo largo de los diferentes estudios incluidos en esta tesis, se encontró evidencia de variación genética intraespecífica y de procesos de adaptación local en ambas especies. Los rasgos adaptativos mostraron un control genético moderado y estructurado entre poblaciones. La diferenciación genética cuantitativa entre poblaciones fue consistentemente superior a la diferenciación en marcadores moleculares para todos los caracteres y ambientes de ensayo, a diferentes niveles jerárquicos, sugiriendo la existencia de adaptación local (*Capítulos 3 y 4*). También se identificaron *loci* presumiblemente adaptativos, subyacentes a importantes caracteres de aptitud biológica (*Capítulos 2 y 3*). Los factores ambientales influyeron en la variabilidad genética y en la arquitectura genética detectada de los caracteres adaptativos estudiados (*Capítulos 2, 3 y 4*).

En lo que respecta al estudio de la relación entre genotipo y fenotipo, en primer lugar, se abordó una cuestión evolutiva clásica: la correlación entre la heterocigosidad y la aptitud biológica (*Capítulo 1*). El objetivo era averiguar si episodios de endogamia, debidos a un aumento del apareamiento entre familiares o a la deriva genética después de un cuello de botella, conducen a una disminución de la aptitud biológica en el pino rodeno. Se vio que la heterocigosidad genómica tiene poco impacto en la aptitud biológica del pino marítimo, aunque la variación en la endogamia parece estar asociada con la supervivencia. Esto podría explicarse por el sistema reproductivo de la especie y su historia evolutiva. Sin embargo, la heterocigosidad en marcadores presumiblemente adaptativos parece jugar un papel crítico bajo fuertes presiones selectivas, como las que podrían derivarse del cambio climático.

Los estudios de asociación genotipo-fenotipo permitieron identificar *loci* concretos subyacentes a diferentes caracteres adaptativos en ambas especies de pino (*Capítulos 2 y 3*). Para el pino carrasco se encontraron 15 SNPs significativamente asociados con caracteres de crecimiento, extractivos de madera, contenido en lignina y floración,

explicando una variación moderada a nivel fenotípico (<11%). Se demuestra que la inclusión de fenotipos integrados en las pruebas de asociación genotipo-fenotipo puede aumentar ligeramente la proporción de la varianza fenotípica explicada, así como posibilitar la detección de posibles efectos pleiotrópicos (*Capítulo 2*). En el pino rodeno, los estudios de asociación realizados, con un mayor número de marcadores moleculares (6.100 SNPs) y en ambientes contrastantes, permitieron identificar seis SNPs asociados con la supervivencia en ensayos bajo clima atlántico suave, y dos SNPs asociados con la altura en un ensayo situado en un clima mediterráneo más severo (*Capítulo 3*). En ambos casos, serían necesarios más estudios para confirmar que los *loci* candidatos identificados son realmente genes adaptativos funcionales. Teniendo en cuenta que la proporción de varianza fenotípica explicada fue baja (<10%) a pesar del aumento del número de marcadores moleculares utilizados para el pino marítimo, y dado que se espera que la mayoría de los caracteres adaptativos estén controlados por un gran número de *loci* con efecto pequeño, avanzamos hacia un enfoque de adaptación poligénica. Con ese objetivo, se combinaron la genética de poblaciones y los estudios de asociación, testando la existencia de un exceso de varianza en la diferenciación entre poblaciones, así como la presencia de correlaciones excesivas con variables ambientales. Para ello, se compararon las frecuencias alélicas de un conjunto de *loci* adaptativos candidatos frente a un modelo nulo de deriva genética neutra e historia evolutiva compartida de la población. Se encontró un exceso de varianza entre poblaciones para altura y supervivencia, en consonancia con la existencia de selección direccional, principalmente impulsada por la covarianza entre *loci*, es decir, por desequilibrio de ligamiento. También se encontraron señales de correlaciones excesivas con temperaturas mínimas y precipitación media, respectivamente, para altura y supervivencia, lo que confirma el papel clave de tales variables ambientales en los procesos de adaptación local en la especie. Los resultados apuntan a muchos *loci* de pequeño efecto subyacentes a los caracteres estudiados, estando la adaptación local impulsada por pequeños cambios de frecuencias alélicas (*Capítulo 3*).

En el pino carrasco también se testó la relación entre variables ecogeográficas y diferentes fenotipos adaptativos -incluyendo crecimiento, reproducción, resistencia a la sequía y caracteres de calidad de madera-, así como correlaciones entre caracteres. Los análisis se realizaron carácter por carácter y también para fenotipos integrados. En general, se encontró evidencia de adaptación local intraespecífica. También se comprobó que las relaciones balanceadas en cuanto a asignación a caracteres adaptativos relacionados con crecimiento, reproducción y defensa coinciden con las predicciones de la teoría de la asignación y las teorías de caracteres de historia vital en la especie (*Capítulo 2*). En pino rodeno se corroboró la existencia de variación clinal a lo largo de gradientes de temperatura

y precipitación, y el importante papel del ambiente en la adaptación local de la especie. Una diferenciación genética cuantitativa entre poblaciones consistentemente mayor que la diferenciación molecular neutra, para altura y supervivencia bajo ambientes contrastantes, apuntó también a una selección divergente generalizada. Se encontraron altos niveles de variación genética y plasticidad fenotípica a diferentes escalas espaciales, lo que sugiere la existencia de potencial de cambio evolutivo para hacer frente a nuevas presiones selectivas, aunque variable dentro del rango de distribución de la especie (*Capítulo 4*).

Por último, se demuestra que los conocimientos adquiridos sobre variación genética de caracteres adaptativos pueden utilizarse en programas de conservación genética dinámica de árboles forestales (*Capítulo 4*). Los componentes neutros y adaptativos de la diversidad genética, así como los factores ambientales, deben tenerse en cuenta en la identificación de unidades de conservación a nivel subespecífico, especialmente en especies con una fuerte estructura poblacional y complejas historias evolutivas. El enfoque de zonación ambiental utilizado actualmente por la estrategia paneuropea de conservación genética para los árboles forestales se mejoraría con la integración gradual de información de caracteres moleculares y cuantitativos a medida que estos datos vayan estando disponibles.

ABSTRACT

Forests have great ecological, economic and social value, playing also a critical role both as carbon pools and carbon dioxide sinks. Nevertheless, over the last decades, climate change has modified natural forest geographic ranges and ecosystem functioning. Mediterranean forests are especially sensitive to climate change as its consequences are expected to be experienced more abruptly in dry and cool regions, where, in addition, many tree species are found at their adaptive limit. In response, forest trees can adapt, migrate or go extinct. Coping with future environmental conditions in current distributions could be achieved by adaptation processes or via plastic responses. Local adaptation implies a genetic response to selection pressures, while plasticity could enable adapted phenotypes without underlying genetic change. Overall, forest tree populations are genetically very diverse, and local adaptation is expected to rely on such available genetic variability within and among populations. Understanding past and contemporary adaptive process and ecological drivers shaping genetic diversity within species is of key importance to predict the fate of forests and implement effective conservation strategies under the current climate change scenario. However, despite its utmost importance in species persistence and range shifts facing climate change, the genetic basis of local adaptation remains poorly understood.

This thesis is framed within the field of ecological genetics, and structured in four chapters, corresponding to already published papers or manuscripts in preparation, and two annexes, with complementary information and also already published studies. The main objective of the thesis was to evaluate past and ongoing evolutionary and ecological processes leading to local adaptation in forest trees. I aimed to disentangle the complex interaction between genotypes, phenotypes and environmental conditions, advancing in the knowledge of the genetic basis of adaptive traits. These research questions were then related to practical issues for dynamic conservation of forest genetic resources.

Pinus pinaster Ait. (maritime pine) and *P. halepensis* Mill. (Aleppo pine) are two ecologically and economically important outcrossing Mediterranean pine species, with different demographic histories and levels of intraspecific genetic diversity. They provide the opportunity to integrate information on population differentiation, genetic architecture, plasticity and ecotypic trends, offering numerous advantages for the study of local adaptation. To do so, here I used genetic data from nuclear microsatellites (12 nuSSRs) and single nucleotide polymorphisms (384 and 6,100 SNPs), and phenotypic data, measured in common gardens replicated in multiple environments. Common garden data allowed to estimate to which extent phenotypic expression is genetically based, and trial replicates

under contrasting environments to know whether this expression can also be plastic. Quantitative genetics, population genetics and genotype-phenotype association methods are complementarily used in the four chapters to advance in the understanding of the genetic basis of adaptive traits in maritime and Aleppo pines.

Overall, throughout the different studies included in this thesis, I found evidence of intraspecific genetic variation and local adaptation processes in both species. Adaptive traits exhibited a moderate genetic control and were structured across populations. Among-population quantitative genetic differentiation was consistently above among-population molecular marker differentiation across traits and testing environments, at different hierarchical levels, suggesting local adaptation (*Chapters 3 and 4*). I was also able to identify putatively adaptive *loci*, underlying important fitness traits (*Chapters 2 and 3*). Environmental factors influenced genetic variability and the detected genetic architecture of the studied adaptive traits (*Chapters 2, 3 and 4*).

Particularly addressing the relationship between genotype and phenotype, I first approached a traditional evolutionary question: the correlation between heterozygosity and fitness (*Chapter 1*). I aimed to find out whether inbreeding events, due to increased mating between relatives or genetic drift after a bottleneck, lead to decreased fitness in maritime pine. Genome-wide heterozygosity has little impact on maritime pine fitness, even though variation in inbreeding seems to be associated with survival. This could be explained by maritime pine mating system and evolutionary history. However, heterozygosity at putatively adaptive markers may play a critical role under increased selective pressures, such as the ones that could be derived from climate change.

Genotype-phenotype association studies allowed the identification of particular *loci* underlying different adaptive traits in Aleppo and maritime pines (*Chapters 2 and 3*). For Aleppo pine, I found 15 SNPs significantly associated with growth traits, wood extractives, lignin content or flowering, explaining a moderate variation in phenotypic traits (<11%). I demonstrate that integrated phenotypes in genotype-phenotype association tests can slightly enhance the proportion of phenotypic variance explained, as well as give the chance of detecting possible pleiotropic effects (*Chapter 2*). For maritime pine, association tests performed with an increased number of molecular markers (6,100 SNPs) at contrasting environments, allowed the identification of six SNPs associated with survival in trials under mild Atlantic climate, and two SNPs associated with tree height in one of the trials under a harsher Mediterranean climate (*Chapter 3*). In both cases, future studies would be needed to confirm the identified candidates as functional adaptive genes. As the proportion of phenotypic variance explained remained low (<10%), despite the increased number of molecular markers used for maritime pine, and provided that most adaptive traits are

expected to be controlled by a large number of *loci* with small effect, we moved forward towards a polygenic adaptation approach. For this, I combined population and association genetics, and tested for excess variance in among-population differentiation and signals of excessive correlations with environmental variables, when allele frequencies of a set of candidate adaptive *loci* were compared to a null model of neutral genetic drift and shared population history. I found an excess of variance among populations for height and survival, consistent with directional selection, mainly driven by covariance among *loci* (i.e. linkage disequilibrium). I also found signals of excessive correlations with minimum temperatures and mean precipitation, respectively, for height and survival traits, confirming the key role of such environmental variables in local adaptation processes in the species. Our results point to many *loci* of small effect underlying the studied traits, with local adaptation driven by small allele frequency changes (*Chapter 3*).

In Aleppo pine, I also tested the relationship between ecogeographical variables and different adaptive phenotypes -including growth, reproduction, drought resistance and wood quality traits-, as well as among-trait correlations. Analyses were done trait-by-trait and also for integrated phenotypes. Overall, I found supporting evidence for intraspecific local adaptation and that allocation to adaptive traits related to growth, reproduction and defence trade-off against each other, matching theoretical predictions from allocation and life-history theories in the species (*Chapter 2*). In maritime pine, I corroborated the existence of clinal variation along temperature and rainfall gradients in the species, and the important role of the environment in maritime pine local adaptation. Consistently higher among-population quantitative genetic differentiation than neutral molecular differentiation, for height and survival under contrasting environments, point to widespread divergent selection too. I found high levels of genetic variation and phenotypic plasticity at different spatial levels, suggesting potential for evolutionary change to cope with new selective pressures, although variable within the species range (*Chapter 4*).

Finally, I show that the acquired knowledge about the genetic variation of adaptive traits can be used to improve dynamic genetic conservation programs for forest trees (*Chapter 4*). Neutral and adaptive components of genetic diversity, as well as environmental factors, should be accounted for in the identification of conservation units below the species level, especially in species with strong population structure and complex evolutionary histories. The environmental zonation approach currently used by the pan-European genetic conservation strategy for forest trees would be improved by gradually integrating molecular and quantitative trait information, as data become available.

INTRODUCTION

Background

The evidence for human-driven global climate change seems overwhelming. Climate change is increasingly recognized as one of the most important challenges faced globally by ecosystems and societies (Thuiller 2007). In this framework, the European Union is committed to a progressive decarbonisation of the world economy by the end of the century, focusing on mitigation through global agreements to reduce greenhouse gases, as the one achieved in the COP21 conference (Paris, December 2015). Forests play a key role, both as carbon pools and carbon dioxide sinks. Today, forests cover nearly 40% of the European land surface (Figure 1) and host much of the continent's biodiversity (FAO 2015). They have great ecological, economic and social value, playing also a critical role in mitigation, wood production and the provision of many ecosystem services. Unfortunately, forest ecosystems seem to be particularly susceptible to climate change, as the long life-span of trees can limit rapid adaptation to environmental changes (Alley 2003; Lindner *et al.* 2010; Milad *et al.* 2011). Nonetheless, over the last decades, climate change has produced numerous shifts in species' distributions and abundance (Thomas *et al.* 2004), modifying the geographic ranges of natural forests, ecosystem functioning and interactions between species. Nowadays European forest ecosystems face multiple natural and anthropogenic threats. Forecasted future increasing temperatures combined with varying degrees of increasing or decreasing precipitation add to current habitat degradation and fragmentation, biotic exchange, poor silvicultural practices and use of low quality or poorly adapted forest reproductive material (Leone and Lovreglio 2004; Koskela *et al.* 2007). Moreover, Mediterranean forests are generally expected to experience the consequences of climate change more abruptly, as climate change impacts are expected to be more severe in dry, high-temperature regions, where in addition many tree species are at their adaptive limit, in the rear edge of the distribution (Hampe and Petit 2005; Lindner *et al.* 2010; Fady *et al.* 2015).

Figure 1. Surface covered by forests in Europe in 2011. Information based on remote sensing technologies and forest inventory statistics. Source: European Forest Institute (<http://www.efi.int/>).

In response to rapidly changing environments, tree populations can persist, either through migration to track spatially new ecological niches or through adaptation to new conditions in current locations, or become extinct (Aitken *et al.* 2008). Although forest distribution ranges have been and continue to be highly dynamic, current estimates of forest tree migrations are far below what would be necessary for species to track future climate warming (Petit *et al.* 2008; Corlett and Westcott 2013). Fast adaptation driven by new selection pressures may also be hindered by tree long-generation times (Lindner *et al.* 2010; Milad *et al.* 2011). Phenotypic plasticity also plays a critical role in the response of plant populations to climate change, enabling adapted phenotypes without genetic change (Matesanz *et al.* 2010; Chevin *et al.* 2012; Valladares *et al.* 2014). Provided that trees are nowadays adapted to the current climate, the study of past forest changes and adaptive processes can provide the necessary knowledge to predict future responses of trees to human-induced climate change (Savolainen *et al.* 2007; Petit *et al.* 2008; Lindner *et al.* 2010; Alberto *et al.* 2013). Disentangling the mechanisms of past adaptation events in forest trees becomes of paramount importance for future forest preservation.

A population is considered locally adapted (*sensu* Kawecki and Ebert 2004) when it has the highest fitness at its home site compared to other non-local populations. Local adaptation is

believed to play a key role in long-lived forest trees because they usually inhabit large and environmentally heterogeneous ranges and host high levels of genetic variation (Hamrick *et al.* 1992; Petit and Hampe 2006). Although it has been recently suggested that local adaptation may be less common in plant populations than generally assumed, and independent of life history, spatial or temporal habitat heterogeneity and demographic scale (Leimu and Fischer 2008), local adaptation is more common and widespread in trees than in other plant species (see Alberto *et al.* 2013 for a review). A long history of common garden experiments has revealed high among-population levels of genetic variation for adaptive quantitative traits, geographic structuring of that variation along climatic gradients, and genotype-by-environment interaction, providing strong evidence of local adaptation of tree populations in response to climate (Howe and Aitken 2003; Savolainen *et al.* 2007). These patterns of local adaptation likely occurred together with forest range expansions after the last glacial maxima (Bucci *et al.* 2007; Petit *et al.* 2008). The challenge now is greater, as human driven climate change will impose stronger and faster selection pressures to tree populations (Davis and Shaw 2001).

Case studies in Mediterranean forest trees: *P. pinaster* and *P. halepensis*

Besides the fundamental role of forest trees in the global impact of climate change, their rich background in quantitative genetics and recent growing genomic resources make them particularly appealing models for adaptation, evolution and climate change research (Alberto *et al.* 2013). Among forest tree species, conifers, and particularly pines, exhibit additional advantages, such as low levels of domestication, high levels of genetic variation, large natural populations, outcrossing mating systems and already established common gardens for accurate phenotype evaluation (Neale and Savolainen 2004). The diverse genus *Pinus* comprises more than 100 species, widely distributed across many forest types in the Northern hemisphere (Keeley 2012). *Pinus pinaster* Ait. (maritime pine) and *Pinus halepensis* Mill. (Aleppo pine) are two Mediterranean pine species belonging to section *Pinus*, subsection *Pinaster*. Both are monoecious outcrossing long-lived trees, with great ecological and economical importance.

P. pinaster has a fragmented distribution in the western Mediterranean region (Figure 2), covering about 4 million hectares distributed across Southern France, Italy, Northern Africa and the Iberian Peninsula, where glacial refugia have been suggested (Bucci *et al.* 2007; Lucas and Martínez 2009). It has a very broad ecological niche, spanning from Atlantic to continental and semi-arid climates, ranging from sea level to 2000 m a.s.l. elevations, in both calcareous and siliceous soils (Alía and Martín 2003). The species shows high intraspecific variability for life-history traits and high levels of genetic diversity (Tapias *et al.* 2004; Soto *et*

al. 2010). It has a thick bark, relatively large cones in groups of two or three, presence of serotinous cones and long needles (Alía *et al.* 1996; Alía and Martín 2003) (Figure 3). Maritime pine is the main tree species used for reforestation in south-western Europe, being used to secure coastal and inland sand dune areas and to create new forests for resin, bio-industry compounds and timber production (Bouffier *et al.* 2013).

Figure 2. *P. pinaster* and *P. halepensis* distribution maps. Blue shaded areas correspond to the species natural distribution. Source: www.euforgen.org.

Figure 3. *P. pinaster* morphology of needles and cone (© Ruíz De la Torre, 1993) and trees from a natural stand (© tabuyodelmonte.wordpress.com).

P. halepensis has a broader geographical distribution, from Syria and Turkey in the East to the Eastern Iberian Peninsula in the West (Figure 2). Present-day distribution is thought to be derived from a range expansion from Eastern Mediterranean populations westwards all around the Mediterranean basin (Grivet *et al.* 2009). The species grows on all substrates and at variable altitudes (0-2600 m a.s.l.), living in contrasting environments including almost all bioclimates of the Mediterranean region (Fady *et al.* 2003). Its ecological niche, though also broad, is more restricted to drier and warmer areas, with a greater tolerance to drought stress than *P. pinaster*. Various genetic studies have shown low levels of genetic diversity both in neutral molecular markers and in quantitative traits in different parts of its distribution (Bucci *et al.* 1998; Tapias *et al.* 2001; Climent *et al.* 2008; Grivet *et al.* 2009). Compared to maritime pine, Aleppo pine has smaller, single, pedunculated cones and shorter, fine, flexible, light green needles (Fady *et al.* 2003) (Figure 4). It has great ecological importance in Mediterranean dry ecosystems and is planted for restoration, as an ornamental tree, and occasionally for timber production (Chambel *et al.* 2013).

Figure 4. *P. halepensis* morphology of needles and cone (left, © Ruíz De la Torre, 1993) and tree picture (right).

Genetic variation at the base of adaptation

Understanding the genetic basis of local adaptation is increasingly important and one of the major goals in evolutionary biology, as it plays a key role in species persistence and range shifts facing climate change (Pritchard and Di Rienzo 2010; Savolainen *et al.* 2013). Despite recent theoretical advances, genomic tools and field experiments, the genetic basis of local adaptation remains relatively unknown (Savolainen *et al.* 2013). What is known since long ago is that for a trait to evolve in response to selection, first there must be genetic variance and an underlying heritable genetic basis upon which to select. Then, the selective agent (e.g. new environmental pressures) must be able to efficiently select genetic variants (i.e. alleles) that increase fitness, based on phenotypes (Lynch and Walsh 1998; Le Corre and Kremer 2003).

Overall, forest tree populations are genetically very diverse, conferring them a great potential for adaptation and resilience (Fady *et al.* 2015). Given their long lifespan, local adaptation is expected to largely rely on the available genetic variability within and among tree populations, rather than on new variability generated by mutations. Phenotypic and genetic differentiation of tree populations along environmental gradients or contrasting habitats has been usually interpreted as indicative of local adaptation. However, this should be taken with caution, as local adaptation can be confounded with neutral genetic effects or with non-genetic effects such as plasticity, maternal effects or temporally varying selection (reviewed in Savolainen *et al.* 2013).

Genetic variation can be studied at the phenotypic or molecular marker level (Figure 5). Quantitative genetics focus on the phenotype, usually without information about the underlying genotype. Based on phenotypic measurements of putatively adaptive traits, statistical parameters such as variances, correlations and heritabilities are used to help understand the genetics of complex phenotypes (Falconer and Mackay 1996). Techniques developed from quantitative genetics have traditionally been used for the study of natural selection in forest trees, helping to disentangle the strength, direction and possible ecological causes of selection (Neale and Ingvarsson 2008). Conversely, the starting point for population genetics approaches is variation at the DNA level. Population genetics is the study of molecular marker variation within and among populations, focusing on processes that affect genotypic and allele frequencies, such as inbreeding, mutation, migration, drift and selection. Such approaches fail to consider phenotypes, since the genes and alleles underlying most phenotypic traits are unknown, especially in non-model species.

Figure 5. Approaches for the study of adaptive genetic variation. Research areas are in CAPS within the triangle and analytical methods along triangle sides between the two data types they use. In red, research areas and methods addressed in this thesis. Adapted from (Sork *et al.* 2013).

Ecological genetics links phenotypes and genotypes with environments (Figure 5). Broadly speaking, ecological genetics is at the interface of ecology, evolution, and genetics. It can be defined as the study of the process of evolution of ecologically important traits (i.e. those traits related to fitness) occurring in present-day natural populations (Conner and Hartl 2004). We say that adaptation has occurred when a phenotypic trait (with its concomitant genomic changes) has evolved to help an organism deal with its environment. Among the four processes that can cause genetic change (mutation, genetic drift, migration and natural selection), natural selection is the only one considered to lead to adaptation. The other processes can just speed it up or constrain it. We observe natural selection when biotic and abiotic factors cause differences in fitness among individuals with different phenotypes/genotypes within populations. For natural selection to occur and change the genetic composition of the population, genetic variation is needed. This variability can come from previous mutations, recombination and/or gene flow from other populations with different genetic composition. The genetics of the population can also be affected by genetic drift if the population size is small. These genetic changes are likely to feed back and affect

phenotypic variation, changing the average phenotype in the population across generations. The resulting phenotypic changes can lead to a fitness improvement of the population; that is, it can lead to adaptation.

This thesis is framed within the field of ecological genetics, using as case study maritime and Aleppo pines, two Mediterranean pines with identified signals of local adaptation and very different evolutionary histories. Across its four chapters, I analyze phenotypic and molecular marker data using complementary approaches for the study of the evolutionary processes that enable local adaptation in forest trees.

Accurate phenotype measurements are essential for these studies. Unlike what one might expect, nowadays our ability for phenotyping high number of individuals lags behind our ability to characterize their genomes (Houle *et al.* 2010). Recent advances in genotyping techniques have dropped genotyping prices, while measuring phenotypes with accuracy and precision in natural populations or field trials continues to be both costly and time consuming (Ingvarsson and Street 2011). The phenotype is the outward appearance of an organism. The environment can make the same genotype lead to different phenotypes, and also different genotypes lead to the same phenotype. It is thus of paramount importance to evaluate phenotypes in different environments to be able to tell apart genetic from environmental effects and genotype by environment (GxE) interactions. Moreover, different traits do not usually evolve independently, but exhibit correlated responses to selection or trade-offs, most notably in forest trees (*see Chapter 2*).

In this thesis I evaluate fitness-related and life history traits using clonal common gardens (for maritime pine) and provenance common garden experiments (for Aleppo pine), located in contrasted environments. Measured traits include survival (*Chapters 1 to 4*), growth traits such as height (*Chapters 1 to 4*), diameter and stem volume, carbon isotopes (C δ 13), wood quality traits and reproduction (*Chapter 2*). A detailed description on how and where were traits measured in both species can be found in the general materials and methods section. In brief, maritime pine phenotypic traits were evaluated at five clonal common gardens (Asturias, Madrid and Cáceres -Spain-, Fundão -Portugal- and Bordeaux -France-), located in contrasted environments. Each replicated trial comprised 650 clones (from 37 provenances), with eight replicates per clone, adding up to a total of 26,000 trees phenotyped. For Aleppo pine, I used two identical progeny trials, located in Madrid and Valencia. A total of 685 trees from 28 different provenances sampled across the species range were phenotyped. Traits measured in this species included survival, growth, carbon isotopes, wood quality traits and reproduction (Figure 6).

SPECIES	PHENOTYPIC TRAIT	THESIS CHAPTER	MOLECULAR MARKERS
<i>P. pinaster</i>	Survival (0/1)	CHAPTER 1	9k SNP array 12 nuSSRs 384 SNP array <i>P. pinaster</i>
	Height	CHAPTER 3	
		CHAPTER 4
<i>P. halepensis</i>	Survival (0/1)		384 SNP array <i>P. halepensis</i>
	Height		
	Diameter	CHAPTER 2	
	Stem volume		
	Reproduction		
	Cδ13		
	Lignin content		
	Wood extractives		

Figure 6. Phenotypic traits measured for each species and molecular marker data used in each chapter.

Molecular marker information gained from genotyping is the other fundamental source of data for the study of the genetic basis of evolutionary processes and adaptation. Conifers have large and complex genomes (>20 Gbp) characterized by a large nuclear volume and rDNA repeat units, the proliferation of repetitive sequences from transposable elements and perhaps larger genes and abundant pseudogenes (Ahuja and Neale 2005; De La Torre *et al.* 2014). Pines are known to have very slow chromosomal evolution and all species feature 12 pairs of morphologically similar chromosomes (Guevara *et al.* 2005). Their DNA is also characterized by a rapid decay of linkage disequilibrium (LD) which make them especially interesting organisms for association studies (*see below*).

In maritime pine, several studies have dealt with molecular markers, but up to date most of them analyzed neutral genetic information (i.e. not affected by natural selection; e.g. Bucci *et al.* 2007; but see Grivet *et al.* 2011). Here, building on the pioneering work of Lepoittevin *et al.* (2010), who designed the first multiplex single nucleotide polymorphism (SNP) genotyping oligo pool assay (OPA) in maritime pine, I used two OPAs including 384 and 6,100 SNPs provided by the *Illumina VeraCode* and *Illumina Infinium* platforms, respectively. Both OPAs

included multiple candidate genes for adaptive traits in pines and other forest trees, as well as several SNPs coming from gene sequence of unknown function. A more detailed description of the maritime pine OPAs is provided in *Chapters 4* (for the 384-SNP OPA) and *Chapters 1 and 3* and *Annex 1* (Plomion *et al.* 2015; for the 6,100-SNPs OPA), and references therein. In addition, 12 nuclear microsatellites (nuSSRs), considered neutral, were genotyped following Santos-del-Blanco *et al.* (2012) and references therein. In Aleppo pine, molecular studies are scarcer (e.g. Grivet *et al.* 2009), given the relatively lower economic importance of this species. In this case, I used the 384-SNP OPA described in Pinosio *et al.* (2014). Briefly, this genotyping assay contained 240 SNPs selected from an Aleppo pine transcriptome assembly involving trees with highly contrasted phenotypes and 144 additional SNPs from candidate genes for adaptive traits in this species.

Knowledge on the genetic architecture of adaptive traits can help address outstanding questions such as, how many *loci* are involved in adaptation, and what is the relative contribution of major and minor effect alleles to adaptive evolution (Stinchcombe and Hoekstra 2008). But this can be a tough challenge, as most adaptive plant traits are quantitative polygenic traits (see below: 5. *Polygenic adaptation*). Ideally, an allele should be considered adaptive only when connections are made between its genotype, phenotype and fitness (Barrett and Hoekstra 2011). The identification of such *loci* is especially challenging in species with limited genomic resources, as in the Mediterranean pines studied here (Savolainen *et al.* 2013). However, candidate gene approaches and growing genetic resources have already led to very interesting results in the species, and look promising to advance in the study of adaptive evolution.

Studying the relationship between genotype and phenotype

1. Heterozygosity-fitness correlations

The link between molecular and phenotypic variation is often complex. One of the traditional approaches to address this issue and gain insights in how natural selection acts in contemporary populations, is the study of statistical associations between multilocus heterozygosity and phenotypic traits (i.e. heterozygosity-fitness correlations or HFCs). It is well known that a decrease in heterozygosity, for example, due to mating between relatives or population bottlenecks, can lead to a concomitant reduction in population fitness. Molecular markers (e.g. allozymes, microsatellites, SNPs) can be used to score individual heterozygosity, which can then be correlated to different fitness traits. This tool has been broadly used in the last decades in multiple organisms to study the impact of inbreeding in natural populations. However, their underlying mechanisms are often hotly debated (reviewed in Hansson and Westerberg 2002; Szulkin *et al.* 2010). Three main hypothesis

have been proposed: the *direct* (i.e. heterozygosity in a locus is directly associated with fitness) and *local* (i.e. a locus affects fitness through linkage disequilibrium with the associated locus) *effect hypotheses*, which explain HFCs in terms of heterozygote advantage at particular *loci* (David 1998; Hansson and Westerberg 2002), and the *general effect hypothesis*, which assumes that molecular marker heterozygosity reflects genome-wide heterozygosity, which, in turn, is correlated with the individual inbreeding level and the expression of deleterious recessive alleles. The proposed causative hypotheses are not mutually exclusive. Meta-analyses have shown that HFCs, although positive overall, are weak (explaining around 1% of the variance of fitness traits in studies with microsatellites), far from universal, and context-dependent (Coltman and Slate 2003; Chapman *et al.* 2009; Szulkin *et al.* 2010). The local and direct hypothesis tend to predominate in the literature, although it has been recently argued that inappropriate statistical testing may have biased the results towards the predominance of these hypotheses (Szulkin *et al.* 2010). Several authors have also pointed out a possible publication bias towards reporting positive, significant HFCs (Coltman and Slate 2003; Chapman *et al.* 2009). Another highly controversial issue is to what extent heterozygosity measured at a small number of molecular markers (typically less than twenty) can capture whole-genome heterozygosity (Balloux *et al.* 2004). Understanding how molecular marker diversity (i.e. heterozygosity) correlates with fitness traits, at different spatial scales, and the underlying mechanisms, remain a key issue today in the study of the complex interaction between genetics, ecology and evolution (Szulkin *et al.* 2010).

In *Chapter 1* I address HFCs in maritime pine at different spatial scales and under contrasting environmental conditions, using 6100 SNPs and survival as a fitness proxy. I test for the possible underlying mechanisms and introduce a novel heterozygosity-phenotype association test to look for direct/local effects. The use of a high number of SNP markers with a wide heterozygosity range, and the performance of analyses at different spatial levels and environments, accounting also for population structure, maximized the chances to detect HFCs in the species, overcoming previous drawbacks of traditional HFC studies.

2. Genotype-phenotype association studies

The main goal of association genetics is the identification of statistical associations between phenotypic variation of ecological or economically important traits and allelic polymorphisms in certain genes. Identifying links between phenotype and genotype can increase our understanding of the genomic architecture of adaptive traits and thus of adaptive evolution (Zhu *et al.* 2008; Jaramillo-correa *et al.* 2015). It can also help to provide a more holistic

understanding of the link between evolutionary and ecological processes. As DNA sequencing technologies come down in price, a larger number of individuals and molecular markers can be genotyped even in non-model organisms. Then, these data can be used to scan for phenotype-genotype links using a variety of statistical techniques (Li *et al.* 2008). To date, two main approaches have been developed to achieve this goal: Quantitative Trait Locus (QTL) mapping and association (or linkage disequilibrium) mapping. The main difference between traditional QTL and association mapping is the nature of the mapping population (Cardon and Bell 2001; Neale and Savolainen 2004). The QTL approach uses a segregating population created artificially by crossing different genotypes (e.g. F2 and backcross) to generate linkage disequilibrium between genetic markers and QTLs. The number of crossover events is limited and the map resolution of QTLs is determined by the size of the progeny array. Conversely, in association mapping, individuals are selected from non-structured populations, taking advantage of lower levels of LD due to longer histories of recombination existing within natural populations (Neale and Savolainen 2004; Fournier-Level *et al.* 2011). Indeed, after many generations of recombination and random mating only tightly linked *loci* will show statistical association, allowing a finer mapping than standard QTL approaches (Savolainen *et al.* 2013). Moreover, association studies often use a higher number of molecular markers (SNPs), adding to advantages that have made association mapping the preferred approach for dissecting quantitative traits in numerous organisms (Hirschhorn and Daly 2005).

Depending on the scale and focus of the study, association mapping falls into two broad categories: (1) genome-wide association mapping (GWAS), which surveys genetic variation in the whole genome to find signals of association; and (2) candidate-gene association mapping (Figure 7), which relates polymorphisms in selected candidate genes that have alleged roles controlling phenotypic variation for specific traits (Zhu *et al.* 2008). Most GWAS have been conducted in model species (e.g. Yang *et al.* 2010 in humans, Atwell *et al.* 2010 and Fournier-Level *et al.* 2011 in *Arabidopsis thaliana*) or economically important crops (e.g. Li *et al.* 2013 in maize, Hwang *et al.* 2014 in soybean) with well-known genomes. For non-model species with limited genomic resources such as forest trees, the preferred approach is candidate-gene association mapping, which was pioneered in *Drosophila* long ago (Mackay and Langley 1990). Relevant candidate genes must be selected previously, based on the evaluation of available results from genetic, biochemical, or physiology studies in the same or related species. The candidate-gene-based approach has proven effective in conifers (e.g. González-Martínez *et al.* 2007, 2008; Eckert *et al.* 2009; Quesada *et al.* 2010; Beaulieu *et al.* 2011) and might even be optimal and certainly most feasible and cost-effective, given the characteristics of conifer genome variation (Neale and Savolainen 2004). Conifers, and

particularly pines, are well-suited for association studies due to (1) high levels of genetic variation, (2) rapid decay of LD due to high population recombination rates, (3) low levels of domestication, (4) large natural populations and outcrossing mating systems and (5) accurate evaluation of phenotypes, due to the possibility of establishing clonal replicates or large progenies.

Figure 7. Schematic diagram of candidate-gene phenotype-genotype association tests. Population structure (Q) and the relative kinship (K) can be included in the model to avoid false positives. E stands for residual variance. Adapted from (Zhu *et al.* 2008).

Among the available statistical methods for association, the mixed linear model (MLM) developed by Yu *et al.* (2006) has been the most widely used, flexible and powerful controlling false positives. False positives generated by population structure and the species demographic history are indeed regarded as the main hurdle for association mapping in plants (Zhu *et al.* 2008). In this method, random markers are previously used to estimate population structure (Q) and a relative kinship matrix (K), which are then fit into a MLM to test for marker trait associations (Figure 7). This way, genetic relatedness is accounted for, preventing spurious associations if population structure and/or unequal individual relatedness exist (Pritchard *et al.* 2000; Yu *et al.* 2006; Zhu *et al.* 2008). Strong associations between a SNP and the studied phenotype are interpreted as indicating that the SNP, or a nearby correlated SNP, likely affecting the phenotype (e.g. Figure 8). Multiple testing correction is also needed to control for false positives when a large number of markers and traits are tested. Besides single-SNP analysis, which tests each SNP one at a time, recently developed multi-SNP Bayesian approaches (e.g. Guan and Stephens 2011; Li *et al.* 2012),

seem promising to identify higher numbers of associated SNPs. However, the current statistical approaches used to associate genetic markers to phenotypes can only identify *loci* of moderate to large effect, lacking the statistical power to detect all causative variants and leading to the so-called problem of missing heritability (Lipka *et al.* 2015). Fortunately, genotyping and phenotyping are progressing extremely fast, and so do statistical approaches, continuously evolving at a breakneck pace, offering great promise for association genetics in complex trait dissection.

Figure 8. Genotypic effects of SNPs that showed significant genetic association ($P < 0.0005$) with female flowering (SNP 163), stem diameter (SNP 211) and total wood extractives (SNP 264) in candidate-gene single-locus association tests in *P. halepensis* (see Chapter 2).

By examining associated SNPs in more detail we may glean important insights into the biology of the phenotype under study. Association mapping can be also useful to select nucleotide polymorphisms for population genetic studies aiming at revealing adaptive divergence in natural populations (see Chapter 4). Plasticity and environmental effects affect phenotypic measurements, and thus can also blur association test results. Association studies of fitness traits measured in multiple natural environments, combined with geographic and climatic analyses, constitute a powerful approach to identify environment-specific candidate genes for local adaptation (Fournier-Level *et al.* 2011; see methodological approach in Chapters 1, 2 and 3). Association mapping studies in conifers should greatly improve our understanding of the architecture of complex traits in plants and other organisms (Neale and Savolainen 2004).

I use genotype-phenotype association tests in Chapters 1, 2 and 3, combined with quantitative genetics and population genetics approaches for an enhanced understanding of the genetic architecture of the studied traits. Association tests are performed over genotypes replicated in contrasting environments, with accurate assessment of environmental effects.

Polygenic adaptation

Most ecologically, evolutionary and economically important traits are polygenic i.e. controlled by many *loci* of relatively small effect (Mackay *et al.* 2009; Pritchard *et al.* 2010; Rockman 2012). Adaptive traits show even more continuous patterns of variation and are frequently controlled by a more complex and polygenic genetic architecture, as exemplified by studies in *Arabidopsis*, maize or *Drosophila* (Comeault *et al.* 2014). Thus, a quantitative genetics framework seems more feasible than the traditional “selective sweep” hypothesis to detect *loci* under local adaptation (Pritchard and Di Rienzo 2010; Le Corre and Kremer 2012). Indeed, the molecular evidence of selection is expected to be weak because the signal of selection is distributed across many *loci* (Le Corre and Kremer 2012; Turchin *et al.* 2012; Kemper *et al.* 2014). The concept of polygenic adaptation is rooted in classical quantitative genetics, with well-established theoretical bases (see reviews by Pritchard *et al.* 2010; Le Corre and Kremer 2012). It has been recently suggested that we need to move from searching for alleles of strong effect and resume the use of population genetic data to advance in our understanding of local adaptation as a process (Turchin *et al.* 2012; Tiffin and Ross-Ibarra 2014). Hancock *et al.* (2010) found that subtle shifts in allele frequencies were responsible for human adaptations to diet, subsistence and ecoregion, identifying SNPs that showed concerted allele frequency changes among populations with respect to specific environmental traits. Later, Turchin *et al.* (2012) published a landmark study in which they found weak selection coefficients and widespread allele frequency differences in *loci* associated with human height across Northern and Southern European populations. But polygenic adaptation studies are still scarce and almost nonexistent outside humans, although approaches that investigate the signal of selection aggregated across *loci* show great promise (Tiffin and Ross-Ibarra 2014). For example, Berg and Coop (2014) developed a method which uses candidate associated *loci* from association mapping results to identify selection on quantitative phenotypes. They use a quantitative genetics approach and test for variance excess in associated *loci* when compared with neutral models built from available non-associated marker data. Polygenic adaptation could allow effective adaptation to environmental changes, but would do so with little effect on substitution rates and without selective sweeps (Pritchard *et al.* 2010) (Figure 9). Future studies in this direction should shed light on the relative role of hard sweeps on new mutations and polygenic adaptation processes from standing genetic variation in local adaptation and evolution.

Figure 9. Illustration of the hard sweep, soft sweep and polygenic adaptation models. Circles indicate alleles that are favoured following an environmental change. In the hard sweep model, selection drives a new mutation to fixation, creating a large region with reduced polymorphism. In the soft sweep, models favoured alleles to increase in frequency, but diversity is maintained. In the polygenic model, previously to selection, favourable alleles exist at modest frequencies at various *loci* across the genome. After selection, they are more abundant, but do not become fixed at any locus. Modified from (Pritchard *et al.* 2010).

In *Chapter 3* I address polygenic adaptation in maritime pine. Local adaptation across the species range is approached from a quantitative genetics perspective, adding up the advantages of GWAS and traditional quantitative genetic indexes for the identification of adaptation events driven by many *loci* of small effect. I show that, although association tests may fail to identify all markers underlying the studied polygenic traits, their results combined with quantitative genetic analyses can help us identify signals of local adaptation also in non-model species.

Environmental association: ‘reverse ecology’

Reverse ecology aims to obtain novel insights into the ecology of an organism through genomic and environmental data (Figure 10). It uses genomic tools and population genomics methods to detect selection. Genotype-environment associations try to identify underlying adaptive genetic variants (i.e. discover the genetic markers that are associated with a particular habitat) as well as putative environmental drivers without prior information about the relevance of ecological traits (Li *et al.* 2008). These methods can allow us to find the targets of natural selection with little prior knowledge on how selection targeted that trait. Most commonly used methods to identify SNP-environment associations include multivariate logistic regression (Grivet *et al.* 2011; De Mita *et al.* 2013), latent factor mixed models (Frichot *et al.* 2013) and testing for unusual allele frequency patterns and correlations between environmental variables and allele frequencies using Bayesian analysis (Günther

and Coop 2013). If performed at different spatial scales, they can be used to test for local adaptation: based on differences in allele frequencies at the population level, they can reveal if species distributed across heterogeneous environments at the regional scale are under differential selection. Ideally, locally adapted SNPs could then be tested for correlations with phenotypes for validation, to confirm geographically defined molecular proxies for climate adaptation (Jaramillo-Correa *et al.* 2015).

Although promising, reverse-ecology approaches also have limitations. The main one is the geographical basis of both climate and genetic variation. If the climatic variables used are spatially structured or partially match the genetic structure of the species, correlations may be biased and confound the interpretation of the analyses. Patterns of gene flow and genetic drift can also lead to gene frequency clines at neutral *loci* (Vasemagi 2006). Thus, these methods need to take into account background levels of population structure if these correlations are used to identify genes affected by natural selection along climatic gradients. They are also limited in non-model species by incomplete knowledge of gene function. Moreover, just as it happens in genotype-phenotype association studies, the molecular evidence of selection is expected to be weak because the signal of selection in polygenic quantitative traits is distributed across many *loci* influenced in turn by multiple environmental factors (Tiffin and Ross-Ibarra 2014).

In *Annex 2*, I aimed to identify genetic polymorphisms related to climate adaptation in maritime pine. After single-locus genotype-environment association tests, I used a common garden to evaluate survival under extreme climatic conditions to validate SNP–climate associations and demonstrate the utility of these genetic markers to predict adaptation to future climates. Provided the polygenic basis of adaptive traits, moving towards a polygenic model also in genotype-environment correlation tests seems promising. Thereby, in *Chapter 3* I test for unusual allele frequency patterns and correlations between environmental variables and allele frequencies under a polygenic adaptation framework. Two sets of SNPs (neutral and putatively associated with adaptive traits) are used to detect unusually strong correlations between genetic values of adaptive SNPs and specific environmental variables, when compared to a neutral drift model. A significant correlation was found, consistent with the hypothesis that populations are locally adapted, via the phenotype, to local conditions that are correlated with environmental variables.

Applications for conservation

How to best ensure species' persistence, now that climate change touches every part of the globe, is a difficult challenge (Hoffmann and Sgrò 2011). Mediterranean forest ecosystems are particularly sensitive to rapid climate changes and should be considered prime targets for conservation efforts (Myers *et al.* 2000; Fady-Welterlen 2005; Lindner *et al.* 2010). Over the last decades European countries have made considerable efforts to conserve the genetic diversity of tree species (Vries *et al.* 2014; see www.euforgen.org). Although most European tree species are not threatened (IUCN category of 'least concern'), there is a recognized need for intraspecific conservation in many of them, including Mediterranean pines (Koskela *et al.* 2013; Lefèvre *et al.* 2013). The current preferred approach for preserving trees' genetic diversity is dynamic conservation, which aims to maintain evolutionary processes and adaptive potential in tree populations, to ensure their long-term sustainability (Eriksson *et al.* 1993; Namkoong 1997). Under an evolutionary perspective, local adaptation patterns and correlations between fitness traits and environmental variables are crucial to identify distinct evolutionary significant units for conservation (Leimu and Fischer 2008). The usefulness of molecular markers and quantitative trait data in conservation and forest management has been broadly recognized. However, to date, genetic information is largely neglected in conservation planning and management policies.

Petit *et al.* (1998) suggested the use of molecular markers to select target populations for conservation, based on partitioning the contribution of the population to total diversity into allelic diversity of the population and its divergence from the other populations. The alternative approach (the one currently used in the European conservation program), is based on climatic and geographical information. It assumes that environmental classifications are closely linked with genetic differentiation, under the hypothesis of local adaptation and clinal variation of tree species (Lefèvre *et al.* 2013). However, local adaptation is not necessarily driven by obvious environmental differences and, most likely, different biotic and abiotic factors drive local adaptation in different species (Leimu and Fischer 2008). Molecular markers can be used as a starting point in conservation policies, as they provide information about historical evolutionary divergence and demographic processes (Fraser and Bernatchez 2001). But molecular genetic markers alone do not provide a reliable tool for conservation geneticists wanting to measure a population's adaptive potential, as patterns identified in quantitative adaptive traits normally differ from those observed using neutral or weakly selected *loci* (Fraser and Bernatchez 2001; Frankham 2010; Mittell *et al.* 2015). A combined approach including molecular, quantitative and ecological data, although rarely employed, is fundamental and would allow the integration of evolutionary processes in conservation planning (Rader *et al.* 2005; Ouborg *et*

al. 2006; Sgrò *et al.* 2011). Precise and integrated data in neutral and adaptive genetic patterns are essential to design effective tree dynamic conservation strategies to minimize biodiversity losses in the face of climate change. Thus, applying gained knowledge in the genetic basis of adaptive traits coming from quantitative and population genetics, association and environmental studies, could make a difference in the effectiveness of pine species conservation. It is thus urgent to bridge the gap between basic research in ecology and evolution and conservation management (Shafer *et al.* 2015).

In *Chapter 4*, I address maritime conservation strategies using environmental, molecular marker and quantitative trait data. Gained knowledge on the genetic basis of fitness related traits is applied to define potential dynamic conservation units in the species. I introduce key molecular and quantitative trait statistics to assess population differentiation and identify candidate population groups for conservation. Environmental variation is a good starting point for *in-situ* gene conservation strategies, but neither a purely environmental nor a neutral molecular marker classification are able capture all variability that should be considered in conservation planning. The described approach represents a step forward in setting priorities for upcoming conservation efforts, stressing the effectiveness of gradual integration of genetic information at different scales and time-frames as it becomes available for the species.

Thesis overview and organization

This thesis includes firstly a general introduction, a thesis overview note and the thesis objectives. A general materials and methods section follows, describing in detail the phenotyping procedures used to obtain pine phenotypes of different adaptive traits. Chapters 1 to 4 are based upon already published papers (*Chapters 1 and 4*) and manuscripts in preparation (*Chapters 2 and 3*), and include a brief introduction, materials and methods, results and discussion. The general phenotyping materials and methods section contains information about plant material and phenotyping shared by the four chapters, but extended particular materials and methods sections are included in each chapter. The thesis also includes a brief discussion concluding remark and the thesis conclusions. Finally, you will find supplementary material, with additional tables and figures for the four chapters, and two annexes.

Annex 1 reproduces explicitly a publication I co-authored, describing the development of a 9,000 *Illumina Infinium* SNP array for maritime pine. Genotyping data from this SNP array was used in *Chapters 1 and 3*. *Annex 2* also reproduces explicitly a publication I co-authored, which explores single-locus genotype-environment associations. Both publications in annexes 1 and 2 are consistently cited throughout the thesis chapters, and are important for their understanding.

The thesis is framed within the study of local adaptation, addressing the complex interaction between genotypes, phenotypes and the environment (Figure 10).

Figure 10. Overview of the thesis organization: type of data (orange), methodological approaches (blue), and chapters/annexes (green) addressing the study of the complex interaction between genotypes, phenotypes and the environment, which condition tree fitness in a global change scenario.

OBJECTIVES

The main aim of this thesis is to improve the understanding of the evolutionary processes leading to local adaptation in forest trees, through disentangling the genetic basis of important adaptive traits in the Mediterranean forest tree species *Pinus pinaster* and *P. halepensis*.

I aimed to assess distinct sources of intraspecific phenotypic variation, unravelling local adaptation from phenotypic plasticity, and to identify particular genomic regions underlying fitness-related traits in the studied species, as well as identifying possible ecological drivers of local adaptation. Throughout my work I also intended that the research questions posed here could be related to practical issues related to conservation of forest genetic resources under current climate change.

The specific objectives of this thesis were (see also summary in Table 1):

1. Study intraspecific standing variation and population differentiation in adaptive traits, such as survival and growth, and estimate quantitative genetic parameters to elucidate the genetic control of the target adaptive traits (*Chapters 1 to 4*).
2. Identify putative genes underlying the studied adaptive traits (*Chapters 1 to 3*).
3. Test for correlations between marker heterozygosity and fitness (HFCs), at different spatial scales and environments, to find out if inbreeding leads to reduced fitness in long-lived forest trees (*Chapter 1*).
4. Identify the causes underlying significant HFCs. Is genome-wide heterozygosity correlated with fitness, so that increased individual inbreeding levels lead to the expression of deleterious recessive alleles? Or does heterozygosity at particular *loci* increase fitness through heterozygote advantage? (*Chapter 1*).
5. Assess the adequacy of using integrated phenotypes including different traits related to fitness to identify genes underlying adaptive traits and possible pleiotropic effects (*Chapter 2*).
6. Test the hypothesis that fitness components are related to and modulated by the environment, and that correlations and trade-offs among traits match predictions from allocation and life history theories in pines (*Chapter 2*).
7. Prove the effectiveness of a polygenic adaptation approach, combining population genetics and genotype-phenotype association tests to identify signals of local adaptation (*Chapter 3*).
8. Assess the importance of considering both molecular and phenotypic variation in adaptive traits for forest tree conservation under climate change (*Chapter 4*).

Table 1. Overview of main objectives, materials and methods and results in publications and manuscripts included in the thesis.

Chapter	Specific Objectives	Species	Materials and methods			Results
			Genetic markers	Phenotypes	Statistical methods	
1	<ul style="list-style-type: none"> Correlation between heterozygosity and fitness Identify underlying process 	<i>P. pinaster</i>	6100 SNPs	Survival	<ul style="list-style-type: none"> REML GMLM* for BLUP** estimation Standard correlations Bayesian estimation of individual <i>loci</i> heterozygosity effects 	Rodríguez-Quilón <i>et al.</i> (2015). <i>Proc Roy Soc B</i>
2	<ul style="list-style-type: none"> Ecotypic trends and trait trade-offs Genotype-phenotype associations for integrated phenotypes 	<i>P. halepensis</i>	384 SNPs	Wood quality Cδ13 Growth Reproduction	<ul style="list-style-type: none"> REML MLM*** for BLUP estimation Bivariate Bayesian MLM for genetic correlations among traits Principal Component Analysis Standard correlations MLM for single-locus association tests Bayesian Variable Selection Regression for multi-locus association tests 	Rodríguez-Quilón <i>et al.</i> (In preparation)
3	<ul style="list-style-type: none"> Genotype-phenotype associations Detection of signals of polygenic adaptation 	<i>P. pinaster</i>	6100 SNPs	Survival Height	<ul style="list-style-type: none"> Bayesian MLM and GMLM for trait BLUP and variance components estimation Quantitative genetics indexes MLM for single-locus association tests Bayesian estimation of allelic effects Variance and covariance excess tests 	Rodríguez-Quilón <i>et al.</i> (In preparation)
4	<ul style="list-style-type: none"> Asses intraspecific genetic and phenotypic variation Consider genetic and phenotypic variation for conservation genetics 	<i>P. pinaster</i>	12 nuSSRs + 384 SNPs	Survival Height	<ul style="list-style-type: none"> Bayesian MLM and GMLM for trait BLUP and variance components estimation Quantitative genetics indexes Hierarchical clustering analysis Standard correlations 	Rodríguez-Quilón <i>et al.</i> (2016). <i>Ecol Appl</i>
Annex						
1	Develop a 9000 <i>Illumina Infinium</i> SNP array for maritime pine	<i>P. pinaster</i>	9000 SNPs	-	<ul style="list-style-type: none"> Bioinformatic processing (SNP redundancy, alignment, SNP detection, annotation) 	Plomion <i>et al.</i> (2015) <i>Mol Ecol Res</i>
2	Identify SNPs related to climate adaptation	<i>P. pinaster</i>	12 nuSSRs+ 384 + 2646 SNPs	Survival	<ul style="list-style-type: none"> Multi-variate logistic regressions Latent factor Mixed Models Environmental Bayesian assoc. tests MLM for BLUP estimation Principal Component Analysis Standard and nonparametric correlations 	Jaramillo-Correa <i>et al.</i> (2015) <i>Genetics</i>

* Generalized Mixed Linear Models adjusted using Restricted Maximum Likelihood ** Best Linear Unbiased Predictor *** Mixed Linear Models

MATERIALS AND METHODS: PHENOTYPING

Here I describe the plant material, measured traits and methods used to phenotype trees throughout the four following chapters. Data analysis and statistical methods are described in depth in the materials and methods sections in each particular chapter. Molecular marker materials and methods are also included in each chapter. The development of the Illumina Infinium SNP array for maritime pine is described in Annex 1.

A total of 25,400 *Pinus pinaster* trees and 700 *P. halepensis* trees were phenotyped for different adaptive traits. Different tree samples and trial sites were used for each chapter, according to the study design and genotypes and phenotypes available at the time of data analysis (see Figure 6 in the *Introduction* section).

Tree phenotypes have been measured for decades in common gardens and provenance trials, but usually in a limited number of easy to measure and non destructive traits (e.g. tree height or diameter). However, our ability to characterize phenomes (i.e. the full set of phenotypes of an individual) begins to lag behind our ability to characterize genomes, as genotyping prices keep dropping (Alfá and Majada 2013). Since typical genotype–phenotype genetic studies usually involve a relatively large number of accessions, phenotyping traits with high accuracy and precision can be now both more costly and time-consuming than genotyping (Houle *et al.* 2010; Ingvarsson and Street 2011). For this thesis I measured survival, growth and reproductive traits in field trials, for a very large number of trees, in the traditional way. But many other traits of interest in forest trees, such as wood quality, are complex in nature and difficult to evaluate (Groover 2007). When it comes to large numbers of trees, destructive sampling and wet laboratory analysis can become extremely time-consuming and owing to practical constraints, so that the number of trees that can be characterized becomes limited. Nevertheless, new non-destructive high-throughput phenotyping technologies have been developed (summarized in table 4.1 from Alfá and Majada, 2013) and their use is becoming increasingly widespread. From this set of techniques, in this thesis I used Near Infrared Spectroscopy (NIR) for wood quality analysis (see below).

Common garden experiments

All phenotypic data used in this thesis originate from several common garden experiments installed under contrasting environmental conditions. I used clonal common gardens and provenance common gardens for *P. pinaster* and *P. halepensis*, respectively. Clonal testing and provenance testing have enabled precise and thorough evaluation of tree phenotypes for

a long time (Neale and Savolainen 2004). The opportunity to replicate individual genotypes or plant related individuals, both within and across multiple environments, allows for far more precise phenotyping as the amount of environmentally induced variation can be estimated and partitioned out (Ingvarsson and Street 2011).

P. pinaster clonal common gardens comprised trees from 37 provenances, sampled across the species natural distribution range. Open-pollinated seeds were collected in natural stands, afterwards planted, and seedlings grown in nursery. A variable number of trees (ranging from 1 up to 28, average of 15 trees per population) was obtained from each provenance. Later, each tree was propagated vegetatively by cuttings (see protocol in Majada et al. 2011), to get forty replicates of each genotype. These clones were used to establish five clonal trials, with 650 clones and eight ramets per clone, set in a randomized complete block design in 2010 (Figure A1.1). This clonal common garden design allowed multiple phenotypic measurements for each genotype and accurate evaluation of genetic and environmental effects, as we have the same genotypes replicated and growing in contrasting environments. The four clonal trials located in Asturias, Portugal, Madrid and Cáceres, constitute identical copies (5,200 trees in each trial), while the fifth trial, located in Bordeaux, was a bit smaller (4,600 trees). Trials in Asturias, Portugal and Bordeaux are located in the Atlantic region, with high annual rainfall and mild temperatures. Nevertheless, Portugal shows less than half the summer rainfall than Asturias and Bordeaux. By contrast, trials in Madrid and Cáceres are located in continental areas under Mediterranean influence, with large seasonal temperature oscillations and a marked summer drought. Apart from its marked precipitation seasonal oscillation, clay soils make Cáceres a worse site than Madrid for maritime pine survival and growth (Table A1.1). A different subset of clones from the complete clonal common garden was selected in each chapter (*Chapters 1, 3 and 4*), depending on the research question and phenotypes and genotyping data available by the time data analyses were performed.

Table A1.1. Environmental conditions (precipitation and temperature) at *Pinus pinaster* clonal common garden sites.

Trial	Annual rainfall (mm)	Rainfall in June, July and August (mm)	Mean temperature (°C)	Temperature range (°C)
Asturias	1160	152	11.3	11.4
Portugal	985	63	14.2	16.3
Bordeaux	859	196	11.9	15.3
Madrid	443	63	13.7	19.3
Cáceres	898	46	15.5	18.7

Figure A1.1. Plant material and design of *Pinus pinaster* clonal common garden, with five replicates in contrasting environments. Genotypes were obtained from mother plants, while phenotypes were measured in the field, in the five trial sites. The experimental design allowed multiple measurements of each genotype at each location and under different environmental conditions.

Figure A1.2. *Pinus pinaster* clonal common gardens: A, Asturias trial, one year after the establishment and B, Fundão (Portugal), four years after the establishment.

The *P. halepensis* provenance common garden comprised trees from 52 native range-wide populations, originated in continental Spain, Balearic Islands (Spain), Italy, France, Tunisia and Greece. It consists in two identical provenance trials, located in Valdeolmos (Madrid) and Altura (Valencia; Figure A1.3). Seeds were sampled in natural populations, collecting cones from about 30 trees, at least 100 m apart from each other, in each population. Then, seeds were bulked and seedlings derived from those seeds were grown under identical conditions in nursery, and planted into the final common garden sites after one or two years. Both trial sites provided contrasting environmental conditions. Valdeolmos study site has a deep alluvial soil and the area has a continental climate with Mediterranean influence, and marked summer drought. By contrast, the trial in Altura has a Mediterranean climate, with mild winters and warm summers (Table A1.2). Both common gardens were measured periodically since their establishment in 1997, so that wide data series were already available prior to this work.

Figure A1.3. *Pinus halepensis* provenance trials located in Altura (A) and Valdeolmos (B).

Table A1.2. Environmental conditions (precipitation and temperature) at *Pinus halepensis* provenance trial sites.

Trial	Annual rainfall (mm)	Rainfall in June, July and August (mm)	Mean temperature (°C)	Temperature range (°C)
Altura	580	141	14.6	14.5
Valdeolmos	498	156	13.5	19.2

Studied traits

Survival

Tree survival was evaluated in the field along the years, and recorded as a binary variable (1 for live trees and 0 for dead trees, Figure A1.4).

Figure A1.4. Survival assessment in *Pinus pinaster* Portugal common garden: A, dead tree from Petrocq provenance; B, three-year old live tree from Petrocq provenance and C, common garden overview four years after the establishment (five-year old trees).

Growth traits

Tree growth was estimated through total height, stem diameter, and derived stem volume. Total height was measured with a ruler or with a telescopic pole, depending on tree total height, to the nearest cm. Diameter at breast height was measured with a caliper to the nearest mm. Stem volume was estimated indirectly from height and survival measurements using the following formula: $V = 1/12\pi \cdot h \cdot d^2$

Wood quality

Near infrared spectroscopy (NIR) provides a rapid, non-destructive method for the estimation of wood properties (Schimleck 2007; Isik *et al.* 2011). The technology is particularly well suited for studies where large numbers of samples need to be analyzed, providing an indirect chemical characterization of wood samples. NIR spectroscopy has been widely used in agriculture, pharmaceutical and food industries, but its application in wood science is recent (first results were published in the 90's) and provides an efficient alternative to costly and time-consuming wet chemistry methods (Schimleck 2007; Alía and Majada 2013).

To obtain each wood sample, wood shavings were collected by drilling a 1 cm wide and 6 cm deep hole into the tree at breast height. When the tree stem was very narrow, we reduced the hole depth and made up to three holes in the stem to get enough wood shavings (> 7 g). The samples were then oven-dried at 60°C for 18 h, and afterwards kept at 30°C in a silica gel desiccator. Then, each sample was cleaned, removing impurities (bark, phloem, needles), weighted and introduced inside an airtight vial. Afterwards, at the ISA-IICT Wood Laboratory (<https://www.isa.ulisboa.pt/en/cef>), samples were grinded using two mills, consecutively, to get wood dust for NIR analysis. Samples were then oven-dried again, at 60°C for 8h. Finally, samples were introduced in the NIR bucket to get NIR spectra (Figure A1.5).

Near infrared spectra acquisition was carried out using a Bruker Vector 22 N spectrometer (Bruker Optics, Ettlingen, Germany). To get wood quality data (e.g. extractives, lignin content) from NIR spectra, correlations with standard destructive laboratory procedure data should be performed for each particular species. Thus, labwork for dichloromethane, ethanol and water extractives, lignin content and H/G ratio determination was performed in a subsample of 72 wood samples (with highly variable spectra) to set up the calibration curves. A partial least squares regression model was implemented for calibration, using the OPUS Quant software from Bruker Optics (see details in Alves *et al.* 2010). Finally, we were able to obtain accurate data on wood extractives, lignin content and H/G ratio. Wood density estimates are still pending due to calibration problems.

Figure A1.5. Sample preparation for NIR analysis of wood property traits: A, debarking and drilling to get wood shavings; B, wood sample sorting and cleaning at the laboratory; C, wood samples inside airtight containers; D, mills; E, wood samples after first and second grinding; F, sample introduction into NIR bucket and G, NIR spectrometer getting spectra of a wood sample.

Carbon isotopes

As reviewed by Warren *et al.* (2001), $\delta^{13}\text{C}$ may be a useful indicator of water use efficiency (WUE) for conifers in seasonally dry climates. Carbon isotope discrimination ($\delta^{13}\text{C}$) measures the ratio of stable carbon isotopes ($^{13}\text{C}/^{12}\text{C}$) in the plant dry matter compared to the ratio in the atmosphere (Tuberosa 2012). The accuracy of $\delta^{13}\text{C}$ from *P. halepensis* wood as a proxy for WUE estimation has been previously confirmed by (Ferrio *et al.* 2003; Voltas *et al.* 2008, 2015).

The same samples that were used for NIR wood analyses, were also used for $\delta^{13}\text{C}$ analysis. After grounding, a small subsample from each sample was kept for carbon isotope analysis. Each sample was homogenized and introduced in a 5x8 mm tin capsule. The exact weight of each sample was measured using a precision scale (Figure A1.6). Then, the carbon isotopic composition ($\delta^{13}\text{C}$) of wood samples was determined by mass spectrometric analysis at the University of California Davis Stable Isotope Facility (<http://stableisotopefacility.ucdavis.edu/>). Carbon isotope discrimination was later calculated from the isotope compositions of the plant material (δ_p) and that of the air (δ_a , assumed to be -8‰), as defined by (Farquhar *et al.* 1982):

$$\Delta^{13}\text{C} = \frac{\delta_a - \delta_p}{1 + \delta_p}$$

Figure A1.6. $\delta^{13}\text{C}$ sample preparation: A, wood samples; B, 5x8 mm tin capsules; C, weighting in a precision scale; D, closing and compressing the tin capsule and E, wood samples prepared for their shipment for $\delta^{13}\text{C}$ analysis.

Reproduction

Total number of female cones per tree were approximated from counts during 15 seconds. This time interval was chosen after several tentative counts by different observers, aiming at combining a reasonable correlation with full cone count and a reasonable speed needed for measuring hundreds of trees with affordable field work schedules. Male pollen cone clusters were also counted in 15 seconds. Quantitative count data and a qualitative assessment of their size and density were then used to classify male reproduction in a qualitative 1-4 scale. For more details on reproduction measurements see Santos-del-Blanco (2014) and references therein.

Trait measurements

All trait measurements performed in the common gardens used for this thesis can be found in Table A1.3 (for *P. pinaster*) and Table A1.4 (for *P. halepensis*). For *P. pinaster*, only survival and height measurements were used for the studies included in this thesis (*Chapters 1, 3 and 4*). Trials in continental-Mediterranean environments (Madrid and Cáceres) were only measured after one year in the field, because of very high mortality. Unfortunately, most trees died in both locations because of severe drought during the first summer after field establishment. For *P. halepensis*, all traits included in the table (growth, reproduction, wood quality and carbon isotopes) were used (*Chapter 2*), but only the last measurement was retained for the study. See Figure 6 in the *Introduction* for a summary of the traits included in each chapter.

Table A1.3. Trait measurements of *Pinus pinaster* clonal trials 2011-2015. Only survival and height measurements were used for the studies included in this thesis (Chapters 1, 3 and 4).

Trial / Trait	Survival	Height	Phenology	Nº bud elongations/ season	Specific Leaf Area Index	Cδ13+ Nδ15 (Leaves)
Asturias	Dic 2011 Nov 2012 Mar 2014	Dic 2011 Nov 2012 Mar 2014	Growth index (height for different buds in same season)	Nov 2012 Mar 2014	-	-
Portugal	Jan 2012 May 2012 Oct 2012 May 2013 Nov 2013	Jan 2012 May 2012 Oct 2012 May 2013 Nov 2013	Growth index (height for different buds in same season)	Oct 2012 May 2013 Nov 2013	Mar 2014 (for all trees alive, ~3000 samples)	2005 Leaf samples (all clones, 1 to 4 replicates)
Madrid	Dic 2011	Dic 2011	-	-	-	-
Cáceres	Dic 2011	Dic 2011	-	-	-	-
Bordeaux	Nov 2012 Nov 2013 Nov 2014	Nov 2013 Nov 2014 Nov 2015	1-5 scale: Spring 2013 Spring 2014 Spring 2015	-	-	-

Table A1.4. Trait measurements of *Pinus halepensis* provenance trials 1997-2015. In bold, measurements used in this thesis (Chapter 2).

Trial / Trait	Surv	Growth		Reproduction		Wood quality (NIR): Extractives Lignin H/G ratio	Cδ13 Wood
		Height	Diameter	♂ (1-4)	♀ counts		
Valdeolmos (Madrid)	1997						
	1999						
	2000	2004	2006	2004	2003		
	2003	2006	2009	2005	2004	2013	2012
	2006	2009	2010	2010	2005		
	2009	2010			2008		
	2010				2009		
	2011				2010		
Altura (Valencia)	2000	2000	2000			2013	2012
	2003	2003	2003				
	2007	2007	2007	2010	2010		
	2010	2010	2010

CHAPTER 1

HETEROZYGOSITY-FITNESS CORRELATIONS IN MARITIME PINE

ABSTRACT

Heterozygosity-fitness correlations (HFCs) have been used to understand the complex interactions between inbreeding, genetic diversity and evolution. Although frequently reported for decades, evidence for HFCs was often based on underpowered studies or inappropriate methods, and hence their underlying mechanisms are still under debate. Here, we used 6,100 genome-wide Single Nucleotide Polymorphisms (SNPs) to test for general and local effect HFCs in maritime pine (*Pinus pinaster* Ait.), an iconic Mediterranean forest tree. Survival was used as a fitness proxy, and HFCs were assessed at a four-site common garden under contrasting environmental conditions (total of 16,288 trees). We found no significant correlations between genome-wide heterozygosity and fitness at any location, despite variation in inbreeding explaining a substantial proportion of the total variance for survival. However, four SNPs (including two non-synonymous mutations) were involved in significant associations with survival, in particular in the common gardens with higher environmental stress, as shown by a novel heterozygosity-fitness association test at the species-wide level. Fitness effects of SNPs involved in significant HFCs were stable across maritime pine gene pools naturally growing in distinct environments. These results led us to dismiss the general effect hypothesis and suggested a significant role of heterozygosity in specific candidate genes for increasing fitness in maritime pine. Our study highlights the importance of considering the species evolutionary and demographic history and different spatial scales and testing environments when assessing and interpreting heterozygosity-fitness correlations.

INTRODUCTION

It is well known that a decrease in heterozygosity, for example due to mating between relatives or genetic drift after population bottlenecks, can lead to a concomitant reduction in fitness. This is believed to be caused by the expression of recessive deleterious alleles and the lower occurrence of beneficial overdominance effects (Hansson and Westerberg 2002; Waller *et al.* 2008; Charlesworth and Willis 2009; Olano-Marin *et al.* 2011). However, understanding how heterozygosity, as evaluated by molecular markers, correlates with fitness (Heterozygosity-Fitness Correlations or HFCs), at different spatial scales, remains a key issue in the study of the complex interaction between genetics, ecology and evolution (Reed and Frankham 2003; Szulkin *et al.* 2010; Luquet *et al.* 2011). The mechanisms underlying HFCs are still hotly debated, since the hypotheses proposed so far involve alternative causative agents and have different evolutionary consequences (Szulkin *et al.* 2010; Mueller *et al.* 2011).

Three main hypotheses have been proposed to explain HFCs (reviewed in Hansson & Westerberg, 2002; Szulkin *et al.* 2010). The first two explain HFCs in terms of heterozygote advantage at particular loci: (i) the *direct effect* hypothesis, in which a molecular marker itself has a functional influence on a fitness trait; and (ii) the *local effect* hypothesis, which assumes HFCs to be caused by heterozygosity in genes that are in linkage disequilibrium with the studied molecular markers. Finally, (iii) the *general effect* hypothesis assumes that molecular marker heterozygosity reflects genome-wide heterozygosity, which, in turn, is correlated with the individual inbreeding level and the expression of deleterious recessive alleles. Direct, local and general effects are not mutually exclusive (Balloux *et al.* 2004).

Numerous organisms show evidence for either direct/local fitness effects (e.g. cattle (Amos and Acevedo-Whitehouse 2009) and fish (Lieutenant-Gosselin and Bernatchez 2006)), general fitness effects (e.g. amphibians (Lesbarrères *et al.* 2005), or both (e.g. birds (Voegeli *et al.* 2013), humans (Govindaraju *et al.* 2009) and plants, see (Leimu *et al.* 2006) for a review). HFCs for specific loci rather than for genome-wide heterozygosity tend to predominate in the literature. However, it has been recently argued that expected effects are small and hard to detect, and that inappropriate statistical testing (e.g. single-locus regression instead of multilocus models incorporating specific effects for each locus) may have biased the results towards the predominance of the direct and local effect hypotheses (Szulkin *et al.* 2010). Meta-analyses have shown that HFCs, although positive overall, are weak (explaining around 1-5% of the variance of fitness traits in studies with microsatellites), far from universal, and context-dependent (Coltman and Slate 2003; Reed and Frankham 2003; Chapman *et al.* 2009; Szulkin *et al.* 2010). Several authors have also pointed out a

possible publication bias towards reporting positive, significant HFCs (Coltman & Slate, 2003; Chapman *et al.* 2009; Szulkin *et al.* 2010, but see Miller & Coltman, 2014). Furthermore, HFC studies could have also been hindered by the choice of fitness traits and molecular markers, incomplete knowledge on demographic history and population structure for target species and lack of statistical power (Szulkin *et al.* 2010; Miller and Coltman 2014). For example, under current theory, only life-history traits involving many different loci under directional dominance, such as survival or reproductive success, are expected to correlate with heterozygosity, while many studies have focused on less appropriate growth or behavioural traits (Chapman *et al.* 2009; Szulkin *et al.* 2010).

Another highly controversial issue is the degree to which heterozygosity measured at a small number of molecular markers captures whole-genome heterozygosity (Slate *et al.* 2004; Balloux *et al.* 2004; De Woody and De Woody 2005), challenging the support for the general effect hypothesis in studies from the 'pre-genomics era'. Moreover, for marker loci to be able to estimate general inbreeding, the studied population must have gone through some mechanism generating variation in inbreeding, such as consanguineous matings, genetic drift or a recent bottleneck (Szulkin *et al.* 2010; Agudo *et al.* 2012). Rates of both local and genome-wide recombination are also relevant, as they influence the effect of selection on neighbouring neutral sites through linkage (e.g. Corbett-Detig *et al.* 2015) and the long-term maintenance of diversity (Jaramillo-Correa *et al.* 2010). Finally, the type and genome location of molecular markers used to test for HFCs have great importance, as fitness effects are expected to be different depending on whether the markers are neutral or potentially functional, and local or widespread in the genome (Da Silva *et al.* 2009; Olano-Marin *et al.* 2011; Szulkin and David 2011; Agudo *et al.* 2012; Voegeli *et al.* 2013). In this context, genome-wide Single Nucleotide Polymorphisms (SNPs) may help to overcome the inherent limitations traditionally associated to the study of HFCs (Hoffman *et al.* 2014; Miller *et al.* 2014; Miller and Coltman 2014). Here, we used a large SNP dataset (6,100 SNPs), including both putatively neutral control SNPs and potentially functional polymorphisms from candidate genes, to address HFCs in maritime pine (*Pinus pinaster* Ait.), an outcrossing conifer with a fragmented but widespread distribution in south-western Europe and North Africa.

Outcrossing species with long generation times, such as forest trees, can accumulate substantial genetic load (Byers and Waller 1999) and are thus suitable for detecting general HFCs. They also often have high levels of standing genetic variation for adaptive genes (e.g. Grivet *et al.* 2011 for Mediterranean pines), which makes them ideal to identify loci involved in direct or local HFCs too. Moreover, forest trees, given their typically high population differentiation over extensive distribution ranges, can also be useful to study HFCs at different spatial scales. However, to date, only scarce data exist on HFCs for forest tree

species (Leimu *et al.* 2006) and the few available studies were based on only a handful of molecular markers (typically less than 20). These studies reported disparate results, from positive HFCs (Mosseler *et al.* 2003) to negative (Abrahamsson 2011) or null correlations (Savolainen and Hedrick 1995), or even positive and negative HFCs within different populations of the same species (Bush *et al.* 1987).

In this study, individual HFCs were assessed for distinct maritime pine populations and gene pools in a multisite common garden, thus providing an ideal scenario to test whether HFCs remain stable or vary across the species range or under distinct environmental conditions (Leimu *et al.* 2006). HFCs considering distinct selection pressures have seldom been tested or suffered from limitations due to reduced environmental variability or lack of replication (Lesbarrères *et al.* 2005 but see Richardson *et al.* 2014). Using survival as a fitness proxy, we tested for both general effects, correlating neutral SNP heterozygosity and fitness data, and local effects, as revealed by a novel heterozygosity-fitness association test. The interpretation of our results in the light of the relatively well-known evolutionary history of maritime pine and the use of complementary analytical approaches at different spatial scales provided comprehensive insights into the mechanisms underlying HFCs in this outcrossing, long-lived forest tree.

MATERIAL AND METHODS

Study species

Maritime pine (*Pinus pinaster* Ait.) is a long-lived, outcrossing conifer, native to the occidental Mediterranean Basin and the European Atlantic front (Figure 1.1). This species has great ecological and economic importance, being an iconic tree of Mediterranean landscapes as well as the main commercial softwood source in southern Europe. It grows in a variety of substrates, from siliceous to calcareous, in elevations ranging from seashore to over 2,000 m a.s.l., and under different climate regimes (Alía and Martín 2003). In the Mediterranean area, maritime pine is found in scattered populations, from the coast to the Iberian central plateau. In contrast, Atlantic populations, which grow in a moist-temperate climate with low seasonality, are large and continuous, partly due to afforestation.

Maritime pine survived the last glaciations in multiple refugia located in southeastern Spain, northern Africa, southern Italy and the Atlantic coast of Portugal (Burban and Petit 2003), where it presumably overcame severe bottlenecks (see Fig.3 in Grivet *et al.* (2011) and Naydenov *et al.* (2014)). Based on molecular markers, six extant gene pools have been described in the species (French Atlantic coast, Iberian Atlantic coast, central Spain,

southern Spain, Corsica and Morocco (Santos-del-Blanco *et al.* 2012; Jaramillo-Correa *et al.* 2015); see figure 1), which also showed evidence of genetic differentiation for adaptive traits (González-Martínez *et al.* 2002; Santos-del-Blanco *et al.* 2012). High molecular and phenotypic variability among gene pools, together with population adaptive divergence, make maritime pine particularly interesting for HFCs studies at different spatial scales.

Plant material

Open-pollinated seeds were collected from 34 natural populations selected across maritime pine natural range. The sampling included representations of the six gene pools previously described using neutral molecular markers (*see above* and Figure 1.1). Seedlings were grown in nursery and each individual was propagated vegetatively by cuttings (Majada *et al.* 2011). These cuttings (1 to 28 genotypes per population, Table 1.1) were used to establish a common garden experiment (total of 16,288 trees) at four locations that covered the species' climatic breadth (Figure 1.1).

Figure 1.1. Location of the four common gardens (stars) and the 34 source populations used to test for HFCs in *Pinus pinaster*. Populations are represented by numbers, see key in Table 1.1. Symbols represent different gene pools as indicated in the legend. The size of the symbol is proportional to population mean heterozygosity (H_o) for 6,100 SNPs. Shaded grey areas represent the species' natural distribution range.

Common gardens in Asturias and Portugal are located in the Atlantic region, with high annual rainfall (1,160 and 985 mm, respectively) and mild temperatures (11.3 and 14.2 °C, respectively). Nevertheless, Portugal shows less than half the summer rainfall than Asturias, indicating higher seasonality (63 mm and 152 mm, respectively). By contrast, trials in Madrid and Cáceres are located in continental areas under Mediterranean influence (annual rainfall of 443 and 898 mm, and mean temperature of 13.7 and 15.5 °C, respectively), with large seasonal temperature oscillations and a marked summer drought. Apart from its climate, clay soils make Cáceres a worse site than Madrid for maritime pine survival and growth. Each of the four replicated trials comprised 4,072 trees, with 509 genotypes and eight ramets per genotype, set in a randomized complete block design. This clonal common garden design allowed multiple survival measurements for each genotype and accurate performance evaluation.

Molecular markers

Illumina Infinium technology was used to genotype 8,949 SNPs. After standard quality filters (based on genotype clustering scores, SNP call frequency and Hardy-Weinberg equilibrium tests) and careful visual inspection, 6,100 polymorphic SNPs from 3,511 distinct amplicons/unigenes with lower than 10% missing data, were kept for this study. The successfully scored SNP set included 2,661 control SNPs (see Chancerel *et al.* 2013 and *Annex 1*: Plomion *et al.* 2015 for details). These SNPs were randomly chosen from transcriptome sequence data originated from cDNA libraries of different tissues without any prior experimental treatment, and are thus expected to represent neutral or quasi-neutral variants. The remaining 3,439 SNPs were obtained from mixed sources (see Table S3 in *Annex 1*), including potentially functional markers from candidate genes (e.g. 18 SNPs associated with climate adaptation (Jaramillo-Correa *et al.* 2015) and 17 SNPs associated with serotiny, a fire-response adaptive trait (Budde *et al.* 2014)). SNPs were also selected to provide wide genome coverage and low linkage with other SNPs from the same amplicon/unigene. Thus, they provide independent genome-wide samplings of the genome. Further details on SNP selection, assay construction, DNA extraction and SNP typing methods are provided in *Annex 1*.

Fitness and heterozygosity estimates

1. Fitness

Tree survival was scored as a binary variable (0/1) three years after plantation and used as a fitness proxy (David 1998; Szulkin *et al.* 2010). Survival Best Linear Unbiased Predictors (BLUPs) for each population and genotype were estimated by restricted maximum likelihood

(REML), as implemented in ASREML (Gilmour *et al.* 2009). We fitted our data using Generalized Linear Mixed Models, GLMM (logit link function, binomial error structure) according to the formula:

$$Y_{ijk} = \mu + b_k + P_i + C_{j(i)} + \varepsilon_{ij}$$

where Y_{ijk} is the vector of observations at the ramet level (i.e. eight observations per genotype) for survival, coded as either 0, dead, or 1, alive; μ is the overall mean, b_k is the fixed effect of the k^{th} block, P_i is the random effect of the i^{th} population, $C_{j(i)}$ is the random effect of the j^{th} clone within the i^{th} population and ε_{ij} the random error.

We fitted an independent GLMM to data from each of the four common gardens (Asturias, Portugal, Madrid and Cáceres). Overdispersion was accounted for by including a multiplicative overdispersion scaling parameter in the sites where data variance was greater than the variance predicted by the model (Gilmour *et al.* 2009). Pseudoreplication derived from having multiple (eight) observations of the same genotype in each common garden, was accounted for by including genotype as a random factor. Spatial heterogeneity was tested by including a first order autoregressive component in the models (Pinheiro and Bates 2000). This approach is commonly used in field trials with small-scale environmental heterogeneity and consists in correcting a given observed value by comparing it with the value of the previous individual in the spatial grid (Dutkowski *et al.* 2002). However, this procedure was later discarded as it did not improve the model fit (data not shown).

Subsequently, we obtained the Best Linear Unbiased Predictors (BLUPs) for each level of the “genotype” random factor, i.e. one estimate of survival per genotype. This value represents the deviation from the mean survival for each genotype. As such, it represents the genetic merit of each genotype for survival under given specific common garden conditions. This is because the observed phenotypic co-variance among replicates is due to their genetic resemblance, which in the case of ramets of the same genotype equals 1 (Falconer and Mackay 1996; Lynch and Walsh 1998), i.e. they are perfect genetic copies, barring mutations. Environmental variance is then absorbed by the error term in the model.

This way we obtained one survival BLUP for each genotype, subsequently used as the individual fitness estimate for HFCs. These fitness estimates represent the genetic merit of each genotype for survival under the testing conditions of each common garden and have reduced environmental noise, thus increasing the power to detect HFCs. To avoid spurious HFCs due to population structure when merging genotypes from different populations (Szulkin *et al.* 2010), genotype survival estimates did not include population effects (i.e. population mean estimates were subtracted from the genotype survival estimates), as if all trees formed a unique population.

2. Heterozygosity

Individual multilocus observed heterozygosity (H_o) was determined as:

$$H_o = \frac{1}{m} \sum_{j=1}^m H_{ij},$$

where m is the number of loci successfully scored, and H_{ij} is a binary variable that takes the value of 1 if the j^{th} locus is heterozygous within the i^{th} individual, and 0 otherwise (Govindaraju *et al.* 2009). We preferred multilocus observed heterozygosity (H_o) over other standardized measures of marker heterozygosity (e.g. adaptive distance or internal relatedness) because H_o is, in general, more robust (Szulkin *et al.* 2010) (see (Chapman *et al.* 2009) for a discussion). To combine data from different populations for HFC analyses, individual genotype heterozygosity estimates were standardized with respect to the population mean, in a similar way to the survival data (Szulkin *et al.* 2010). An inbreeding coefficient (s) based on $g2$, an estimate of identity disequilibrium (ID), was also computed for each population using RMES software (David *et al.* 2007). The procedure estimates s from $g2$ (the covariance in heterozygosity standardized by average heterozygosity), which depends only on the mean and variance of inbreeding in the population, and not on locus-specific characteristics (for details see David *et al.* 2007).

Heterozygosity-Fitness Correlations (HFCs)

1. General fitness effects

To test for general fitness effects using individual multilocus heterozygosity, the molecular markers assayed must carry information about genome-wide levels of heterozygosity and inbreeding (De Woody and De Woody 2005). Thus, prior to the analysis of general effect HFCs, we used two methods based on ID to assess the ability of the SNP markers used to capture individual inbreeding levels. First, we ran a heterozygosity–heterozygosity correlation test (Balloux *et al.* 2004; Alho *et al.* 2010) with 1,000 randomizations of the marker partition, using *Rhh* package (Alho *et al.* 2010) in R v. 3.1.2 (R Core Team 2015). This test is based in comparing random partitions of the marker set. If independent subsets of loci are able to produce correlated heterozygosity estimates, then we could assume that the marker set has enough power to capture genome-wide information. Second, we computed $g2$, species-wide and for each population and gene pool, running 1,000 iterations in RMES (David *et al.* 2007). The parameter $g2$ is an estimator of ID that uses all markers simultaneously, and is thus expected to be only affected by demographic history and not by the particular set of markers used (Szulkin *et al.* 2010; Miller and Coltman 2014). Then, we assessed the expected power

of our SNP marker set to detect general-effect HFCs using formulae in Miller *et al.* (Miller *et al.* 2014) (equation 5).

To test for the general effect hypothesis, the HFCs significance level was assessed with standard Pearson correlation coefficients (r) between individual heterozygosity based on 6,100 SNPs and survival estimates (BLUPs), after Bonferroni corrections for multiple testing. These coefficients were first computed species-wide considering all individuals (genotypes) as if they were part of a single population (i.e. after subtracting population effects, see *above*). Maritime pine has a strong population genetic structure resulting from its complex evolutionary history and high level of fragmentation (see *Plant material*). Thus, as HFCs significance and sign may vary among distinct geographical groups, HFCs were also studied at different spatial levels: (i) the gene pool level, which also considered individual data without population effects and (ii) the population level, where HFCs were determined considering only the genotypes tested in each population (see Table 1.1 for details on sample sizes). Three populations with five or less genotypes (Madisouka, Carbonero and Cmpeta) were excluded for these later analyses. As phenotypic correlations across common gardens were low (see *Results*), general HFCs were tested separately for each one of the four common gardens. Finally, the expected correlation between heterozygosity and inbreeding (f) and between inbreeding and fitness, as gauged by survival BLUPs, were computed following Szulkin *et al.* (Szulkin *et al.* 2010).

Table 1.1. Heterozygosity and inbreeding for 34 *Pinus pinaster* populations evaluated in four common gardens with contrasted environments (Asturias, Portugal, Madrid and Cáceres; see Figure 1.1). Numbers in brackets after population names denote population location in Figure 1.1. N, number of individuals (genotypes); H_o , observed heterozygosity based on 6,100 SNPs; g_2 , a multilocus estimator of Identity Disequilibrium; $s(g_2)$, inbreeding coefficient for the population estimated from marker data (David *et al.* 2007); $s(g_2)$ values were not computed for the three populations with less than 5 genotypes.

Gene pool	Population	N	H_o	g_2	$s(g_2)$
French Atlantic coast	Hourtin (1)	26	0.274	0.003	0.013
	Leverdon (2)	26	0.260	0.023	0.083
	Mimizan (3)	18	0.269	0.013	0.050
	Olonne sur Mer (4)	23	0.266	0.024	0.086
	Petrocq (5)	22	0.269	0.013	0.048
	Pleucadec (6)	20	0.278	0.004	0.014
	St-Jean des Monts (7)	26	0.266	0.015	0.056
	<i>Overall</i>	161	0.268	0.014	0.051
Iberian Atlantic coast	Alto de la Llama (8)	9	0.245	0.006	0.025
	Armayán (9)	8	0.236	0.001	0.003
	Cadavedo (10)	10	0.215	0.033	0.113
	Castropol (11)	10	0.221	0.014	0.053
	Lamuño (12)	9	0.235	0.001	0.005
	Leiria (13)	23	0.258	0.001	0.005
	Puerto de Vega (14)	8	0.219	0.019	0.069
	San Cipriano (15)	8	0.249	0.006	0.022
	Sergude (16)	21	0.246	0.006	0.023
	Sierra de Barcia (17)	8	0.225	0.022	0.079
	<i>Overall</i>	114	0.239	0.012	0.045
Central Spain	Arenas de S. Pedro (18)	17	0.264	0.002	0.009
	Bayubas de Abajo (19)	18	0.277	0.000	0.000
	Boniches (20)	9	0.267	0.004	0.015
	Carbonero (21)	5	0.259	NA	NA
	Cenicientos (22)	9	0.271	0.005	0.019
	Coca (23)	17	0.270	0.003	0.011
	Cuellar (24)	28	0.270	0.002	0.008
	Olba (25)	21	0.265	0.016	0.058
	Quatretonda (26)	17	0.266	0.021	0.076
	San Leonardo (27)	14	0.267	0.004	0.016
	Valdemaqueda (28)	12	0.269	0.000	0.001
	<i>Overall</i>	167	0.268	0.006	0.021

Table 1.1. (Continued)

Gene pool	Population	N	Ho	g2	s(g2)
Corsica	Pinia (29)	9	0.225	0.002	0.010
	Pineta (30)	12	0.205	0.036	0.121
Overall		21	0.217	0.013	0.049
Southern Spain	Cómpeta (31)	4	0.270	NA	NA
	Oria (32)	26	0.261	0.004	0.014
Overall		30	0.262	0.003	0.013
Morocco	Madisouka (33)	1	0.185	NA	NA
	Tamrabta (34)	15	0.175	0.008	0.032
Overall		16	0.176	0.009	0.035
SPECIES		509	0.256	0.014	0.053

2. Local fitness effects

Local fitness effects were tested using a novel association test between heterozygosity and fitness. As for the case of general HFCs, an independent test was conducted for each common garden. SNPs were coded as 0 if the locus was heterozygous and as -1 otherwise. Heterozygote effects were computed using a Bayesian mixed linear model that imputes missing data and fits simultaneously additive SNP effects and population structure (Quesada *et al.* 2010), as implemented in BAMD software v. 3.5 (Li *et al.* 2012). The kinship matrix was inferred from the full SNP dataset in SpaGeDi v. 1.3 (Hardy and Vekemans 2002). Negative genetic covariances between individuals were set to zero as in Yu *et al.* (2006). To avoid false positives and control for confounding associations between marker heterozygosity and survival, individual assignment probabilities to each gene pool were included in the model as covariates (i.e. the Q matrix, as obtained from STRUCTURE software v. 2.2 (Pritchard *et al.* 2000); details on STRUCTURE runs are provided in Supplementary Information, Table S1.1). A total of 500,000 iterations were performed on BAMD and the last 200,000 were kept to estimate the distribution of additive SNP effects (γ). Then, a one-sided test ($\alpha=0.05$) was used to identify significant local HFCs (i.e. those with $\gamma>0$). Finally, fitness effects of SNPs involved in significant HFCs were also computed separately for trees belonging to the two gene pools with higher sample sizes, French Atlantic Coast ($N=135$) and Central Spain ($N=168$), which represent two highly contrasted growing conditions (Atlantic vs. Mediterranean) within the species range. In this case, the distributions of additive SNP effects were used to identify significant differences (i.e. non-overlapping 95% Confidence Intervals) with SNP effects on fitness at the species-wide scale.

RESULTS

Fitness and heterozygosity estimates

Tree survival ranged from 8% to 95% across test sites (average of 8% in Cáceres, 25% in Madrid, 68% in Portugal and 95% in Asturias; Supplementary Information, Table S1.2), evidencing highly contrasted levels of environmental stress at each site. Site-to-site correlations for individual genotype survival were not significant (see Supplementary Information, Table S1.3). Multilocus observed heterozygosity (H_o) ranged from 0.085 to 0.316, with mean of 0.256 ± 0.033 (Table 1.1). Populations from Mediterranean and Atlantic regions did not show significant differences in heterozygosity, with mean of 0.268 for both the French Atlantic Coast and Central Spain gene pools (Table 1.1). Inbreeding coefficients (s) ranged from 0 to 0.121 (Table 1.1 and Figure S1.1 in Supplementary Information) and neither showed any remarkable differences across the six gene pools or between the Mediterranean and Atlantic regions.

Heterozygosity-Fitness Correlations (HFCs)

The multilocus heterozygosity-heterozygosity mean correlation coefficient was 0.95 (95% CIs: 0.94-0.96). Such highly positive and significant value indicates that our SNP markers are suitable to detect general effect HFCs in this species, if present. Accordingly, the g^2 estimator of ID differed significantly from zero species-wide ($g^2=0.014$, $SD=0.003$, $P<0.001$) and also for the Atlantic and Mediterranean regions separately ($g^2=0.015$, $SD=0.004$, $P<0.001$, and $g^2=0.013$, $SD=0.004$, $P<0.001$, respectively; see Table 1.1 for population estimates). The expected power to detect HFCs according to Miller *et al* [27] was very high ($r^2=0.97$). But correlations between heterozygosity (H_o) and fitness, as evaluated by survival BLUPs, were not significant (after Bonferroni corrections) at any of the spatial levels tested (i.e. species, gene pool or population) in any of the four common gardens (see Supplementary Information, Table S1.4). The same results were obtained when only the 2,661 control SNPs were used (see Supplementary Information, Table S1.5). Among all computed correlations, 55% were positive and 45% were negative, and no particular sign bias was observed. Variation in inbreeding among individuals accounted for 22-51% of the total variance in survival, depending on the common garden. However, heterozygosity was only moderately correlated with inbreeding ($r^2=0.17$; see also Figure S1.1 in Supplementary Information).

Using a heterozygosity-fitness association test, we found that four SNPs (SNP1646, SNP2431, SNP4734 and SNP8781; see Annex 1: Plomion *et al.* 2015 for notation), two of them non-synonymous (SNP1646 and SNP2431), were involved in local effect HFCs (Table

1.2 and Figure 1.2). Interestingly, one of the non-synonymous mutations (SNP2431) was significantly associated with survival in two common gardens (Madrid and Cáceres). Most of the significant positive associations (four out of five, involving three SNPs out of four) were found in the common gardens with higher environmental stress: one in Madrid and three in Cáceres, the common garden with highest mortality. Fitness effects of SNPs involved in HFCs were stable across maritime pine gene pools naturally growing in distinct Atlantic and Mediterranean environments (i.e. they had overlapping 95% Confidence Intervals with species-wide SNP effect estimates on fitness; Table 1.2).

Table 1.2. Additive SNP effects (γ) on fitness and one-sided test for local HFCs ($\gamma > 0$) based on SNP effect distributions obtained using a Bayesian heterozygosity-fitness association test at the rangewide geographical scale. Additive SNP effects and a two-sided test for significant differences with species-wide SNP effect estimates are also shown for two maritime pine gene pools growing under highly contrasted environments (Atlantic versus Mediterranean climate); ns: not significant (i.e. overlapping 95% Confidence Interval).

SNP code ^a	SNP type ^b	Annotation (BLASTx)	Accession number ^c	SNP effects (γ) [P-values]	Regional SNP effects (γ)	
					French Atlantic Coast	Central Spain
Portugal trial						
SNP1646	[A/G] ^{nonsyn}	V-type proton ATPase	501311935	0.04144 [0.0053]	0.02918 ns	0.04673 ns
Madrid trial						
SNP2431	[T/G] ^{nonsyn}	Expansin 2	1485845574	0.05451 [0.0125]	0.04153 ns	0.02739 ns
Cáceres trial						
SNP2431	[T/G] ^{nonsyn}	Expansin 2	1485845574	0.04435 [0.0342]	0.03969 ns	0.02922 ns
SNP4734	[A/C] ^{NA}	Serine-type endopeptidase	501314366	0.04923 [0.0071]	0.03989 ns	0.01850 ns
SNP8781	[C/G] ^{NA}	GPCR-type G protein 1	1516987379	0.05752 [0.0020]	0.05905 ns	0.04158 ns

^aSee notation in *Annex 1: Plomion et al. 2015*.

^bAbbreviations: *nonsyn*, nonsynonymous; *NA*, unknown/data unavailable.

^cAs provided by NCBI's dbSNP.

Figure 1.2. Box-plots for significant local HFCs at the species-wide geographical scale. A: Common gardens under mild Atlantic environmental conditions (Portugal, there were not any significant HFC in Asturias), B: Common gardens under harsh Mediterranean environmental conditions (Madrid and Cáceres). Boxes denote the interquartile range and horizontal lines within boxes the phenotypic means (survival BLUPs).

DISCUSSION

HFCs were examined in the keystone tree maritime pine, under contrasted environments, using 6,100 genome-wide SNP markers. Variation in inbreeding explained a substantial proportion of the total variance in survival, but no significant correlation was found between genome-wide heterozygosity and fitness at any spatial level (i.e. species, gene pool or population). However, heterozygosity in four SNPs was positively associated with survival, in particular in the common gardens with higher environmental stress, as shown by an association test. Local HFCs were stable across maritime pine gene pools naturally growing under Mediterranean and Atlantic climates. These results led us to dismiss a global relationship between heterozygosity and fitness in maritime pine (i.e. general effects). Alternatively, we suggest that not enough variance in inbreeding depression and demographic history may have decoupled heterozygosity and inbreeding.

Lack of general fitness effects

Previous studies in conifers found no clear HFC patterns (Bush *et al.* 1987; Savolainen and Hedrick 1995), but their results may have been swayed by a lack of statistical power, due to a reduced number of markers, inappropriate analytical methods, and the confounding effects of environment and population structure (Szulkin *et al.* 2010; Fox and Reed 2011). It can even be argued that previous HFC studies in forest trees were all unwittingly looking for local effects, as they used a very small number of markers that were poor estimators of genome-wide heterozygosity (Balloux *et al.* 2004). Our multisite clonal common garden design –with several hundreds of individuals from across the species natural range, planted in contrasted environments–, our use of a high number of SNP markers with wide heterozygosity range, and the performance of analyses at different spatial levels, accounting for population structure, should have overcome previous drawbacks, maximizing the chance to detect HFCs in the species, if present (David 1998; Waller *et al.* 2008; Szulkin *et al.* 2010). The expected power to detect HFCs (Miller *et al.* 2014) was very high and this is one of the first studies reporting significant g^2 values (but see García-Navas *et al.* 2014) and a high heterozygosity-heterozygosity correlation (Grueber *et al.* 2011; Miller and Coltman 2014). These results endorse the greater ability of a high number of SNP markers (in our case 6,100), despite incomplete coverage of the genome, for the estimation of genome-wide heterozygosity, as recently suggested by Hoffman *et al.* (2014).

Common gardens covered a wide range of environments from very harsh in Cáceres to very mild in Asturias, as evidenced by survival rates ranging 8-95%, which allowed us to assess HFCs in highly contrasted growing conditions. The correlation between individual heterozygosity and fitness is expected to be particularly strong under high stress conditions (Richardson *et al.* 2004; Armbruster & Reed, 2005; Lesbarrères *et al.* 2005 but see Waller *et al.* 2008 and Da Silva *et al.* 2009). This is because the deleterious effects of inbreeding (i.e. inbreeding depression) tend to increase under environmental stressful conditions, as the increased variance in fitness in the inbred population allows natural selection to act against the least fit genotypes, reducing the cumulative loss of fitness predicted to occur in more benign environments (Armbruster and Reed 2005). However, we did not find significant correlations between genome-wide multilocus heterozygosity and fitness under any testing environment, including those with low survival.

This lack of correlation could be explained by maritime pine mating system and evolutionary history. Forest trees, and maritime pine in particular, are mostly outcrossing, with little variation in outcrossing rates across populations, and show low levels of correlated paternity (De Lucas *et al.* 2008) and high heterozygosity (Jaramillo-Correa *et al.* 2015). As a mainly

outbreeding species, maritime pine may not have enough variance in inbreeding as to result in significant general effect HFCs. However, when compared to other *Pinus* species, maritime pine has reduced inbreeding depression (although still significant), in particular for fecundity traits and general vigour (Durel *et al.* 1996). Evidence of historical demographic bottlenecks in the species (Grivet *et al.* 2011; Naydenov *et al.* 2014) suggests that some genetic load may have been purged during maritime pine survival in isolated glacial refugia (Byers and Waller 1999; Richardson *et al.* 2004; Bucci *et al.* 2007; Charlesworth and Willis 2009), contributing to the observed lack of general effect HFCs (Byers and Waller 1999; Reed and Frankham 2001, 2003; Chapman *et al.* 2009). This hypothesis is supported by a relatively low correlation between heterozygosity and inbreeding ($r^2=0.17$), which renders HFCs non-significant despite a substantial proportion of the total variance in survival being explained by variation in inbreeding (22 to 51%, depending on common garden). Similar low or non-significant HFCs despite relatively high association between inbreeding and fitness have been reported for both plants and animals by Szulkin *et al.* (2010).

Agricultural hybrids where two homozygous inbred lines are crossed to produce heterozygote F1 individuals with outstanding performance (i.e. heterosis), can be considered an extreme case of general HFCs, e.g. in maize or rice (Tester and Langridge 2010). However, naturally occurring differences in individual heterozygosity are normally much lower and, as we showed in this study, did not lead to significant fitness differences in maritime pine. To our knowledge, efforts to produce high-performance forest trees by crossing double haploid lines in breeding programmes have been generally unsuccessful (Andersen 2005).

Significant local fitness effects

Heterozygosity at four SNP markers was positively associated with fitness, in particular in the two common gardens with higher mortality (Madrid and Cáceres), suggesting that heterozygosity at specific loci can confer selective advantage to maritime pine under stressful environmental conditions. Apart from including a large set of control SNPs, our SNP assay was enriched for polymorphisms from candidate genes, which could have fostered the identification of local HFCs, despite the huge size of the pine genome (~20-25 Gbp). These local HFCs were stable across maritime pine gene pools naturally growing under Atlantic and Mediterranean climates, suggesting broad, species-wide, adaptation processes. Individuals with increased heterozygosity at candidate genes may well possess the necessary diversity of alleles required to adequately cope with environmental stochasticity (Chapman *et al.* 2009). Changing selection pressures (i.e. drought, extreme temperatures) could favour one allele or the other, alternatively, along the extensive reproductive period and long life of trees

(e.g. wet or dry years or decades). It has been recently argued that heterozygote advantage should be very common during adaptation if selection is stabilizing and at least some mutations are large enough to overshoot the optimum (Sellis *et al.* 2011). Although there is scarce experimental proof of particular loci showing heterozygote advantage (Hedrick 2012), it has been suggested that single alleles can confer general stress-resistance for some types of stress (Armbruster and Reed 2005; Amos and Acevedo-Whitehouse 2009). For example, heterozygote advantage might include some important cases of recent adaptation under strong selection (e.g. mutants in livestock and companion animals, warfarin resistance in rats; see Hedrick, 2012 for discussion and more examples). However, it is generally not possible to differentiate between heterozygote advantage and other types of balancing selection (Hedrick 2012).

Additive SNP effects in significant local HFCs were modest, ranging from 0.041 to 0.058 (Table 1.2), as it would be expected for highly polygenic fitness traits such as survival. Integrated life history traits (e.g. survival, reproductive success) involve many different loci, all of which are targets for deleterious recessive mutations. This genetic architecture favours the expression of HFCs (Szulkin *et al.* 2010). However, the expected effect sizes for those traits are very small, as they have a more complex polygenic architecture than morphological or physiological traits (Merilä and Sheldon 1999; Chapman *et al.* 2009). Hence, it is not surprising that previous studies based on microsatellites and invoking local effect HFCs explained very little variance of the studied traits (Chapman *et al.* 2009; Szulkin *et al.* 2010). Recently, García-Navas *et al.* (2014) found heterozygosity at a single microsatellite locus to explain a disproportionately large variation in offspring size in great tits (37%). Although they suggest a possible heterozygote advantage for this locus, they also argue that this result could be biased by an underlying general inbreeding effect. It is interesting to note, however, the much stronger mean effect sizes found for plants than for animals (Chapman *et al.* 2009). Anyhow, because purging of genetic load and decay of linkage disequilibrium are slow processes, effect sizes tend to be also smaller in populations that underwent historical bottlenecks, as is the case of maritime pine, than in those that suffered more recent ones (Chapman *et al.* 2009).

Remarkably, a non-synonymous polymorphism, SNP2431, was involved in HFCs for the two Mediterranean common gardens with high environmental stress (Madrid and Cáceres). This SNP is located in a candidate gene encoding an expansin, which belongs to a family of closely related plant cell wall proteins involved in cell growth (Li *et al.* 2003; Cosgrove 2005). The heterozygote advantage observed at this locus may then be explained by direct selection, as it appears to play an important functional role in plant growth and development (Muller *et al.* 2007). Some previous studies have also reported significant HFCs for markers

located in genes of known function but not for neutral markers (Luikart *et al.* 2008; Da Silva *et al.* 2009). However, most of the local HFCs found in this study were located in poorly annotated genes and it is likely that they are in linkage with causative polymorphisms rather than being the causative polymorphisms themselves (i.e. they can be considered 'local' but not 'direct' HFCs). Thus, further work is needed to validate or identify causative polymorphisms and fully understand the functional basis of the significant HFCs found in our study.

CHAPTER 2

INSIGHTS INTO THE GENETICS OF INTEGRATED ADAPTIVE TRAITS IN ALEPPO PINE

ABSTRACT

Unravelling adaptive processes under varying environments is critical to foretell the fate of forests in a climate change scenario. A keystone to achieve this aim is the availability and combined analysis of phenotypic and genetic data. This will help to shed light on the genetic architecture of adaptive traits, which remains poorly understood in non-model tree species. In this study we address phenotypic standing variation and look for putative underlying loci in key traits related to fitness in Aleppo pine. Growth, drought resistance, wood extractives, lignin content and reproduction were evaluated in a common garden including range-wide populations (total of 684 trees) to study trait-environment correlations and trade-offs among traits. We found significant correlations between growth, drought resistance and reproductive traits and various ecogeographical variables, supporting the hypothesis of local adaptation. We also found that allocation to adaptive traits related to growth, reproduction and defence traded-off against each other. The same trees were genotyped and 294 polymorphic SNPs were used for single-locus and multi-locus genotype-phenotype association analysis. We identified 15 SNPs significantly associated with growth traits, wood extractives, lignin content or flowering, explaining a moderate variation in phenotypic traits (<11%). Using integrated phenotypes from multivariate analysis, we found one SNP in gene *XYL1* associated with all wood extractive traits. Multi-locus association tests identified some of the previous significant associations from single-locus tests, and new SNPs showing significant associations with multiple correlated traits. This could be interpreted as a signal of pleiotropy, but can also be entangled by phenotypic correlations. This study highlights the utmost importance of accounting for possible correlations and trade-offs among traits, and the advantages of considering integrated phenotypes, when association studies for multiple adaptive traits are addressed.

INTRODUCTION

Forest persistence under climate change depends on future environmental conditions and the ability of forest trees to cope with them. Sessile long-lived trees will be able to adapt to upcoming new environments depending on the species' phenotypic plasticity and standing genetic variation, which will determine their response to selection pressures. The relationship between selection acting at the phenotypic level and the genetic variation underlying adaptive phenotypes has long been a major focus in evolutionary genetics. Environmental forces enable natural selection, which acts at the phenotypic level as a whole. To the extent that phenotypic variation comes as the result of genetic variation, evolutionary change and local adaptation will take place. Thus, the evolutionary response to selection will depend ultimately on genetic variation, conditioned also on how that variation is structured within and among traits, i.e. on the genetic architecture of adaptive traits.

In this framework, it is of key importance to disentangle genetic control from environmental effects and to study the relationship and possible trade-offs among adaptive traits. Phenotypic traits do not vary independently, but instead reflect webs of developmental, physiological and functional interactions of varying strengths (Armbruster *et al.* 2014). Phenotypic integration has particularly commanded the interest of modern evolutionary biologists because of its potential to constrain the course of phenotypic evolution. The concept of integrated phenotype has acquired a very broad usage, covering observed patterns of covariation, the capacity or tendency for covariation, the underlying organismal architecture that gives rise to the tendencies and, ultimately, the observed patterns and their evolutionary causes and consequences. Here, we will just refer to statistical integration, understood as patterns of strong phenotypic or genetic correlations in standing population variation. Unravelling phenotypic from genetic correlations is fundamental, as phenotypic trade-offs between traits do not necessarily, or even usually, lead to negative genetic correlations (Houle 1991; Roff and Fairbairn 2007; Armbruster *et al.* 2014). If there are strong negative phenotypic correlations between two traits that have however no genetic basis, even under strong selection pressures, no evolutionary response is possible (Stearns 1989).

Many studies have described extensive genetic variation in life-history traits in forest tree species, likely reflecting adaptation to varying environmental conditions along large distribution ranges (Alberto *et al.* 2013). Within species, the study of populations that have evolved under a range of environments enables the detection of two important adaptive patterns. First, trade-offs, emerging as strong correlations, mainly among growth and other life-history traits (Loehle 1988; Obeso 2002; Santos-del-Blanco *et al.* 2012; Grivet *et al.* 2013). And second, ecotypic trends or clinal variation, i.e. continuous genetically-based phenotypic variation along environmental gradients (Santos-del-Blanco *et al.* 2013; De La

Mata *et al.* 2014; Gauli *et al.* 2015). The combined study of growth, wood quality, stress response and reproductive trait phenotypes at the individual and population level can shed light on the complex patterns of adaptation in forest trees. In addition, the genetic architecture of adaptive traits reflects the evolutionary history of populations and also shapes the divergence among them (Comeault *et al.* 2014). It can also help to address outstanding questions such as how many loci are involved in adaptation, and what is the relative contribution of major and minor effect alleles to adaptive evolution (Stinchcombe and Hoekstra 2008). However, despite their central role in evolution, the genetic basis of adaptive traits is still widely unknown (Savolainen *et al.* 2013).

In the last decades, the identification of polymorphisms responsible for phenotypic variation has been one of the major goals of population and quantitative genetics (Stinchcombe and Hoekstra 2008; Wegrzyn *et al.* 2010). Given the forecasted future climate conditions, it becomes of key importance to identify genes causing natural variation in traits that affect fitness. Genotype-phenotype association studies have proved to be an effective tool to identify single loci underlying important traits in multiple species, including conifers (e.g. González-Martínez *et al.* 2007, 2008; Eckert *et al.* 2009; Budde *et al.* 2014; see Neale and Savolainen, (2004) and Grattapaglia *et al.* (2009) for reviews). Association studies typically look for simple, marginal effects of SNPs on phenotypic variation (Marjoram *et al.* 2014). Strong associations between a SNP and the phenotype are interpreted as indicating that the SNP, or a nearby SNP in linkage disequilibrium (LD), likely affects the phenotype. The choice between whole-genome approaches (GWAS) and a candidate gene-based approach is usually based on the extent of LD, as this determines the mapping resolution that can be achieved and also the number of markers needed to ensure an adequate coverage of the genome (Nordborg and Tavaré 2002). The candidate-gene association approach is particularly useful in species with a rapid LD decay (Neale and Savolainen 2004; González-Martínez *et al.* 2006) and large genomes (~30 Gbp in pines; Ahuja and Neale 2005; De La Torre *et al.* 2014), such as conifers.

Candidate-gene association mapping relates polymorphisms in selected candidate genes that have supposed roles in the control of phenotypic variation in specific traits to observed phenotypes (Zhu *et al.* 2008). Candidate genes are selected based on prior knowledge from model species, mutational analysis, biochemical pathways or linkage analysis of the traits of interest. Then, Linear Mixed Models (LMM) can be used as a powerful approach attributing effects to single-loci. The underlying LMM used to detect statistical associations has remained unchanged since Yu *et al.* (2006), although many approaches have been developed afterwards to reduce the computational burden (e.g. EMMA, Kang *et al.* 2008; EMMAX, Kang *et al.* 2010). Two main confounding factors can make false positives arise:

population/family structure and LD (Neale and Savolainen 2004; Lipka *et al.* 2015). However, statistical procedures such as including in the model the kinship matrix inferred from molecular markers and covariates from STRUCTURE analysis (a Bayesian clustering approach to quantify population structure), and the application of multiple testing correction procedures, have proved effective to reduce false positives (Balding 2006; Lipka *et al.* 2015).

The primary rationale for association studies is the idea that, by examining the associated SNPs in more detail afterwards—for example, examining to which genes do they belong or lay close to—we may glean important insights into the biology of the phenotype under study. However, the proportion of variance explained by single loci is usually very low, and so has been the case for forest trees (e.g. González-Martínez *et al.* 2007, 2008; Holliday *et al.* 2010; Cumbie *et al.* 2011; Prunier *et al.* 2013). Lately, new methods considering the interaction between markers have been developed (Guan and Stephens 2011; Segura *et al.* 2012). The goal is to extract more information from signals that exist in the data, particularly marginal associations, than do standard single-SNP analyses (Guan and Stephens 2011). Multi-SNP approaches, such as Bayesian Variable Selection Regression (BVSr), generally outperform single-SNP approaches in detecting true causal variants even in the absence of interactions between variants (Guan and Stephens, 2011; Ehret *et al.* 2012).

Forest trees are often seen as complex organisms for genetic studies, but they exhibit compelling advantages for genotype-phenotype association studies: (i) adequate levels of nucleotide diversity for SNP markers, (ii) rapid decay of LD, (iii) large and random-mating populations with limited population structure and (iv) access to large clonal or family-based genetic tests for precise evaluation of phenotypes (Neale and Ingvarsson 2008). In a broader sense, the fact that they are often found in large natural populations (connected by extensive gene flow) that harbour a wealth of genetic diversity, makes long-lived trees ideal for disentangling selective from stochastic evolutionary forces. Additionally, they often have large geographical distributions covering different environmental gradients, enabling the study of local adaptation processes, and further possibilities of tree populations to adapt to changing environments (Petit and Hampe 2006; Neale and Kremer 2011). Aleppo pine (*Pinus halepensis* Mill.) is a long-lived conifer, widely distributed across the Mediterranean Basin, where it can be found across environments encompassing a broad range of moisture and thermal conditions. It has also remarkable capacity to cope with dry and low-fertility environments (Voltas *et al.* 2008; Santos-del-Blanco *et al.* 2013). Previous studies suggest that Aleppo pine populations expanded westwards all around the Mediterranean basin from some refugia located in the southern part of the Balkan Peninsula (Grivet *et al.* 2009). Within the current distribution of the species, three large gene pools have been differentiated using

microsatellites: Western (including Spain), North African (including Tunisia) and Eastern (including Greece and eastern Italy; Grivet *et al.* 2009, 2011).

Here we investigate the standing phenotypic variability across the Aleppo pine natural range for 14 adaptive traits. The studied traits include defense related, stress response, growth and reproductive traits, enabling an integrated study of tree functions related to adaptation. Namely, the traits addressed in the study were wood chemistry (dichloromethane, ethanol and water extractives, total extractives, Klason and total lignin content and HG ratio), water use efficiency (evaluated by carbon isotopes (C δ 13), total carbon content), height, stem diameter, volume, and female and male flowering. Many traits of interest in forest trees, such as the ones described above, are complex in nature and occur late in development (Groover 2007). Wood property traits combine fundamental roles in the evolution of plants, plant growth, and resistance to environmental stress in nature with practical importance in wood and pulp production (González-Martínez *et al.* 2007). Intrinsic water-use efficiency (the amount of carbon acquired per unit of water lost) is another component of adaptation that is specially relevant to plant capacity to thrive in dry areas. As reviewed by (Warren *et al.* 2001), C δ 13 in wood samples is a useful indicator of water availability for conifers in seasonally dry climates. Growth and flowering are also widely recognized key traits related to fitness in forest trees (e.g. Santos-del-Blanco *et al.* 2015 for maritime pine). Phenotypic variation and its genetic basis for growth, wood quality and drought resistance traits have been broadly studied in pines (e.g. González-Martínez *et al.* 2007, 2008; Correia *et al.* 2008; Cumbie *et al.* 2011; Marguerit *et al.* 2014). However, to our knowledge this is the first study considering jointly such a large number of adaptive traits, including also reproductive traits. First, we quantified correlations among traits and between traits and environmental conditions found in the areas where populations originated. This allowed us to identify trade-offs and ecotypic trends in the species, within and among populations. The same trees were genotyped using a 384-plex SNP assay. Then, we used the phenotypic measurements of the 14 traits and successfully genotyped SNPs to conduct single-locus and multi-locus association tests, looking for loci underlying the traits under study. Integrating basic quantitative genetics and association results allowed us to disentangle the genetic basis of the studied adaptive traits in the species.

MATERIALS AND METHODS

Plant material and phenotypes

Two provenance common garden experiments were phenotyped for this study (Figure 2.1). Both common gardens have a complete block design, with trees in groups of 4 trees of the

same provenance within blocks. The kinship among trees of the same provenance remains unknown (i.e. trees from the same provenance are assumed to be unrelated).

Figure 2.1. *Pinus halepensis* provenance trials located in Castellón (A) and Madrid (B), Spain. Both common gardens constitute identical copies.

We measured 14 traits related to wood quality, drought resistance, growth and reproduction (Table 2.1) in 684 trees from 28 provenances, covering the species natural distribution range (Figure 2.2). The studied provenances include populations from the three large gene pools previously described in the species (Grivet *et al.* 2009, 2011): Western, Eastern and North African. Growth and reproductive traits were measured in 2010 in 13 year-old trees. Wood samples for wood quality and isotope analysis were collected in 2012. Phenotyping methods are described in the general *materials and methods* section.

Table 2.1. Traits measured in the *Pinus halepensis* provenance trials.

Trait group	Trait	Description	Measurement	Method
Wood quality	Dichloromethane	Dichloromethane extractives	Indirect	NIR
	Ethanol	Ethanol extractives	Indirect	NIR
	Water	Water extractives	Indirect	NIR
	Total extract	Total extractives	Indirect	NIR
	K Lignin	Klason lignin content	Indirect	NIR
	Total lignin	Total lignin content	Indirect	NIR
	H/G ratio	Ratio of p. hydroxyphenylpropane (H) and gaiacyl (G)	Indirect	NIR
Water use efficiency	Carbon δ^{13}	Carbon delta 13 content	Indirect	Isotope analysis
	Total Carbon	Total C content	Indirect	Isotope analysis
Growth	Height	Total height	Direct	Pole
	Diameter	Normal stem diameter	Direct	Caliper
	Stem volume	Derived from height and diameter	Indirect	$V=1/12\pi \cdot h \cdot d^2$
Reproduction	Female flowering	Female cones	Direct	Cone count
	Male flowering	Male cones	Direct	Qualitative scale

Figure 2.2. Distribution map of *Pinus halepensis* source populations (circles) and common gardens (triangles). Green areas indicate the species' natural distribution range.

Quantitative genetics

Phenotypic data were retrieved from two different trials (*see above*). To be able to mix phenotypic data from both sites we had to correct field measurements to minimize the environmental site effects. To do so, we fitted a MLM independently for each trait using 'lme4' package (Bates *et al.* 2015) in R (R Core Team 2015), with site and block as fixed factors and provenance as random factor. For male reproduction, which was measured in a 1 to 4 scale, we fitted a GLM (Poisson family, log link function). Then, we tested the significance of site, including just site as fixed effect, and the significance of block, including just block as fixed effect. If site and block were significant, we subtracted the fixed effect of the corresponding level to the individual field measurement of each individual, to remove the environmental effect as much as possible. We tested for spatial autocorrelation following (Dutkowski *et al.* 2006), but finally discarded the autoregressive components, as the procedure did not improve the model fit. This way we were able to get corrected values for all traits, for each individual and population, to be further used in quantitative analysis and genotype-phenotype association studies.

To get an estimate of the genotype by environment interaction (GxE) in the studied traits we fitted a new MLM for each trait, with block as random factor and the interaction between site and provenance as a fixed factor. We considered the *p*-value of the interaction (Satterthwaite approximation for degrees of freedom) as an indicator of GxE.

Trait trade-offs

Using population corrected means (BLUPs) derived from previous models, we tested for possible correlations and trade-offs among traits. First, we computed Pearson's correlation coefficient for each pair of traits, at the individual tree and population levels. These correlations can be considered "raw" genetic correlations, but running bivariate models is highly recommended to unravel the environmental effect too (see Hadfield *et al.* 2010 for discussion). Phenotypic, genetic (broad sense) and environmental correlations among traits were then jointly estimated through bivariate models in *MCMCglmm* (Hadfield 2010). A model was fitted for each pair of traits. MCMC chains were run 550,000 times, sampling every 500 and with a burn in period of 50,000 iterations, yielding a total of 1,000 estimates per parameter per model.

To have a more comprehensive and graphical view of trait trade-offs we also ran a Principal Component Analysis (PCA) both for individual tree data and for population means of the 14 traits. We used the parallel analysis implemented in the *psych* package (Revelle 2011) in R to decide the number of principal components to retain. We selected only those principal components with eigenvalues for observed data larger than those obtained in simulations (Hayton *et al.* 2004). Then, we ran the PCA based on the correlation matrix with the retained number of components and *varimax* rotation.

Trait-environment correlations

We considered six climatic variables: annual precipitation, precipitation during the warmest quarter (June, July, August), annual mean temperature, mean temperature of the warmest month, mean temperature of the coldest month, and a continental index (difference between mean temperature of the warmest month and mean temperature of the coldest month). Climatic data for each source population was obtained from the *WorldClim* database (Hijmans *et al.* 2005). Three spatial variables were also recorded for each population: longitude, latitude and altitude. We computed graphical pairwise correlations among these 9 spatial and environmental variables.

We also computed correlations between provenance adjusted means (BLUPs) for each trait and the 9 variables described. Additionally, we carried out a principal component analysis (PCA) to allow a more synthetic interpretation, reducing the number of environmental variables. The analysis was implemented by the *psych* package (Revelle 2011) in R (R Core Team 2015). To decide the number of principal components to retain we ran a parallel analysis (implemented in the *psych* package; Revelle, 2011) and afterwards ran a PCA with

the retained number of components and varimax rotation. The PCA loadings were subsequently used in correlation analysis with plant traits.

Molecular markers and population structure

A total of 685 trees were genotyped using a 384-plex SNP assay. The assay included polymorphisms from functional candidate genes related to stress responses, wood properties, phenology and growth, as well as expressional candidate genes for a wide range of stress responses and those differentiating trees with marked phenotypic differences in fire adaptation traits (for details on genotyping and the SNP assay see Pinosio *et al.* 2014).

Based on the scored polymorphic loci, we computed Wright's F-statistics (F_{ST} , F_{IS} , F_{IT}) in the 28 studied populations using SpaGeDi 1.3 (Hardy and Vekemans 2002). Then, we used the Bayesian clustering method implemented in STRUCTURE software (Pritchard *et al.* 2000) to perform 10 independent runs for $K=7$ (the known number of genetic groups in the species, unpublished results, personal communication from MJ Serra-Varela and D Grivet), with a burning period of 100,000 steps, followed by 10^6 MCMC replicates. Resulting admixture coefficients were averaged across replicated runs using the software CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007). These averaged coefficients (*Q-values*) were further used as covariates in genotype-phenotype association tests.

Association tests

For each phenotypic trait, we conducted a single-locus genotype-phenotype association test. A MLM was fitted independently for each marker following Yu *et al.* (2006), as implemented in Tassel 3.0 (Bradbury *et al.* 2007). The kinship matrix was inferred in SpaGeDi 1.3 (Hardy and Vekemans 2002), using (Loiselle *et al.* 1995) kinship coefficient. Negative genetic covariances between individuals were set to 0 as in Yu *et al.* (2006). To avoid false positives and control for population structure, individual assignment probabilities to each population cluster were included in the model as covariates (i.e. the Q matrix, as obtained from STRUCTURE software). Corrections for multiple testing were performed using the false discovery rate (FDR) method (Storey and Tibshirani 2003), as implemented in R *qvalue* package (Dabney and Storey 2012).

Additionally, we performed an association test using as phenotypes the PCA scores of the three most important PCs from the multi-trait PCA (see *above*). The joint analysis of multiple related phenotypes has recently become popular in association studies because it can increase statistical power to detect target loci (Aschard *et al.* 2014). One of the proposed methods is to use PCA to transform all phenotype data into a single or few new phenotypes,

that are linear combinations of the original phenotypes (Yang and Wang 2012; Suo *et al.* 2013). We used these integrated phenotypes to perform standard univariate regression at each SNP marker as described above.

We also carried out multilocus genotype–phenotype association tests for each trait using BVSR analysis, as implemented in the computer software *piMASS* 0.9 (Guan and Stephens 2011). BVSR models applied to genotype–phenotype association studies use the phenotype as the regression response and genetic variants as regression covariates to evaluate variants or genomic regions that are strongly associated with a particular phenotype (Guan and Stephens 2011). BVSR implements a linear model:

$$y \mid \gamma, \mu, \tau, \beta, X \sim N_n(\mu + X_\gamma \beta_\gamma, \tau^{-1} I_n)$$

where y and X are the observed data, with y denoting the phenotypic scores of individuals and X containing the genotypes of individuals at all loci. The vector γ is a set of binary variables that indicate which SNPs have nonzero values for β ; μ is the intercept, and τ the precision (inverse of the variance) of the residual errors. *piMASS* evaluates multiple predictor models, which incorporate different subsets of genetic variants and produces a model average of parameter estimates. SNPs statistically associated with phenotypic variation are identified by the posterior distribution of γ . The model contains additional parameters describing higher level aspects of the genetic architecture of the trait. These include the phenotypic effect associated with each SNP (β), the proportion of variance explained by the SNPs (PVE), the conditional prior probability of a SNP being included in the model (p_{SNP}), the number of SNPs in the regression model (N_{SNP}), and the average phenotypic effect associated with a SNP that is in the model (σ_{AVE}). We normal quantile–transformed all our phenotypic trait data before the analyses, as suggested by Guan and Stephens (2011). This previous transformation standardizes phenotypic traits to have a mean of 0 and variance of 1 and facilitates comparisons of the phenotypic effect of SNPs across traits. For each trait, we ran an independent chain in *piMASS*, each with a burn-in of 10,000 steps followed by 100,000 steps with sampling every tenth step.

RESULTS

Quantitative genetics

MLMs with site as fixed factor and provenance as random factor, showed that site and block were not significant for total extract and HG ratio. For the remaining traits, the fixed effect for each block was subtracted from the original field measurement, to get corrected estimates and be able to mix data from the two trial sites. Tree provenance explained a substantial

amount of trait variation in C δ 13 (variance explained of 26.6%), growth (23.9% for height, 19.5% for diameter and 25.8% for stem volume) and reproductive traits (22.1% of variance explained in female flowering). By contrast, tree provenance explained only little of wood quality traits variation (0 to 7.1%).

Trait mean values at the species level are provided in Table 2.2 and Figure 2.3. For detailed means and boxplots at the population level see Figures S2.1, S2.2 and S2.3. We found no significant differences among populations for dichloromethane, ethanol and water extractives or other wood quality traits (Figure 2.3). The same occurred with total carbon content. However, populations from Greece and Italy and even more the two populations from Tunisia showed a significantly lower (more negative) C δ 13 content (Figure 2.3). Trees from Greek populations grew higher and had thicker stems, resulting in significantly higher stem volumes. Populations from the three large gene pools showed significant differences for all growth traits. Regarding reproduction, populations from Tunisia were significantly different from the others, exhibiting a higher number of female cones.

Figure 2.3. Boxplots for 12 traits measured in 28 *P. pinaster* provenances, pooled by gene pool. Each black dot represents a mean value for a given provenance. The red dotted line corresponds to the trait mean estimate across populations.

MLMs including the site by provenance interaction as fixed effect, revealed high plastic effects (site effects) for all traits except for total extractives, HG ratio and total carbon content. We also found that provenance was a non-significant factor explaining lignin content and HG ratio (Table 2.2). The same occurred with total carbon content. The genotype by environment interaction was highly significant for all growth and reproductive traits, Cδ13 and wood extractives, but not significant for the remaining wood-quality traits (Table 2.2).

Table 2.2. Mean values and SE for the 14 traits studied in *Pinus halepensis* and factor significance based on MLMs for each trait, including site, provenance and site x provenance (GxE) as fixed effects. Significance: *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$.

Trait group	Trait	Units	Mean	SE	Site effect	Prov. effect	GxE
Wood quality	Dichloromethane	%	2.05	0.02	2.2E-16***	0.000178***	3.21E-07***
	Ethanol	%	1.77	0.01	2.2E-16***	1.04E-05***	1.23E-06***
	Water	%	1.75	0.01	6.34E-08***	4.23E-05***	0.02471*
	Total extract	%	5.36	0.03	0.1607	0.000854***	0.000187***
	K Lignin	%	25.98	0.04	2.85E-12***	0.10563	0.010959*
	Total lignin	%	26.46	0.04	2.89E-12***	0.105375	0.010998*
	H/G	ratio	0.08	0.00	0.3179	0.3988	0.106
Water use efficiency	Carbon δ13	‰	25.08	0.02	2.2E-16***	2.20E-16***	8.28E-05***
	Total Carbon	µg	1044	10.03	0.000153	0.8663	0.4635
Growth	Height	cm	522.76	3.55	2.2E-16***	2.20E-16***	2.91E-07***
	Diameter	mm	90.86	0.97	2.2E-16***	2.20E-16***	0.002772**
	Stem volume	dm ³	13.99	0.29	2.2E-16***	2.20E-16***	2.28E-12***
Reproduction	Female flowering	n° cones	24.50	0.49	2.2E-16***	1.05E-13***	3.16E-10***
	Male flowering	1-4 scale	0.85	0.02	2.2E-16***	2.20E-16***	0.000546***

Trait trade-offs

At the individual level (all individual tree data pooled together), we found significant correlations among traits of the same functional group, except for total carbon and Cδ13 content (Figure 2.4). K-lignin and total lignin content exhibited a perfect positive correlation ($r=1$), so we just kept total lignin to avoid duplicity in subsequent analysis. The lignin content was also highly correlated with the HG ratio, and those two with total carbon content ($r=0.14$ and $r=0.17$, respectively). Lignin content also showed a significant trade-off with stem diameter ($r=-0.12$). Cδ13 showed a significant negative correlation with all growth traits, and a positive correlation with female reproduction (Figure 2.4).

Figure 2.4. Graphical pairwise correlations, correlation coefficients and significance of the correlation at the individual tree level for the 14 studied phenotypic traits in *P. halepensis*. Red boxes mark traits of the same functional group: A. Wood quality, B. Drought resistance, C. Growth and D. Reproduction traits. Below diagonal, graphical pairwise relationships. Above diagonal, pairwise Pearson's correlation coefficients and level of significance of the correlation (*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, · $P < 0.1$).

The multi-trait PCA based on individual tree data confirmed a positive relationship among traits of the same functional group and trade-offs between growth traits and reproduction, and wood extractives and lignin content (Figure 2.5). Cδ13 content is negatively related to growth traits, and total carbon positively with lignin content. The first PC (23% of variance explained) is mainly influenced by extractive traits, PC2 (22% of variance explained) by growth traits and Cδ13 and PC3 (12% of variance explained) by lignin content (Supplementary Information Table S2.1).

Using bivariate models, we found a highly significant negative correlation between Cδ13 and total extractives ($r = -0.10$, $CI: -0.19 - -0.04$), and between Cδ13 and stem volume ($r = -0.23$, $CI: -0.33 - -0.17$). The remaining pairwise genetic correlations among traits were not significant.

Figure 2.5. PCA based on individual tree data for the 14 wood quality, growth, drought resistance and reproductive traits studied in *Pinus halepensis*. Different colors represent trees from the three large gene pools, as stated in the key. Details on PCA results can be found in Table S2.1 in Supplementary Information.

At the population level, a number of statistically significant correlations were also observed between traits (Figure S2.4 in Supplementary Information). Besides the expected correlations within wood quality traits and growth traits, we found significant negative correlations between growth and reproductive traits. We also found a positive significant correlation between female and male flowering. Cδ13 was negatively correlated with growth traits, and positively with female reproduction, as identified at the individual level. PCA plots show these positive relationships and trade-offs graphically (Figure 2.6). The parallel analysis suggested that the optimal number of PC to retain was three (70% of variance explained). Details on PCA loadings can be found in Table S2.2 in Supplementary Information.

Bivariate models confirmed a highly significant positive genetic correlation between Cδ13 and female flowering at the population level ($r = 0.83$, CI : 0.52-0.97), and a negative one between stem volume and male flowering ($r = -0.58$, CI : -0.81– -0.13). The remaining pairwise genetic correlations among traits were not significant.

Figure 2.6. PCA based on population trait means for the 14 wood quality, growth, drought resistance and reproductive traits studied in *Pinus halepensis*. Different colors represent trees from the three large gene pools, as stated in the key. Details on PCA results can be found in Table S2.2 in Supplementary Information.

Trait-environment correlations

Pairwise correlations among the nine geographic and environmental variables are shown in Figure S2.5 in Supplementary Information. As expected, altitude showed a negative correlation with mean annual temperature and mean temperature of the warmest quarter. Longitude exhibited a highly significant positive correlation with mean temperature of the warmest quarter and the continentality index. Mean temperature of the warmest quarter was also highly significantly correlated with mean temperature of the coldest quarter and the continentality index.

The PCA parallel analysis revealed an optimum number of 3 principal components to retain (Figure S2.6 in Supplementary Information). These three first PC explained 81% of the total variance. PC1 was mainly influenced by mean annual temperatures, mean temperature of the coldest quarter and altitude, PC2 by rainfall and latitude and PC3 by the continentality index, mean temperatures of the warmest quarter and longitude.

Correlations between environmental factors and plant traits, reflecting ecotypic trends of variation, were higher for growth, Cδ13 isotopes and reproductive traits (Table 2.3). For wood quality traits only dichloromethane extractives were significantly correlated with the continentality index ($r = -0.40$, $P < 0.05$). Cδ13 was negatively correlated with annual rainfall ($r = -0.44$, $P < 0.05$) and rainfall during the warmest quarter ($r = -0.40$, $P < 0.05$). A positive

significant correlation was found with mean temperature of warmest quarter ($r= 0.62$, $P<0.005$) and the continentality index (CI; $r= 0.65$, $P<0.005$). Growth traits were significantly correlated with longitude, latitude and annual rainfall (Table 2.3). Female flowering was correlated with mean temperature of the warmest quarter ($r= 0.60$, $P<0.005$) and CI ($r= 0.54$, $P<0.005$). Male flowering was positively correlated to annual mean temperatures ($r= 0.53$, $P<0.005$), and negatively with summer rainfall ($r= -0.38$, $P<0.05$).

Table 2.3. Correlation of trait population means with geographic and environmental variables at the site of origin and the three most important PC from the PCA. Only traits showing significant provenance effects are included in the table. PWC: precipitation in the warmest quarter (June, July, August); AP: annual precipitation; MT: mean annual temperature; MTW: mean temperature of the warmest quarter; MTC: mean temperature of the coldest quarter; CI: continentality index. Significance: * $P<0.05$ ** $P<0.005$.

Trait	Long.	Lat.	Alt.	PWC	AP	MT	MTW	MTC	CI	PC1	PC3	PC2
Dichloromet.	-0.40*	-0.40*	-0.06	-0.29	-0.12	0.30	-0.19	0.15	-0.40*	0.21	-0.41*	-0.34
Ethanol	-0.33	-0.18	-0.12	-0.11	0.01	0.15	-0.12	0.07	-0.23	0.14	-0.28	-0.13
Water	-0.02	-0.19	-0.08	-0.19	0.15	0.20	0.14	0.25	-0.05	0.20	-0.01	-0.13
Total extract	-0.09	-0.16	-0.22	-0.19	0.11	0.29	0.03	0.25	-0.19	0.30	-0.16	-0.10
Cδ13	0.28	-0.26	0.25	-0.40*	-0.44*	0.09	0.62**	0.22	0.65**	-0.08	0.66**	-0.52*
Height	0.51*	0.49*	-0.11	0.32	0.55**	-0.23	-0.01	-0.03	0.01	0.01	0.14	0.59*
Diameter	0.59**	0.50*	-0.11	0.27	0.51*	-0.20	0.08	0.01	0.10	0.02	0.24	0.56*
Stem volume	0.66**	0.49*	-0.09	0.20	0.45*	-0.20	0.15	0.05	0.16	0.02	0.32	0.51*
Female flowering	0.14	-0.39*	0.14	-0.28	-0.19	0.18	0.60**	0.31	0.54**	0.03	0.51	-0.46*
Male flowering	-0.25	-0.45*	-0.25	-0.38*	-0.14	0.53**	0.23	0.37	-0.03	0.38	-0.08	-0.44*

Molecular markers and population structure

We obtained high quality genotypes for 294 polymorphic SNP loci in the 684 sampled trees. Molecular marker neutral differentiation among populations was moderate ($F_{ST} = 0.1762$, $SE=0.007$). Previous analysis reported the existence of range-wide population structure in the species, with $K=7$ optimal clusters (unpublished results, personal communication Rose Ruiz-Daniels and Delphine Grivet). Details on cluster membership for each individual can be seen in Figure 2.7.

Figure 2.7. Barplot of individual assignment probability to each of the optimal $K=7$ clusters in *Pinus halepensis*, as produced by STRUCTURE 2.3.4 software using 294 SNPs. 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.

Association tests

Using single-locus association tests we identified three SNPs significantly associated with diameter (SNP 211), total extractives (SNP 264) and female flowering (SNP 163; Figure 2.8). These SNPs were located in genes of known function and exhibited moderate marker effects (Table 2.4).

Figure 2.8. Box plots (genotypic effects) of SNPs that showed significant genetic association with diameter (SNP 211), total extractives (SNP 264) and female flowering (SNP 163) in single-locus association tests. In 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 2.4. SNPs that showed significant genetic association in single-locus association tests, after correction for multiple testing.

Trait	SNP	Gene	Annotation	Marker effect (R^2)	P	FDR Q-value
Diameter	211	<i>Pro</i>	profilin-like	0.026	0.00012	0.017
Total extractives	264	<i>XYL1</i>	putative alpha-xylosidase	0.026	0.00024	0.057
Female flowering	163	<i>PP2C</i>	protein phosphatase 2C	0.021	0.00046	0.105

Marker-trait single-locus association tests using as phenotypes the PCA scores of the three most important PCs from the multi-trait PCA (see above), confirmed the association for SNP 264, but differed in the remaining SNPs detected for individual traits. We found SNP 264 to be significantly associated with PC1 ($q\text{-val}=0.050$, $R^2=0.027$; Figure 2.9), primarily composed of extractive traits. The second principal component was related largely to growth traits and carbon isotopes. We found four additional SNPs associated with PC2 ($q\text{-val}=0.081$) exhibiting smaller marker effects (Figure 2.9): SNP 59 ($R^2=0.001$), SNP 71 ($R^2=0.001$), SNP 222 ($R^2=0.001$) and SNP 241 ($R^2=0.002$). These four SNPs are located in genes of unknown function.

Figure 2.9. Box plots (genotypic effects) of SNPs that showed significant genetic association with PC1 (SNP 264) and PC2 (SNP 59, SNP 71, SNP 222 and SNP 241). In 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.

For multi-locus association tests, we modelled trait variation as a function of the 294 successfully-scored SNPs, using BVSR. The proportion of variance explained (PVE) varied among traits (Supplementary Information Table S2.3.). Our SNP data explained a moderate proportion of variation of wood total extractives (PVE = 10.8%) and tree height (PVE = 9.7%). For the remaining wood quality, growth, drought stress and reproductive traits, BVSR models explained a smaller amount of phenotypic variation (1.7 - 8%, see Table S2.3. in Supplementary Information). Estimates of the number of SNPs underlying traits ranged from 4 for HG ratio to 22 for height and total C content (Supplementary Information Table S2.4). 95% credible intervals (CIs) surrounding these point estimates were typically very small (see Table S2.4). The average effect of associated SNPs ranged from 0.086 for water content to 0.412 for total wood extractives.

Multi-locus association tests confirmed the strong association of SNP 264 with total extractives (Bayes Factor>150). Another 18 strong evidences of association (BF>10, strong support according to Jeffreys' scale) were also found (Table 2.5). Among the 13 different SNPs exhibiting significant associations, 11 of them are located in annotated genes of known function in trees (Table 2.6). When considering only the three SNPs recovered with the largest posterior inclusion probability for each trait, we identified a number of SNPs that associated with multiple traits (SNP 22, SNP 227, SNP 259, SNP 264 and SNP 267; Supplementary Information Table S2.5). For example, SNP 227 associated with all traits related with wood lignin content and all tree growth traits. Estimates of $|β|$ were larger for SNPs associated with total wood extractives and stem volume than for the other traits (Table S2.5).

Table 2.5. SNPs that showed significant genetic associations ($BF > 10$) in multiple-locus association tests. In bold, associations with $BF > 20$, with stronger evidence of association according to (Kass and Raftery 1995). Boxplots of genotypic effects for those SNPs can be found in Figure S2.7 in Supplementary Information.

Trait	SNP	Effect size	β
Dichloromethane	SNP 326	1.225	-0.172
	SNP 264	1.159	0.209
Ethanol	SNP 38	1.402	-0.230
	SNP 264	1.196	0.212
Water	SNP 273	1.062	0.211
	SNP 309	1.041	0.144
Total extractives	SNP 264	2.205	0.270
Lignin	SNP 273	1.22	-0.222
	SNP 227	1.033	0.148
Ratio HG	SNP 227	1.212	0.156
	SNP 126	1.105	-0.143
Height	SNP 267	1.283	0.149
	SNP 185	1.073	-0.210
Diameter	SNP 227	1.314	-0.161
Stem volume	SNP 136	1.926	-0.220
	SNP 377	1.904	-0.242
	SNP 227	1.881	-0.183
	SNP 133	1.344	-0.190
Female flowering	SNP 216	1.095	-0.251

Table 2.6. Annotation and function of SNPs associated with wood quality, growth and reproduction traits identified using single-locus and multi-locus association tests.

SNP	Method	Trait	Gene	Annotation	Species	Best blastx	Function	
							Molecular	Biological
snp38	PiMASS	Ethanol extractives	-	unknown	-	-	-	-
snp126	PiMASS	Ratio HG	CLV	clavata-like receptor	<i>Picea glauca</i>	ABF73316	ATP binding, protein kinase activity	Regulation of cell proliferation in meristems
snp133	PiMASS	Volume	PG4	probable polygalacturonase	<i>Eucalyptus globulus</i>	ABG34278	Polygalacturonase activity	Carbohydrate metabolic process, cell wall organization
snp136	PiMASS	Volume	CSE	caffeoylshikimate esterase	<i>Eucalyptus grandis</i>	XP_010066716	Hydrolase activity	Lignin biosynthesis
snp163	Tassel	Female flowering	pp2c	protein phosphatase 2C	<i>Populus euphratica</i>	XP_011010154	Regulator of ABA responses	Stomatal closure, vegetative growth control
snp185	PiMASS	Height	4cl	4-coumarate-CoA ligase, partial	<i>Pinus halepensis</i>	ADL29066	Lygase activity	Lignin biosynthesis, plant defence
snp211	Tassel	Diameter	pro	profilin-like	<i>Phoenix dactylifera</i>	XP_008794098	Actin binding	Cell filament formation and plant cytoskeleton
snp216	PiMASS	Male flowering	BXL1	Beta-D-xylosidase 1	<i>Pinus tabulaeformis</i>	AJP06250	Hydrolase activity	Carbohydrate metabolic process, male and female cone development
snp227	PiMASS	Diameter, Volume, Lignin, Ratio HG	-	unknown	-	-	-	-
snp264	Tassel & PiMASS	Dichloromethane, ethanol and total extractives	XYL1	putative alpha-xylosidase	<i>Pinus pinaster</i>	AF448201_1	Carbohydrate binding, hydrolase activity	Carbohydrate metabolic process, cell wall extension and load-bearing
snp267	PiMASS	Height	Fd-GOGAT	glutamate synthase (ferredoxin)	<i>Pinus sylvestris</i>	CAA56286	Glutamate biosynthetic process	Nitrogen interconversion
snp273	PiMASS	Water extractives, Lignin content	CesA2	cellulose synthase	<i>Pinus radiata</i>	AAQ63936.1	Cellulose biosynthesis	Deposition of secondary cell walls
snp309	PiMASS	Water extractives	CesA1	cellulose synthase catalytic subunit	<i>Pinus taeda</i>	AAX18647	Cellulose synthase (UDP-forming) activity	Cell wall biogenesis/degradation
snp326	PiMASS	Dichloromethane extractives	RbohH	putative respiratory burst oxidase homolog protein H	<i>Musa acuminata</i>	XP_009402243.1	Oxidoreductase, peroxidase activity, Ca binding	Integral component of cell membrane
snp377	PiMASS	Volume	Dof	Dof zinc finger protein Dof2	<i>Pinus pinaster</i>	AGV07537	DNA binding	Regulation of gene expression, lignin production

DISCUSSION

Our results revealed a number of interesting correlations among traits related to growth, drought resistance, reproduction and defense at individual and population levels. These, describe integrated phenotypes matching theoretical predictions from allocation and life history theories. In addition, at the molecular marker level, we unveiled several associations that provide insights on the genetic basis of adaptive traits and possible pleiotropy for correlated traits in Aleppo pine.

Phenotypic variation, environmental correlations and trade-offs

Previous studies had already found intraspecific genetic variability in the species for growth traits (Voltas *et al.* 2008; Esteban *et al.* 2010; Santos-del-Blanco *et al.* 2013; De La Mata *et al.* 2014), water use efficiency (Ferrio *et al.* 2003; Chambel *et al.* 2007; Voltas *et al.* 2008, 2015) and morphometric and anatomical traits (Esteban *et al.* 2010). We did not find intraspecific among-population variation for wood chemical traits (Figure S2.1 in Supplementary Information), but a significant provenance effect, suggesting high within-population variability. Our results confirm previous among-population differences in C δ 13 content, growth and reproductive traits (Figures S2.2 and S2.3 and Table 2.2). Population differentiation may well be the result of adaptive responses, following the postglacial long-range colonization by the species of the western Mediterranean basin, after which genetic differentiation occurred (Grivet *et al.* 2009).

Supporting the hypothesis of local adaptation, we found significant correlations between ecogeographical variables and phenotypic traits. As found in previous studies, areas with milder temperatures and higher rainfall selected for high growth rates, non-conservative water use efficiency and low reproductive investment (Ferrio *et al.* 2003; Voltas *et al.* 2008, 2015; Santos-del-Blanco *et al.* 2013). This could be explained by selective pressures favouring a competitive strategy, involving more investment in vegetative growth in order to outcompete neighbouring trees (Grime 1977; Falster and Westoby 2003). In contrast, harsh environments, with severe drought and short vegetative periods, promote strategies of early and intense reproduction and conservative water use traded off against vegetative growth (Santos-del-Blanco *et al.* 2013). Earlier studies have shown intra-specific variation for C δ 13 in conifers related to climate conditions at the site of origin, such as rainfall for *P. sylvestris* (Cregg and Zhang 2001). Our results indicate that both summer rainfall and temperature oscillations (continentality index, CI) are important selective factors for water-use efficiency in Aleppo pine. Thus, a conservative water-use would be typical of drought-adapted populations from dry areas exposed to prolonged water stress periods in summer. Regarding

wood properties, only dichloromethane extractives were negatively correlated with the continentality index of the environment where the populations had evolved. This could be related to the existence of a cost for the production of extractives, which should be higher in harsher (i.e. more continental) environments.

A number of trade-offs has been previously reported between growth and reproduction in plants (reviewed in Obeso, 2002), and also specifically in Aleppo pine (Santos-del-Blanco *et al.* 2013) and maritime pine (Santos-del-Blanco *et al.* 2012, 2015). Our results are consistent with the hypothesis that energy invested in reproduction detracts from the amount of energy available for growth in long-lived taxa (Obeso 2002). A combination between vegetative and reproductive traits with other traits representative of complementary adaptive strategies is however novel in the literature. Interestingly, the ecotypic trends we found by combining all traits, represent putative adaptive strategies, where allocation to adaptive traits related to growth, reproduction and defence trade-off against each other. This can be seen graphically in Figures 2.5 and 2.6, where growth and reproduction oppose each other, as well as growth and water use efficiency. At the same time, different extractives are placed in vectors perpendicular to both reproduction and growth. Extractives are likely correlated to constitutive defences, thus representing defensive strategies. Trade-offs between growth and defensive investments have already been described in woody plants long ago (Loehle 1988).

Conservative water use strategies and enhanced reproduction were positively correlated and were found in populations where conditions for vegetative growth are harsh. A trade-off between vegetative growth and drought tolerance has been previously reported in this species (Voltas *et al.* 2008, 2015), as well as in other pine species, e.g. Scots pine (Cregg and Zhang 2001). However, Tognetti *et al.* (1997) reported a strong positive correlation between C δ 13 and height in maritime pine, and lodgepole pine populations with high cellulose C δ 13 also showed greater productivity (Guy and Holowachuk 2001). Such divergent association patterns between C δ 13 and productivity are interpreted as differences in the primary source for genetic differentiation in water use efficiency: either photosynthetic capacity or stomatal conductance (Voltas *et al.* 2008). It is hard to hypothesize what the relationship between drought tolerance and reproductive effort might be, and evidence so far is largely missing. To the extent that drought tolerance strategies are costly, e.g. investment in a deep root system (Voltas *et al.* 2015), resources invested in reproduction might decrease. Nonetheless, it is also possible that a deeper root system provides a more constant water and nutrient supply. Then, under stressful environmental conditions, available resource might be invested preferentially to reproduction rather than to growth (Santos-del-Blanco *et al.* 2013). Whether the patterns we observed here obey to hard-wired genetic trade-offs or are the result of selection against them (simultaneous selection aiming to

increase two negatively correlated traits is possible as long as their correlation coefficient is less than -1) would need to be elucidated through the estimation of additive genetic variances and covariances. Four phenomena are likely to be important for the evolution of trade-offs: mutation-selection balance, antagonistic pleiotropy, correlational selection and spatio-temporal heterogeneity (Roff and Fairbairn 2007). All four may be important, but unfortunately our data does not allow to determine whether one or several play a primary role. The existence of pleiotropic SNPs representing trade-offs could however be considered as evidence for their genetic basis (see *below*).

Association genetics

We identified a number of SNPs significantly associated with phenotypic variation in wood extractives and lignin content, growth and female reproduction. These SNPs had small effects on phenotypes (single-SNPs explained <3% of the phenotypic variance; Tables 2.4 and 2.5). The fact that all significant associations account just for a small proportion of phenotypic variance has been repeatedly reported in forest tree studies (Thumma *et al.* 2005; González-Martínez *et al.* 2007, 2008; Ingvarsson *et al.* 2008; Eckert *et al.* 2009; Dillon *et al.* 2010; Holliday *et al.* 2010; Prunier *et al.* 2013). This suggests genetic control by many loci with relatively small individual effects in the studied traits, reflecting a polygenic quantitative model. In forest trees, this polygenic control of continuous traits is supported by a long history of quantitative genetic studies (White *et al.* 2007; Savolainen *et al.* 2007).

Using single-locus association tests, we were just able to identify three SNPs with significant associations, each one associated to a different trait. SNP 163 in gene *pp2c*, was associated with female flowering, SNP 211 in gene *pro* with stem diameter and SNP 264 in gene *XYL1* with total extractives (Table 4 and Figure 8). These SNPs are likely located in close proximity to the causative polymorphisms, or are the quantitative trait nucleotides (QTNs) themselves, as LD appears to decay rapidly in all conifers studied to date (Neale and Savolainen 2004; Brown *et al.* 2004; Heuertz *et al.* 2006; Plomion *et al.* 2014). Their putative function in relation with the identified association is discussed below.

Most published genotype-phenotype association tests analyze each individual phenotype separately, although results on related phenotypes may be reported together. The purpose of analyzing multivariate phenotypes usually include discovery of novel genetic variants with pleiotropy effects, that is, affecting multiple phenotypes, with the ultimate goal of uncovering the underlying genetic mechanism (Yang and Wang 2012). Here, we performed single-locus association tests using multi-trait phenotypes corresponding to the main PCs from a PCA, and found SNP 264 to be significantly associated with PC1, mainly influenced by wood extractive traits. The same SNP was associated with total extractives in single-locus

association tests (see above), and also exhibited highly significant BF in multi-locus associations with dichloromethane and ethanol extractives and total extractives (see below), when tested trait by trait. This result must be considered with caution, as the method cannot determine whether the detected association is truly pleiotropic. That is, we cannot confirm that the marker locus affects all components of the integrated phenotype directly (Yang and Wang 2012). Aschard *et al.* (2014) recently demonstrated that PCA-based strategies achieve a moderate gain in power only in the presence of positive pleiotropy, but have great potential to detect negative pleiotropy or genetic variants that are associated with a single trait highly correlated to others. In our study, the PCA-based approach did not improve the PVE by the SNPs identified as significantly associated for each trait. Besides SNP 264, the four remaining SNPs identified in multi-trait single-locus association tests (associations with PC2) resulted to be in unknown regions of the genome. Thus, no conclusions can be made from this result until a better knowledge of the species' genome is available.

Using multi-locus association tests, we found SNP sets (4 to 22 SNPs depending on the trait) explaining a moderate variation in phenotypic traits (e.g. PVE = 9.7% for height, PVE = 10.8% for total extractives). Analyses based on expanded sets of candidate genes have shown potential to identify sets of markers that jointly explain a large fraction of the phenotypic variance in different traits: 20% of the phenotypic variance in early wood specific gravity was explained by a handful of SNPs in loblolly pine (González-Martínez *et al.* 2007), six SNPs explained 17% of the phenotypic variance in cold damage to stems in Douglas fir (Eckert *et al.* 2009). When association analyses are performed using a single SNP at a time, this implicitly assumes that a multifactorial trait is treated as being caused by a single locus, and all polygenic background variation is effectively ignored in the analysis (Ingvarsson and Street 2011). With the multi-locus approach, we increased the PVE by the SNPs (e.g. from 2.6% to 4.6% for diameter), but still most variance remained unexplained for all the studied traits. The reduced amount of phenotypic variation explained in multi-locus association tests for wood quality traits could be related to the reduced genetic control detected in the study for those traits or to the relatively low number of markers used. The SNPs identified with higher Bayes Factor (BF) included SNP 264, as reported above. Our results uphold the ability of both single-locus and multi-locus methods to identify SNPs with (relatively) high effects on the studied trait, but the multi-locus approach was more effective identifying a greater number of associated SNPs.

Various SNPs showed significant association with multiple correlated traits and different analysis methods (Table 2.6 and Supplementary Information Table S2.5). For example, in single-locus association tests, SNP 264 was associated with all extractive traits. This could be interpreted as a signal of pleiotropy, as the SNP is located in a putative alpha-xylosidase

gene, involved in carbohydrate metabolic processes and cell wall building. SNP 273 was associated with water extractives and lignin, being this SNP located in a gene responsible of cellulose biosynthesis in the deposition of secondary cell walls. SNP 227 was associated with diameter and lignin content, but the gene from which this SNP was obtained is still not annotated in the species. A deeper knowledge of gene functions, gene regulatory networks and expression profiles would be needed to confirm if the multiple associations described here are cases of true pleiotropy.

Some of the detected associations involved SNPs in well-known candidate genes for adaptive traits (Table 2.6). Seven SNPs showed significant associations with wood extractives and lignin content, after multiple testing correction. SNP 264, which was consistently associated using different methods with wood extractives (and thus with wood content in terpenoids, fats, phenolic substances and carbohydrates) is located in *XYL1* gene. This gene codes for an α -xylosidase related to xyloglucan oligosaccharide (the major hemicellulosic polysaccharide in the primary walls) metabolism and growth capacity in maritime pine (Sánchez *et al.* 2003). The same α -xylosidase activity has been previously described in *Arabidopsis* (Sampedro *et al.* 2001). Expressional studies in *Populus* showed that this gene is closely associated with response to photoperiod too (Ruttink *et al.* 2007). SNP 273, located in gene *CesA2*, was associated with water extractives and lignin content. Our result is consistent with previous studies, which found this gene to be involved in the biosynthesis of cellulose in lignified cell walls of the xylem of hardwoods such as *Populus tremuloides* and *Pinus radiata* (Krauskopf *et al.* 2005). Wegrzyn *et al.* (2010) also found one SNP in gene *CesA2* to be significantly associated with lignin content in *Populus trichocarpa*. Control over gene transcripts of the cellulose synthase may be part of a signalling mechanism triggering a stress related compensatory deposition of cellulose and lignin (the two major cell wall components) that is crucial to growth and development in trees. Additionally, one marker from a *CLV* gene (SNP 126) was associated with the ratio of lignin derived monomers. In *Arabidopsis*, *CLAVATA1* is a well-characterized gene integral to the balance between primordial differentiation and meristem proliferation. *CLV*-like genes also play a key role in caulogenesis in *P. pinaster*, *P. pinea* and other conifers (Alvarez *et al.* 2013). Lignin monomer content is fundamental determining mechanical properties of fibre and wood and pulp efficiency. Thus, identification of SNPs associated to this trait could be specially interesting for breeding purposes through marker assisted selection.

We also identified seven SNPs significantly associated to tree growth traits. Tree growth is determined by cell division and expansion in the apical and cambial meristems, developmental and seasonal growth transitions, efficiency of photosynthesis, nutrient and water uptake and transport, and the ability to respond to biotic and abiotic stresses. These

processes are controlled by a myriad of genetic and epigenetic factors that respond dynamically to environmental signals (Grattapaglia *et al.* 2009). These makes the identification of particular SNPs associated with growth traits particularly challenging. However, in the last decade, family-based QTL mapping and association studies have identified a few loci significantly associated with growth traits in pines, albeit controlling small proportions of the total phenotypic variation (e.g. Yu, Li, *et al.* 2006; Lepoittevin *et al.* 2012). Among the identified significant associations with growth traits we found three of them of particular interest: SNP 185, located in the *4CL* gene, SNP 267 in *Fd-GOGAT* gene, associated with height, and SNP 377 from a *Dof* gene, associated with stem volume. Expression profile studies showed that the *4CL* gene is involved in the lignin biosynthetic pathway in Eucalyptus (Paux *et al.* 2005) and in bud formation and dormancy in poplar (Ruttink *et al.* 2007). SNPs in this gene were also identified to be significantly associated with percentage latewood in loblolly pine (González-Martínez *et al.* 2007), with growth, phenology and growth tolerance in Douglas fir (Eckert *et al.* 2009) and with lignin content in poplar (Wegrzyn *et al.* 2010). Grivet *et al.* (2011) highlighted the potential role of the *4CL* gene in local adaptation of *P. pinaster* and *P. halepensis*, pointing out that it may be under selection in natural populations of these two Mediterranean pines. *Fd-GOGAT* is involved in the assimilation of ammonia derived from nitrate reduction and the reassimilation of ammonia released in photorespiration and protein catabolism in plants (García-Gutiérrez *et al.* 1995; Temple *et al.* 1998; Ishizaki *et al.* 2010). It is also related to chlorophyll accumulation and development of green plastids in a light-independent manner in pine seedlings (García-Gutiérrez *et al.* 1995). The *Dof* proteins are members of a major family of plant transcription factors that play diverse roles in regulating gene expression (Yanagisawa *et al.* 2004). Interestingly, it has recently been reported that *Dof* factors might also be involved in the regulation of lignin production (Rogers *et al.* 2005). In conifers, the *Dof* gene family plays a key role in the control of ammonium assimilation for glutamine biosynthesis (Rueda-López *et al.* 2008). *Dof* genes are directly involved in regulating primary metabolism in maritime pine, as they are involved in primary ammonium assimilation from the soil, a crucial process for tree growth and development (Rueda-López *et al.* 2008), and our results suggest that this may be also the case for Aleppo pine.

Association tests for reproductive traits in plants are scarce (but see Ballesta *et al.* (2014), who found two ISSR markers to explain 11% of flowering intensity variation in *Eucalyptus cladocalyx*). To our knowledge, this is the first study including flowering in an association test for a conifer species. Although the PVE was small, we found two markers significantly associated with reproduction: SNP 163 from the *pp2c* gene associated with female flowering, and SNP 216 from *BXL1* associated with male flowering. Expression profile studies stated

that gene *pp2c* is involved in photoperiod response, bud formation and dormancy in poplar (Ruttink *et al.* 2007). More recently, it has been confirmed that the gene is a negative regulator of abscisic acid (ABA) responses in the same species (Chen *et al.* 2015). It is widely known that ABA is related to tree flowering and fruiting. Thus, the associated marker could be indirectly controlling female flowering through ABA regulation.

Different SNPs showed significant association with multiple correlated traits (Table 2.6). Most likely, these SNPs truly participate in the control of only one of the traits, but show multiple significant associations because of trait phenotypic correlations. This highlights the utmost importance of considering trait correlations and trade-offs when interpreting association results, as we have done in our study. If previous quantitative analysis reveal high covariance among traits, we recommend to prioritize increasing the sampling effort in association studies, and maybe reducing the number of traits phenotyped, or also increase it if possible but directly work with integrated phenotypes. New insights in discovering the genetic architecture of adaptive traits may come from jointly testing for associations in multiple traits that are correlated due to shared (but hidden) genetic or non-genetic factors in a single multivariate association analyses. Upcoming multi-trait association tests should help us to detect genetic control and identify true pleiotropy (Lippert *et al.* 2014; Furlotte and Eskin 2015). Phenotypic trade-offs are commonly overlooked in association studies considering multiple traits, but this information should be considered when interpreting association results. Integration of quantitative genetic knowledge and novel genotype-phenotype association methods is essential to unravel the genetic basis of adaptive traits in long-lived tree species, which in turn will determine forest's capacity to cope with upcoming environmental conditions derived from human-driven climate change.

CHAPTER 3

POLYGENIC ADAPTATION IN MARITIME PINE

ABSTRACT

The complex interaction between genes, phenotypes, selection pressures and resulting adaptation processes has been studied for decades. However, the genetic architecture of adaptive traits remains poorly understood, especially in non-model species. Although most quantitative traits are assumed to be controlled by many loci of small effect, most studies still look for single-locus major effects, leaving most phenotypic variance unexplained. Here, we study the underlying genetic architecture of two fitness related traits in maritime pine, under a polygenic adaptation perspective. Using a multi-environment clonal common garden (total of 20,000 trees), we assessed phenotypic variation across years, and identified a moderate genetic control of survival and height, which were also structured across populations. Among-population quantitative genetic differentiation (Q_{ST}) was consistently above among-population molecular marker differentiation (F_{ST}) across traits and testing environments, at different hierarchical levels, suggesting local adaptation. Using 6,100 SNPs in a single-locus association test we found little evidence for large effect loci: six SNPs were associated with survival in Atlantic trials, and two SNPs were associated with tree height in one of the trials under Mediterranean climate. Genotype-phenotype association tests were also used to classify the available markers into adaptive and putatively neutral. Then, we tested for excess variance in among-population differentiation and signals of excessive correlations with environmental variables when compared to a null model of neutral genetic drift and shared population history. We found an excess of variance among populations for height and survival, consistent with directional selection, mainly driven by covariance among loci (i.e. linkage disequilibrium, LD). Populations from the Atlantic region contributed more to height variance excess at the species level, while within region among-population differentiation drove the directional selection signal for survival. We also found signals of excessive correlations with lowest temperatures and mean precipitation, respectively, for height and survival traits. Our results point to many loci of small effect underlying the studied traits, with local adaptation driven by small allele frequency changes. Studying the genetic architecture of adaptive traits under a polygenic perspective, combining population and association genetics, allowed us to identify signals of selection that would otherwise have remained hidden.

INTRODUCTION

Under the current climate change scenario, there is growing interest to understand the genetic basis of adaptive traits in forest tree species. Identifying signatures and targets of local adaptation in trees could help us predict the future of forests. But even though there is evidence of widespread local adaptation in plants (Leimu and Fischer 2008), the genetic basis of local adaptation remains poorly understood (Savolainen *et al.* 2013). What appears to be increasingly clear is that most adaptive events in natural populations occur by the evolution of polygenic traits, rather than by ‘hard selective sweeps’, i.e. those involving fixation of single beneficial mutations (Pritchard and Di Rienzo 2010; Le Corre and Kremer 2012). Most complex phenotypic traits of evolutionary, ecological and economic interest are known or thought to have a polygenic basis: they are controlled by a large number of loci dispersed throughout the genome, usually of small phenotypic effect (Mackay *et al.* 2009; Pritchard *et al.* 2010; Rockman 2012; Le Corre and Kremer 2012). In forest trees, most traits related to adaptation are also polygenic (Savolainen *et al.* 2007). But most approaches aiming at finding genes involved in adaptive events have focused on the detection of single outlier loci, which resulted in the discovery of individually ‘significant’ genes with strong effects. However, due to the fact that the studied traits are usually highly polygenic, the proportion of explained variance remains very low (e.g. for forest trees: González-Martínez *et al.*, 2007, 2008; Holliday *et al.*, 2010; Cumbie *et al.*, 2011; Prunier *et al.*, 2013), with most studies statistically unpowered to capture most molecular markers associated with the trait and leading to the so called missing heritability problem (Makowsky *et al.* 2011; Bloom *et al.* 2013).

Within species, populations become differentiated over time, both genetically and phenotypically, depending on the relative strengths of natural selection, genetic drift, migration and mutation. It is particularly interesting to be able to determine to what degree population differentiation is caused by selective (adaptive) or neutral (stochastic) processes (Leinonen *et al.* 2013). Forest trees, and particularly conifers, have a long history of common garden trials, and studies that have provided evidence of within-species local adaptation (e.g. Grivet *et al.* 2011; Santos-del-Blanco *et al.* 2012 for Mediterranean pines), as well as revealing the highly polygenic nature of adaptive traits (Savolainen *et al.* 2007). Polygenic adaptation could allow rapid adaptive shifts, but yet would often go undetected using conventional methods for detecting selection (Pritchard *et al.* 2010), as it is qualitatively different from the models of adaptive substitutions that dominate the population genetics literature (Pritchard and Di Rienzo 2010).

Population genetics approaches for the study of local adaptation include Q_{ST} - F_{ST} comparisons, scanning for F_{ST} outliers and correlation of allele frequencies with environmental variables (Savolainen *et al.* 2013). All methods seek for hints of natural selection by looking for departures from a standard neutral model. In Q_{ST} - F_{ST} comparisons, the proportion of variation between populations at marker loci (F_{ST}) is compared with Q_{ST} , an analogous measure for quantitative traits. If polygenic traits show higher differentiation than neutral markers ($Q_{ST} > F_{ST}$), researchers conclude that divergent selection is responsible for local adaptation (McKay and Latta 2002; Savolainen *et al.* 2007; Leinonen *et al.* 2013). Regarding F_{ST} outlier tests, there are many methods, differing mainly in the underlying demographic model they assume, the statistical approach and whether selection is explicitly included in the method (Luikart *et al.* 2003; De Mita *et al.* 2013). Correlations between allele frequencies and environmental variables are more recent and powerful and usually implemented using Bayesian approaches (see *Annex 2*: Jaramillo-Correa *et al.* 2015). Another set of methods, namely QTL and association mapping, rely on correlating phenotypes with genotypes. Both aim to identify particular loci underlying trait variation (i.e. genes controlling traits, likely under selection if there is local adaptation going on). The main difference between them is that association mapping considers populations rather than pedigrees (Balding 2006). Genotype-phenotype association tests are rapidly becoming the main method for dissecting the genetic architecture of complex traits in plants, helping to elucidate how plants have managed to adapt to disparate environmental conditions (Ingvarsson and Street 2011). Candidate gene based-association mapping has several important advantages for complex trait dissection in conifers (see *below*) and it has proved effective to identify genes underlying adaptive traits in these species (e.g. for pines González-Martínez *et al.*, 2008; Quesada *et al.*, 2010; Eckert *et al.*, 2012; Parchman *et al.*, 2012). However, both population and association genetics methods have rarely looked to polygenic adaptation in forest trees. Indeed, we would not expect these methods to identify selection if the effect is weak at the individual locus level because the trait is controlled by many genes (Le Corre and Kremer 2012; Tiffin and Ross-Ibarra 2014; Kemper *et al.* 2014). When selection acts on a phenotype that is underwritten by a large number of loci, the response at any given locus is expected to be modest. Thus, we would rather expect to observe coordinated small allele frequency shifts across many loci, with the phenotype increasing alleles, shifting in the same direction, if adaptation is truly polygenic (Pritchard and Di Rienzo 2010; Le Corre and Kremer 2012; Turchin *et al.* 2012; Berg and Coop 2014).

It has recently been argued that, to advance in the understanding of polygenic adaptation, it is mandatory to move beyond simply searching for genes. We should instead resume the use of population genetic data to advance our understanding of local adaptation as a process

(Turchin *et al.* 2012; Tiffin and Ross-Ibarra 2014). From this perspective, approaches looking for markers jointly exhibiting signals of selection show great promise (Keller *et al.* 2012). State-of-the-art approaches construct and test polygenic models by combining association and population genetics, as recently suggested by Berg and Coop (2014). For example, in brief, Berg and Coop (2014) suggest to use allele frequency data at significantly associated loci from association studies to test for a signal of selection on the phenotypes they underwrite, while accounting for the hierarchical structure among populations induced by shared demographical history and genetic drift.

Long-lived species are usually regarded as difficult experimental organisms, but pines exhibit particular advantages for the study of local adaptation, such as high genetic diversity and the fact that distribution ranges cover highly contrasting environments, after range expansions from glacial refugia (Hamrick *et al.* 1992; Savolainen *et al.* 2007). Their nearly undomesticated random-mating populations and long history of quantitative genetics and field experiments, add good opportunities for gaining general insights into the genetics of complex traits and polygenic adaptation. Maritime pine expanded from isolated glacial refugia and now can be found in scattered populations, living under contrasting environments across the occidental Mediterranean Basin and the European Atlantic front (Bucci *et al.* 2007; Figure 3.1). The species shows high genetic variation and a strong population structure (González-Martínez *et al.* 2002; Santos-del-Blanco *et al.* 2012; Jaramillo-Correa *et al.* 2015), offering a great opportunity to study the genetic basis of polygenic adaptive traits at different spatial scales (i.e. species-wide, gene pool and population level).

Here we look for polygenic adaptation in maritime pine, integrating both association and population genetics. To our knowledge, this is the first time that this kind of approach is applied to a plant species. Our experimental design, using the largest marker data set developed for the species up-to-date (6,100 SNPs) and a large number of phenotyped trees, growing in contrasting environments, increases the chances to identify signals of polygenic adaptation. To do so, we compute estimated genetic values based on associated SNPs and then test for variance excess in among-population differentiation, to identify geographical regions or populations which contribute to the polygenic signal, and also test if the genetic values are unusually correlated with environmental variables. We take advantage of the species' population structure, with clearly identified gene pools growing under distinct environmental conditions, to look for co-adapted gene complexes and polygenic adaptation signals across the species distribution range.

MATERIALS AND METHODS

Plant material and phenotyping

Five clonal common gardens were phenotyped for the study, including trees from 32 provenances sampled across the species distribution range (Figure 3.1, Supplementary Information Table S3.1). Each common garden comprised 522 clones (except for the Bordeaux trial, with 412 clones), with 8 replicates per clone, in a randomized complete block design (total of 20,000 trees). For more details on the experimental set up see Rodríguez-Quilón *et al.* (2015). The trial sites provided contrasted environmental conditions, with Asturias, Portugal and Bordeaux common gardens growing under Atlantic climate and Madrid and Cáceres common gardens under continental Mediterranean climate. We measured survival and tree height at early stages (3 to 6 year-old trees). Clonal replicates allowed multiple phenotypic measurements of each genotype.

Figure 3.1. Sampled *Pinus pinaster* populations (circles) and common garden trial sites (other symbols). Neutral gene pools outline the species natural distribution range in different colors.

Genotyping

Needles from the 522 genotypes were collected and dried in silica gel for subsequent DNA analysis. Details on DNA extraction and quantification, SNP discovery and genotyping can be found in *Annex 1: Plomion et al. (2015)*. In brief, the design of the *Illumina Infinium* SNP array included SNPs already validated in the species and new SNPs from RNA-seq and amplicon resequencing data. Although it is far from covering the species genome, the assay is enriched in SNPs from candidate genes which displayed signatures of natural selection (Eveno et al. 2008; Grivet et al. 2011), environmental associations (Jaramillo-Correa et al. 2015) and differential expression under biotic and abiotic stress in maritime pine (Plomion et al. 2016). After removing SNPs with uncertain scoring (visual inspection using *GenomeStudio*), we kept 6,100 SNPs for further analysis in the studied populations. Minor allele frequency (MAF) and SNP heterozygosity were computed in R. Only polymorphic loci (5,165 SNPs) were retained for subsequent analysis (Supplementary Information Figure S3.2).

Phenotypic trait analysis

All survival and height measurements in the five trial sites, across multiple years, led to a total of 20 traits (Table 3.1). Independent Bayesian models were fitted for each trait, using MCMCglmm package (Hadfield 2010) in R (R Core Team 2015). We defined block as a fixed factor and gene pool, population and clone as random factors, in order to estimate associated variance components to each hierarchical level. Genotypic best linear unbiased predictors (BLUPs) for each clone were also retrieved from the model to be used as phenotypes for GWAS (see below). We used a Gaussian model for height data and a binary model for survival data. Markov Chain Monte Carlo (MCMC) chains were run a minimum of 550,000 times, sampling every 5,000 and with a burn-in period of 50,000 iterations. Each MCMCglmm model was run three times to check if estimates were robust according to Kolmogorov–Smirnov tests.

Variance components were then used to compute broad-sense within-population heritability (H^2) (Falconer and Mackay 1996):

$$H^2 = (\sigma_{clone}^2) / (\sigma_{clone}^2 + \sigma_e^2)$$

where σ_{clone}^2 is the variance among clones within populations and σ_e^2 the error variance. For the estimation of broad-sense heritability of survival traits, we included an extra term in the denominator ($\pi^2/3$) to account for implicit *logit* link function variance (Nakagawa and Schielzeth 2010). Phenotypic differentiation among populations was computed at different

hierarchical levels, among gene pools (Q_{RT}) and among populations within gene pools (Q_{SR}) in R, according to equations:

$$Q_{RT} = \frac{\sigma_{gpool}^2}{\sigma_{gpool}^2 + \sigma_{pop}^2 + 2\sigma_{clone}^2} \quad Q_{SR} = \frac{\sigma_{pop}^2}{\sigma_{pop}^2 + 2\sigma_{clone}^2}$$

where σ_{gpool}^2 is the genetic variance between gene pools, σ_{pop}^2 the variance of populations within gene pools and σ_{clone}^2 the variance associated to clones from the MCMC models described above. Global Q_{ST} was computed following (Spitze 1993):

$$Q_{ST} = \frac{\sigma_{Bpop}^2}{\sigma_{Bpop}^2 + 2\sigma_{Wpop}^2}$$

where σ_{Bpop}^2 and σ_{Wpop}^2 are the genetic variances between populations and within populations, respectively, derived from a new set of models (20, one for each trait) in which only population and clone were included as random factors.

Q_{ST} - F_{ST} comparisons

Population neutral differentiation was estimated at different hierarchical levels (i.e. overall F_{ST} , among gene pools F_{RT} and within gene pools F_{SR}) according to Weir and Cockerham (1984) using the *hierfstat* package (Goudet 2005) in R. This way we obtained neutral molecular analogs of the previously computed estimates of genetic differentiation among populations for phenotypic traits (Q_{ST} , Q_{RT} and Q_{SR} , see above). Hierarchical $Q_{ST} - F_{ST}$ distributions were compared following Whitlock and Guillaume (2009), implemented in R.

Identification of genotype-phenotype associations (GWAS)

Single-locus genotype-phenotype association tests were performed independently for each trait using mixed linear models (MLM) following Yu *et al.* (2006), as implemented in TASSEL 3.0 (Bradbury *et al.* 2007). The kinship matrix was inferred following the Loiselle method (Loiselle *et al.* 1995) in SPAGeDi 1.3 (Hardy and Vekemans 2002), based on 1000 random SNPs. Negative genetic covariances between individuals were set to 0 as in Yu *et al.* (2006). Population structure was assessed using the Bayesian clustering method implemented in STRUCTURE 2.3.3 (Pritchard *et al.* 2000), and we found $K=6$ to be the optimal number of clusters for the studied individuals (Figure S3.1 in Supplementary Information). Hence, to control for population structure in the GWAS, individual assignment probabilities to each cluster from STRUCTURE analysis (Q1 to Q5) were included in the model as covariates. Multiple testing correction was performed using the false discovery rate (FDR) method from (Storey and Tibshirani 2003), implemented in R.

Single-locus association tests were also used to sort SNPs as candidate SNPs (those with $p < 0.05/20$, a more relaxed threshold than for single-locus association) and putatively neutral SNPs (the rest). Additionally, we ran new single-locus association tests for all traits using *SNPassoc* package (González *et al.* 2007), as TASSEL implements an additive + dominant model, and we needed exclusively additive effects for SNPs to be later included in the Berg and Coop (2014) test. To remove possible loci departing from neutrality from the putatively neutral SNP set, we performed an F_{ST} outlier test using the Bayesian method implemented in *BayeScan* 2.1 (Foll and Gaggiotti 2008). Molecular markers with $\log_{10}(\text{q-value}) < -3$ were considered as outliers and removed from the putatively neutral SNP set.

Polygenic adaptation test

To test for signals of polygenic adaptation, we used the method developed by Berg and Coop (2014), looking for covariance excess and over-dispersion of genetic values among populations when compared to a putatively neutral model also build from our data. The method follows three consecutive steps: First, it uses GWAS to identify SNPs associated with a phenotype and to estimate the additive effect of each variant on the phenotype in a specific population. Then, the estimated effect sizes and allele frequencies at these loci are used to predict the mean value of the trait we would expect to observe in each of the tested populations, in a model in which all effects are additive and the environment constant. Finally, it tests whether these genetic values differ among populations more than we would expect under a model of only genetic drift, i.e. it identifies divergent selection pressures and local adaptation. Signatures of polygenic adaptation are then evaluated using estimated genetic values through: (1) an excess of variance test, (2) the identification of outlier populations contributing to the polygenic signal, and (3) environmental correlations.

We ran two independent tests, one for survival traits and one for height traits. For both tests, SNPs were split into putatively adaptive candidates or neutral markers. We considered one SNP as a candidate if it had an association $p < 0.05/20$ for any trait (i.e. survival or height) at any trial site. We also removed F_{ST} outlier loci from the neutral SNP set (see above) to get rid of markers departing from neutrality to be used to estimate the neutral model. As environmental variables, we used the six main principal components resulting from an environmental principal component analysis (PCA) as implemented and described in Annex 2: Jaramillo-Correa *et al.* (2015). In brief, we used 32 environmental variables including monthly mean, highest and lowest temperatures and precipitation (retrieved from WorldClim (Hijmans *et al.* 2005)) and a regional climatic model (Gonzalo-Jiménez 2010), for non-Spanish and Spanish populations, respectively) to run two PCAs, one for summer and one for winter climate variables. Population scores for the three first PCs of each PCA (see table

2 in Annex 2: Jaramillo-Correa *et al.* 2015), which accounted for 95% of total climatic variation, were used for the analysis. We chose Cuellar as the match reference population for the neutral model. The polygenic adaptation test was performed in R, departing from the code provided by Berg and Coop (2014), available at <https://github.com/jjberg2/PolygenicAdaptationCode>. The same test was performed using randomly chosen SNPs as candidate SNPs to check that the observed signals were not driven by population structure. The method provides an estimation of among-population variance excess (Q_x , divided into F_{ST} -like and LD-like terms), identification of populations contributing to overdispersion (Z scores) and environmental correlations (r^2).

RESULTS

Phenotypic variation

Tree survival ranged between 2.7% (CI : 1.6-3.9, in Cáceres, 2011) and 98.9% (CI : 98.2-99.3, in Bordeaux, 2013) pointing out the marked difference in the environmental conditions among trial sites (Table 3.1). Asturias was the location with the tallest trees (115.1 cm, CI : 111.8-221.9 in 2014) and Cáceres the one with the most limited tree growth (16.2 cm, CI : 14.2-18.3 in 2011). Survival and growth were consistently higher in Atlantic trials (Asturias, Portugal and Bordeaux) compared to trials under Mediterranean conditions (Madrid and Cáceres; Table 3.1). Variance components from MCMC models associated to each hierarchical level (i.e. gene pool, population and clone) can be found in Table S3.2 in Supplementary Information. All traits were heritable, with height exhibiting higher heritabilities (mean $H^2=0.15$; Table 3.1).

Q_{ST} - F_{ST} comparisons

The studied traits were structured across populations, with Q_{ST} ranging from 0.260 in Cáceres 2011 to 0.518 in Bordeaux 2013 for height. However, for survival Q_{ST} ranged from negligible values to 0.523 (in Portugal 2011). Global F_{ST} was moderate ($F_{ST}=0.132$, CI : 0.130-0.135). Q_{ST} values were significantly higher than F_{ST} for all growth traits, and for survival traits in Portugal, Madrid and Cáceres trials.

Table 3.1. Trait means (MCMC model intercepts), broad sense heritabilities (H^2) and global phenotypic differentiation among populations (Q_{ST}) for *Pinus pinaster* clonal common gardens grown in five locations with contrasting environmental conditions. Height means are provided in cm and survival in %.

Trial site	Survival				Height			
	Trait	Mean	H^2	Q_{ST}	Trait	Mean	H^2	Q_{ST}
Asturias	surv2011	98.0	0.038	0.095	ht2011	26.8	0.166	0.328
	surv2012	98.0	0.037	0.104	ht2012	68.5	0.157	0.371
	surv2014	98.1	0.137	0.021	ht2014	115.1	0.090	0.415
Portugal	surv2011	64.9	0.010	0.523	ht2011	22.2	0.143	0.324
	surv2012	64.3	0.032	0.307	ht2012	47.0	0.086	0.406
	surv2013	63.3	0.023	0.396	ht2013	63.9	0.085	0.344
Bordeaux	surv2012	All clones alive			-	Not measured		
	surv2013	98.9	0.155	0.021	ht2013	73.4	0.158	0.518
	surv2014	98.6	0.221	0.000	-	Not measured		
Madrid	surv2011	20.1	0.019	0.369	ht2011	19.2	0.248	0.371
Cáceres	surv2011	2.70	0.030	0.235	ht2011	16.2	0.261	0.260

Regarding hierarchical population differentiation, we got contrasting results for height and survival traits. For height traits, among-region differentiation (max Q_{RT} =0.43 CI: 0.17-0.79 in Bordeaux, 2013) was higher than differentiation among populations within regions (max Q_{SR} =0.25 CI: 0.11-0.39 in Asturias, 2014; Figure 3.2). However, among-region differentiation for survival traits was negligible. Survival results show a stronger differentiation among populations within regions, albeit higher in trial locations with harsh environmental conditions (Q_{SR} =0.37 CI: 0.29-0.47 in Madrid, 2011). Gene pools showed a high genetic structure, with higher molecular marker differentiation found among regions (F_{RT} =0.116). Within region among-population neutral differentiation was markedly lower, with populations accounting for F_{SR} =0.018 of variance in allele frequencies. Phenotypic trait differentiation estimates were consistently above and significantly different from putatively neutral differentiation, overall and both at the among-region and among-population within region levels (i.e. $Q_{ST}>F_{ST}$, $Q_{RT}>F_{RT}$ and $Q_{SR}>F_{SR}$; Figure 3.2).

Figure 3.2. Hierarchical among-population differentiation for height in 32 *Pinus pinaster* populations growing in five trial sites with contrasting environmental conditions (Asturias, Portugal, Bordeaux, Madrid and Cáceres), across years.

Association genetics and marker effects

Using single-locus association tests, we identified 72 SNPs associated with height traits and 83 SNPs associated with survival traits, after multiple testing correction ($p < 0.05/20$; Table S3.3 in Supplementary Information). Among the identified candidate SNPs, only six were shared between height and survival candidate sets. After removing F_{ST} outlier markers from both neutral SNP sets, 4,593 and 4,582 neutral SNPs were kept, respectively, for height and survival polygenic adaptation tests. Effect sizes for the different measurements of the same trait (different year, same trial) were highly correlated ($r > 0.9$). The correlation between Tassel effect sizes (additive + dominant model, more restrictive) and *SNPassoc* effect sizes (only additive model) was $r = 0.89$. Absolute effect sizes exhibited a significant negative correlation with minor allele frequencies, as expected (i.e. less frequent alleles had bigger effects).

When we defined a more restrictive threshold for significance to avoid false positives (FDR, $q\text{-val} < 0.1$), we found only 8 highly significant associations: six SNPs associated with survival traits in Atlantic climate trials, and two SNPs associated with tree height in Cáceres, the Mediterranean trial with the harshest environment (Table 3.2). The effect sizes (EFF) were moderate and the genetic variance of associated SNPs accounted for less than 10% of trait phenotypic variation in all cases.

Table 3.2. SNPs significantly associated ($q\text{-val}<0.1$) with survival and height in *Pinus pinaster*, as identified in single-locus MLM association tests. In bold, SNPs with $q\text{-val}<0.05$.

SNP ID	SNP gene product name	Trait	$p\text{-val}$	$q\text{-val}$	r^2	EFF	MAF	Hobs	Hexp	F_{RT}	F_{SR}
2170	BX254222-1894	survdic11AST	3.71803E-06	0.019	0.049	-1.238	0.061	0.111	0.115	0.004	0.019
3778	CT579300-159	survmar14AST	3.44383E-07	0.002	0.076	-1.202	0.064	0.121	0.121	0.069	0.004
4625	F51TW9001BD684-1093	surv14BOR	2.73459E-05	0.063	0.052	-0.038	0.311	0.354	0.429	0.151	0.000
4627	F51TW9001BD684-880	surv14BOR	2.73459E-05	0.063	0.052	-0.038	0.311	0.354	0.429	0.151	0.000
4628	F51TW9001BD684-937	surv14BOR	3.86809E-05	0.063	0.051	-0.035	0.313	0.356	0.431	0.152	0.000
5592	i12810s468pg	survmay12POR	4.14457E-07	0.002	0.058	-1.512	0.025	0.038	0.049	0.076	0.120
5722	Myb2-1136	htdic11CAC	1.81639E-05	0.047	0.095	2.368	0.049	0.079	0.093	0.024	0.031
5724	Myb2-1811	htdic11CAC	1.81639E-05	0.047	0.095	2.368	0.052	0.084	0.098	0.023	0.038

Polygenic adaptation test

The among-population excess variance test identified Q_X values clearly above the null expectation for all height traits and for six survival traits. Q_X ranged from 68.4 in Bordeaux 2013 to 122.6 in Cáceres 2011 for height (Q_X p -val<0.001 for all traits), and from 33.6 in Asturias 2011 to 58.8 in Bordeaux 2014 for survival ($0.5 < Q_X$ p -val<0.001 depending on trait; Table S3.3). We found Q_X values in the upper tail of the null distributions, pointing to an excess of variance among populations (Figure 3.3). This variance excess was mainly driven by the LD-like component of Q_X (Figure 3.4, Table S3.3), revealing an excess of covariance among loci underlying the trait in a consistent way across populations.

Figure 3.3. Histograms of the empirical null distribution of Q_X and Q_X observed values for 3 height and 3 survival exemplary traits. The mean of each Q_X null distribution is marked with a vertical black dotted bar and the expected Q_X density is shown as a black curve. Observed Q_X values are represented by a red line.

Figure 3.4. The two components of Q_X (F_{ST} -like and LD-like components) for *Pinus pinaster* height in Asturias, 2011. The null distribution of each statistic is shown as a histogram, the mean value as a black dotted bar and the observed value as a red bar.

For height traits, our signal of excess variance arises mainly from a larger differentiation among populations in different regions (i.e. gene pools), with French Atlantic populations exhibiting higher genetic scores in all trial sites. Populations from Corsica and Morocco gene pools showed higher population specific Z scores for height traits, even higher in Atlantic trials (Figure 3.5). Cuellar and Olba, two populations from the Central Spain gene pool, also exhibited consistently higher Z scores for height. By contrast, for survival, the signal of excess variance among populations came from within-region population differentiation in all trial sites, except in Bordeaux 2013 (Figure 3.6).

Regarding environmental effects, we found a signal of excessive correlation for all height traits with summer principal component SUM2, and for height traits from Atlantic trials also with winter WIN1 (Table 3.3). Both of these environmental PCs are related to minimum temperatures in summer (SUM2) and winter (WIN1) (see Table S2 in *Annex 2: Jaramillo-Correa et al. 2015*). We also identified a strong correlation between survival in Portugal and environmental principal components SUM1, SUM2 and WIN3, and between survival in Bordeaux and SUM3. The principal components SUM1 and SUM3 were mainly loaded by mean precipitation, and SUM2 and WIN3 by minimum and maximum temperatures, respectively.

Table 3.3. Climate correlations for all *Pinus pinaster* traits in the global polygenic adaptation test. Asterisks indicate significant correlations ($\text{sign}(\beta)r^2$): * $P < 0.05$. ** $P < 0.05/(20 \times 6) = 0.0004$

Trial	Trait	SUM1	SUM2	SUM3	WIN1	WIN2	WIN3
Asturias	ht2011	0.003	0.102**	0.013	0.049*	-0.005	-0.007
	ht2012	0.005	0.097*	0.009	0.045*	-0.013	-0.004
	ht2014	0.001	0.086*	-0.001	0.041*	-0.011	-0.022
Portugal	ht2011	-0.008	0.085*	0.005	0.027	-0.018	-0.032
	ht2012	-0.008	0.079*	0.010	0.029*	-0.010	-0.029
	ht2013	-0.002	0.090**	0.018	0.034*	-0.005	-0.028
Bordeaux	ht2013	-0.006	0.091**	-0.001	0.044*	-0.004	-0.028
Madrid	ht2011	0.005	0.061*	0.051	0.011	-0.009	0.036
Cáceres	ht2011	-0.014	0.066**	-0.011	-0.022	-0.047	-0.067
Asturias	surv2011	-0.004	-0.015	-0.008	-0.020	-0.003	-0.016
	surv2012	0.004	-0.020	0.006	-0.017	-0.004	0.023
	surv2014	0.031	-0.034*	-0.016	-0.018	-0.026	0.021
Portugal	surv2011	0.055**	-0.004	0.043	0.002	-0.003	0.062*
	surv2012	0.056*	-0.028*	0.060*	-0.012	0.016	0.075*
	surv2013	0.036*	-0.031*	0.039	-0.016	0.012	0.038
Bordeaux	surv2012	0.025	-0.007	0.084*	0.000	0.039	0.065
	surv2013	0.008	0.002	0.059*	0.003	0.004	0.062*
Madrid	surv2014	0.012	-0.003	0.071*	0.000	0.019	0.065*
	surv2011	0.016	0.025	-0.004	0.009	-0.028*	0.005
Cáceres	surv2011	0.011	0.012	-0.007	0.013	0.018	-0.022

Figure 3.5. Representation of outlier analysis at the regional and individual population levels for height measured in five common gardens with contrasting environmental conditions. Populations are sorted by region (i.e. gene pool) according to key. Each circle represents the genetic score of the population, with circle sizes proportional to the population specific Z score.

Figure 3.6. Representation of outlier analysis at the regional and individual population levels for survival measured in five common gardens with contrasting environmental conditions. Populations are sorted by region (i.e. gene pool) according to key. Each circle represents the genetic score of the population, with circle sizes proportional to the population specific Z score.

DISCUSSION

Unravelling the genetic basis of adaptive traits in forest trees is key to understand past adaptive processes and predict the future of forests under climate change. Here, we assess standing phenotypic variation and genetic control of two fitness related traits in maritime pine, look for single-loci genotype-phenotype associations, and move forward towards a polygenic perspective. We demonstrate that covariances among putatively adaptive loci, although commonly overlooked in current methods to detect local adaptation, may be key to identify signals of local adaptation involving polygenic adaptive traits.

Contrasting environmental conditions in the five trial sites led to very different survival and height means, providing an ideal scenario for association and polygenic adaptation testing in the same genotypes under different selection pressures. The moderate heritability for both traits, somewhat higher for height, provided evidence of the genetic control of the traits under study. Maritime pine exhibited substantial phenotypic variation among populations for both survival and height, under any testing environment. This is consistent with previous studies in the species, which proved differentiation among gene pools and populations for different quantitative traits (e.g. Alfa *et al.* 1997; Correia *et al.* 2008; Santos-del-Blanco *et al.* 2012). In our study, phenotypic differentiation was higher among regions than among-populations within regions for height traits, as it would be expected when distinct gene pools grow under contrasted environmental conditions. However, for survival traits, most differentiation occurred within regions ($Q_{SR} > Q_{RT}$). This could be related to selective agents acting at a smaller local scale for survival. We found no evidence of increased population differentiation under more stressful environments (Table 3.1). Global F_{ST} (0.132) was moderate and similar to previous values reported in the species for allozymes (González-Martínez *et al.* 2002), microsatellites (Vendramin *et al.* 1998; Bucci *et al.* 2007; Lamy *et al.* 2011) and smaller SNP sets (see Chapter 4). Variation of allele frequencies among populations within regions accounted for less than 2% of the variance observed, indicating homogeneous neutral gene pools.

Phenotypic among-population differentiation consistently above that estimated with molecular markers, suggests local adaptation in the species. Local adaptation processes based on $Q_{ST} > F_{ST}$ have already been suggested in maritime pine (e.g. González-Martínez *et al.* 2002). However, although widely used, this approach suffers from limitations (Leinonen *et al.* 2013; Tiffin and Ross-Ibarra 2014), and $Q_{ST} > F_{ST}$ results should not be taken as an unequivocal sign of local adaptation and directional natural selection. Broadly, the efficiency of population genetics approaches has been evaluated using experimental and simulated data sets (Beaumont and Balding 2004; Foll and Gaggiotti 2008; Excoffier *et al.* 2009; Narum and Hess 2011) and it is largely recognized that they may lead to a high rate of false

positives, as evolutionary histories and population structures are generally not accounted for deriving expected neutral F_{ST} values, and Q_{ST} may be hindered by environmental effects (Le Corre and Kremer 2012; Leinonen *et al.* 2013). Our experimental design, with multiple clonal replicates, contrasting trial environments and testing at different hierarchical levels should overcome some of these previously noted drawbacks, identifying a true signal of local adaptation for the studied traits in maritime pine.

Local adaptation may result from alleles with environment-dependent fitness tradeoffs (i.e. antagonistic pleiotropy) or alleles that confer a selective advantage in one environment but are neutral in others (i.e. conditional neutrality; Anderson *et al.* 2013). Besides, control of adaptive traits could be driven by few genes of large effect, many genes of small effect, or both. The fact that traits that confer local adaptation are generally polygenic quantitative traits, make identification of loci driving their variation a challenge, even more in non-model species where genomes are less known. Using single-locus association tests, we only found 6 and 2 SNPs significantly associated with height and survival, respectively. Notably, two of the SNPs associated with tree growth are located in gene *Myb2*, belonging to a gene family involved in secondary wall biosynthesis during wood formation in poplar (Zhong *et al.* 2013). Further independent validation of functional importance, although commonly overlooked, would be needed to elevate the status of associated genes from candidates to causative (Tiffin and Ross-Ibarra 2014). Associations with height traits were only found in trials under Atlantic conditions, and survival associations only under harsh continental Mediterranean conditions (Cáceres trial). We increased the chances of including loci underlying the trait jointly at the species level by including together the SNPs selected in all trial sites when performing polygenic adaptation tests. No association was replicated in different testing environments, with variable selection pressures apparently leading to different relevant genes. Most likely, local adaptation largely occurs through different genes in different parts of the species range (Fournier-Level *et al.* 2013). This could be further explored testing for genotype-phenotype association and polygenic adaptation within regions (i.e. gene pools) independently in future studies with higher sample sizes.

The identified associated loci had small effect sizes, and explained a reduced proportion of phenotypic trait variation (<10%), suggesting a polygenic control of the studied traits, i.e. a larger proportion of the heritable genetic variation is explained by multiple SNPs of small undetectable effect rather than by individual SNPs of large effect. An alternative explanation is that our SNP array, despite being the biggest described so far in the species and including important candidate genes, sampled only a small proportion of the genome, and thus it may have failed to integrate a meaningful portion of the genes underlying local adaptation or phenotypic variation for the target traits. But notice that candidate-gene association mapping

has proved effective to identify loci showing evidence of moderate to large effect on fitness-related traits in pines (e.g. González-Martínez *et al.* 2007, 2008; Quesada *et al.* 2010; Eckert *et al.* 2012; Parchman *et al.* 2012; Lepoittevin *et al.* 2012; Budde *et al.* 2014). Nevertheless, the molecular evidence of selection acting on quantitative adaptive traits is expected to be weak if the signal of selection is distributed across many loci (Le Corre and Kremer 2012; Tiffin and Ross-Ibarra 2014; Kemper *et al.* 2014). This signal could be even harder to detect in species with very large genomes, such as pines (typically >30 Gb; Ahuja and Neale 2005; De La Torre *et al.* 2014). In this context, an alternative approach under a polygenic perspective is to compare statistics calculated at loci of interest (candidate loci) to statistics found at non-candidate or reference loci.

The method we implemented here, developed by Berg and Coop (2014), gains considerable power to detect selection over single locus tests by looking for unexpected covariance among loci in the deviation they take from neutral expectations. Our among-population excess variance tests reject the neutral hypothesis, suggesting that our estimated genetic values are overly dispersed compared to the null model of neutral genetic drift and shared population history. This Q_X excess of variance is consistent with the differential action of natural selection on the phenotype among populations. These results are consistent with directional selection acting in concert on multiple small effect alleles, influencing respectively, height and survival, to drive differentiation among populations at the level of the phenotype, and leading to local adaptation. The detected among-population variance excess was mainly driven by LD (Figure 3.4, Supplementary Information Table S3.3). This is precisely what would be expected for local adaptation of a polygenic trait in a species with reduced F_{ST} values, such as maritime pine. With differential selection among populations, we expect loci underlying a trait not only to vary more than expected under a neutral model, but also to covary in a consistent way across populations (Berg and Coop 2014). Both height and survival most likely reached local adaptation via increased covariance of allelic effects rather than via allele frequency changes, although both options should not be regarded as mutually exclusive. Recent models of local adaptation predict that it is this covariance among alleles that is primarily responsible for differentiation at the phenotypic level (Le Corre and Kremer 2003, 2012; Latta 2003). It has been shown that LD is indeed driving local adaptation of traits such as height or skin pigmentation in humans (Turchin *et al.* 2012; Berg and Coop 2014). Most likely, alleles with largest effect sizes may sweep to fixation while much of the adaptive response is caused by smaller allele frequency shifts across many loci (Chevin and Hospital 2008; Pritchard and Di Rienzo 2010).

Traits as the ones studied here, that have standing variation for selection to act on and are heritable, are most likely to evolve in a polygenic way (Pritchard and Di Rienzo 2010). A

polygenic architecture for the measured fitness-related traits was also expected, given the multifaceted and quantitative nature of both traits. Local adaptation via increased covariance of allelic effects is even more likely under high gene flow, as it is the case for maritime pine (González-Martínez *et al.* 2006), and when selection is recent (Le Corre and Kremer 2012).

Among-region differentiation contributed more to the detected overdispersion for height, while, for survival, the signal of excess variance came primarily from among-population differentiation within regions (Figures 3.5 and 3.6). Observation of extreme *Z* scores for a particular population or group of populations may be seen as evidence that that group has experienced directional selection on the trait of interest (or a correlated one) that was not experienced by the other populations in the analyses (Berg and Coop 2014). Outlier regions for height included Corsica and Morocco, two peripheral regions located in a Mediterranean island and North Africa, respectively. Particular populations with increased *Z* scores included San Leonardo, in the northern edge of the Central Spain region, and Olba, in the eastern edge of this same region. Particular and different selection pressures leading to locally adapted genotypes are expected in those areas, given their location at the edges of the species distribution. Marginal populations, growing under extreme distribution edge conditions, usually exhibit significant divergence from nearest populations. They are often long-term stores of species' genetic diversity, and are disproportionately important for the species survival and evolution (Hampe and Petit 2005). For survival, variance excess was mainly driven by among-population differentiation, independently of their region of origin. Local scale selection pressures may be playing a critical role over this highly polygenic trait that is directly connected with fitness (see review in Scotti *et al.* 2015). Highly variable results under contrasting environmental conditions suggest the implication of different gene complexes in trait control under changing environmental pressures. In contrast to species-wide adaptations, local adaptation of particular populations can be due to ongoing or very recent spatially varying selection (Kawecki and Ebert 2004).

Regarding the identification of selection drivers, we found all height traits (5 trial locations, all years) to be correlated with the principal component SUM2, mainly influenced by minimum temperatures in the warmest quarter. Height traits from Atlantic trials (Asturias, Portugal and Bordeaux) were also correlated with WIN1, determined by minimum and mean temperatures in winter. This is consistent with previous single-locus environment-genotype association tests in maritime pine based on the same PCs, which found sixteen SNPs correlated with axes loaded by temperature (SUM2, WIN1, and WIN3; Jaramillo-Correa *et al.* (2015). In our study, mean precipitation (SUM1 and SUM3) was also significantly correlated with survival in two of the tree trials under Atlantic climate. However, Jaramillo-Correa *et al.* (2015) found only five significant SNP correlations with components associated with precipitation (SUM1).

Water availability and appropriate temperatures are crucial for plant survival, even more in Mediterranean forests, where precipitation is often scarce and temperatures over the optimal. Temperatures are involved in growth phenology and cold hardiness, playing a key role in tree growth through their effects on photosynthesis, cell division and expansion (Ryan 2010). Thus, our results were somehow to be expected. However, they should also be taken with caution, as it can be argued that these correlations might be biased due to the spatial structure of climate variables, which partially match the genetic structure of the species. In complex environments, the interpretation of positive results remains challenging due to the important role of GxE interactions (Berg and Coop 2014), making it difficult to disentangle the roles of demographic history and selection.

CHAPTER 4

MARITIME PINE CONSERVATION GENETICS

ABSTRACT

Preserving intraspecific genetic diversity is essential for long-term forest sustainability in a climate change scenario. Despite that, genetic information is largely neglected in conservation planning, and how conservation units should be defined is still heatedly debated. Here, we use maritime pine (*Pinus pinaster* Ait.), an outcrossing long lived tree with a highly fragmented distribution in the Mediterranean biodiversity hotspot, to prove the importance of accounting for genetic variation - at both neutral molecular markers and quantitative traits - to define useful conservation units. Six gene pools associated to distinct evolutionary histories were identified within the species using 12 microsatellites and 266 Single Nucleotide Polymorphisms (SNPs). In addition, height and survival standing variation, their genetic control and plasticity were assessed in a multisite clonal common garden experiment (16,544 trees). We found high levels of quantitative genetic differentiation within previously defined neutral gene pools. Subsequent cluster analysis and post-hoc trait distribution comparisons allowed us to define ten genetically homogeneous population groups with high evolutionary potential. They constitute the minimum number of units to be represented in a maritime pine dynamic conservation program. Our results uphold that the identification of conservation units below the species level should account for key neutral and adaptive components of genetic diversity, especially in species with strong population structure and complex evolutionary histories. The environmental zonation approach currently used by the pan-European genetic conservation strategy for forest trees would be largely improved by gradually integrating molecular and quantitative trait information, as data become available.

INTRODUCTION

Human-driven environmental change is a broadly recognized threat to global biodiversity (Bellard *et al.* 2012) and how to ensure species persistence under rapid climate change remains a challenge in conservation biology (Sgrò *et al.* 2011). Mediterranean forest ecosystems are particularly sensitive to such rapid changes, specially trees, their keystone species (Fady *et al.* 2015). In fact, the latest predictions forecast a shift in their distribution due to climate change (Aitken *et al.* 2008; Benito *et al.* 2011). However, the exact patterns they will follow and their microevolutionary responses remain highly unpredictable, given the multiple factors involved in the process, the uncertainty of future climate scenarios, their long generation times and the spatial and temporal environmental heterogeneity of their distribution ranges (Alía *et al.* 2014; Fady *et al.* 2015). Recurrent wildfires, changes in land use and new biotic risks add to threats directly derived from global warming, such as drought stress and increased temperatures (Leone and Lovreglio 2004). In response, trees can adapt, migrate or become extinct (Aitken *et al.* 2008). But we still lack data integrating approaches to assess the evolutionary consequences of global environmental changes and their impact on genetic variation, phenotypic plasticity and the adaptability of forest tree populations (Chevin *et al.* 2010; Alberto *et al.* 2013), being these essential information to design and implement conservation strategies.

Maintaining species diversity and habitats has been the main objective of conservation policies so far, while intraspecific genetic diversity has been commonly overlooked (Laikre *et al.* 2010). At the same time, conservation of particular populations within species has been a central task in the European Forest Genetic Resources Program (EUFORGEN; see www.euforgen.org), as most widely distributed forest trees in Europe are not endangered at the species level, but often include threatened valuable populations. The reduced level of domestication and widespread local adaptation in forest trees (Wang *et al.* 2010; Alberto *et al.* 2013) have promoted specific *in situ* dynamic conservation strategies worldwide (Koskela *et al.* 2013; Lefèvre *et al.* 2013). Dynamic conservation aims at maintaining evolutionary processes and adaptive potential in natural populations and man-made stands, to ensure their long-term sustainability (Namkoong 1997). Since first applied to forest trees (Ledig 1986), it has been further implemented and translated into conservation management (e.g. Koskela *et al.* 2013; Lefèvre *et al.* 2013). However, despite significant improvements in recent decades, there are still several questions that need to be addressed, such as how to select target units for conservation and how many, how to assess their viability, and how to manage them (Graudal *et al.* 2014).

The importance of forest genetic diversity for forest sustainability is widely recognized, but nevertheless usually ignored when developing indicators and guidelines for forest management under climate change (Sgrò *et al.* 2011; Graudal *et al.* 2014; Fady *et al.* 2015). A reasonable starting point is to consider that a conservation network should span a species' entire standing genetic variation (Crandall *et al.* 2000). Then, two main approaches have been used to define tree conservation units: i) methods based on the contribution of particular populations to the total neutral genetic diversity or differentiation of the species (Petit *et al.* 1998), and ii) methods based on ecological and geographical information (e.g. Hamann *et al.* 2005; Lefèvre *et al.* 2013). The existing European network of dynamic conservation units is largely based on this second approach (Lefèvre *et al.* 2013), assuming that main climate types in Europe are closely linked with genetic differentiation, indirectly including adaptive processes under the hypothesis of local adaptation and clinal or ecotypic variation in tree species (see Alberto *et al.* 2013 for a review). Neutral molecular markers are considered an appropriate first step to designate conservation units (Fraser and Bernatchez 2001; Sagnard *et al.* 2002), as they provide information about historical evolutionary divergence and demographic processes, although they are often unsuitable to quantify genetic changes responding to particular selection pressures (Ouborg *et al.* 2006; Reed and Frankham 2001). Patterns identified in quantitative adaptive traits normally differ from those observed using neutral or weakly selected loci (Fraser and Bernatchez 2001; Sagnard *et al.* 2002; Frankham 2010). However, difficulty - real and perceived - of measuring adaptive genetic diversity has prevented its integration in conservation planning and management. It is urgent to overcome this situation, particularly for fragmented, structured and locally-adapted tree species facing climate change, where intraspecific diversity becomes of paramount importance (Kapeller *et al.* 2012). Moreover, phenotypic plasticity needs to be considered too, as it can buffer selection towards adaptation, or make populations overcome stressful conditions via plastic changes (Chevin *et al.* 2010; Fady *et al.* 2015). Phenotypic traits may thus be more useful than previously thought in selecting populations for conservation.

A combined approach including molecular, quantitative and ecological data, although rarely employed, is fundamental and would allow the integration of evolutionary processes in conservation planning (Sagnard *et al.* 2002; Ouborg *et al.* 2006; Wang *et al.* 2010; Sgrò *et al.* 2011). Here, we use maritime pine (*Pinus pinaster* Ait.) as a case study to show the usefulness of considering both neutral and quantitative genetic variation to identify dynamic conservation units. Maritime pine is an outcrossing, wind-pollinated long-lived tree of great ecological and economic importance that grows along a wide range of elevations, substrates, and climate regimes. It survived last glaciations in multiple refugia (Burban and Petit 2003;

Bucci *et al.* 2007), maintaining large amounts of genetic diversity despite its fragmented distribution, notably in the Mediterranean area (Figure 4.1). As a result of this complex evolutionary history, it has a strong population genetic structure (González-Martínez *et al.* 2004; Santos-del-Blanco *et al.* 2012; Jaramillo-Correa *et al.* 2015) and shows signals of local adaptation (e.g. genetic differences among populations for adaptive traits, Alía *et al.* 1997, Santos-del-Blanco *et al.* 2012; or correlations between allele frequencies of polymorphisms in candidate genes for climate adaptation and climate variables, Jaramillo-Correa *et al.* 2015).

We used previously developed molecular markers to classify populations into gene pools with distinct evolutionary histories. We considered these gene pools inferred from neutral genetic variation as a basis to identify a minimum set of potential conservation units. Based on a multisite clonal test we then assessed levels of standing genetic variation, differentiation among populations and phenotypic plasticity for two adaptive traits. We used this information to identify genetically differentiated population groups and evaluated their viability for dynamic conservation. Finally, based on our results, we assessed the suitability of the climatic zonation currently used by the European Forest Genetic Resources Program for maritime pine conservation. From this case study, we introduce a cascade decision approach to identify candidate populations for dynamic conservation and suggest key genetic parameters to account for in future forest tree conservation, based on both molecular markers and quantitative traits.

MATERIALS AND METHODS

Plant material

We collected open-pollinated seeds from 35 natural populations covering the main climates across the maritime pine natural range (Figure 4.1). About 30-40 trees, separated at least 50 m from each other, were sampled, and one single seed from each tree was planted in nursery. Then, each sapling (the ‘mother plants’) was propagated by cuttings to obtain clones. These clones (average of 15 clones per population, for details see Table S4.1 in Supplementary Information) were used to establish a common garden experiment of 16,544 trees, in 2010. Each of the four replicated trials comprised 4,136 trees, with 517 clones (i.e. genotypes) and eight replicates per genotype, in a randomized complete block design. The four trial sites provided contrasting environmental conditions (Figure 4.1). Trials in Asturias and Portugal are located in the Atlantic region, with high annual rainfall (1,160 and 985 mm, respectively) and mild temperatures (11.3 and 14.2 °C, respectively), although Portugal shows less than half the summer rainfall than Asturias (63 and 152 mm, respectively). Trials

in Madrid and Cáceres are located in continental areas under Mediterranean influence (annual rainfall of 443 and 898 mm, and mean annual temperature of 13.7 and 15.5 °C, respectively), with large seasonal temperature oscillations and a marked summer drought.

Figure 4.1. (A) Climatic stratification of Europe (Metzger *et al.* 2005) and maritime pine distribution, location of the 35 source populations (circles) and clonal common gardens (other symbols). Size of population symbols is proportional to Nei's diversity (H_e) for 266 SNPs. Different symbol and contour colors indicate the six different neutral gene pools derived from 12 nuSSRs and 266 SNP markers. (B) Proportion of each neutral gene pool area covered by different climatic regions.

Molecular markers

Needles from all the 517 genotypes were collected from mother plants in nursery and dried in silica gel for subsequent DNA extraction. Samples were genotyped for 12 nuclear microsatellites (nuSSRs) and 266 Single Nucleotide Polymorphisms (SNPs; data deposited in Dryad with doi: 10.5061/dryad.c289v). Details for DNA extraction, marker typing and scoring can be found in Santos-del-Blanco *et al.* (2012) and Jaramillo-Correa *et al.* (2015), and references therein. Both sets of molecular markers have proved to be mostly neutral in previous studies (7% of markers associated to environmental variables and a fire-related trait, see Budde *et al.* 2014, Jaramillo-Correa *et al.* 2015). Here we used both sets of molecular markers to assess population structure using the Bayesian clustering method implemented in STRUCTURE 2.3.4, running an admixture model with correlated allele frequencies between clusters. Ten runs were performed for each number of clusters from K=1 to K=10, with a burn-in length of 100,000 and a run length of 1,000,000 iterations. The optimal number of population clusters, K, was calculated following Evanno *et al.* (2005). Additionally, we tested for neutral genetic differentiation within each neutral gene pool following the same procedure. Allelic richness and unbiased Nei's index of genetic diversity (H_e ; Nei 1978) for both marker types were computed using SPAGeDi v1.4 (Hardy and Vekemans 2002). F_{ST} was also estimated for both markers at different hierarchical levels (i.e. species, among and within gene pools). F_{ST} distributions were later computed following the parametric bootstrap procedure proposed by Lamy *et al.* (2011), to be compared to Q_{ST} ones (see below).

Quantitative traits

Survival and height were measured three years after establishment in the field. Clonal common gardens allowed multiple phenotypic measurements of each genotype. Recorded phenotypic variables were analyzed with Bayesian mixed models, implemented with MCMCglmm package (Hadfield, 2010) in R (R Core Team, 2013). Data from each common garden was studied separately. We defined block as a fixed factor and population and clone within population as random factors, in order to estimate associated variance components. For height data, we used a Gaussian model, while for survival data, we used a binary model. We applied inverse Wishart priors to both models, but, in the case of survival data, we used parameter expansion to improve mixing, as variance estimates were low and tended to be trapped around zero. By convention, residual variance was fixed to one in binary models, as it is not estimable from data (Sardell *et al.* 2011). Markov Chain Monte Carlo (MCMC) chains were run a minimum of 5.5 million times, sampling every 5,000 and with a burn in period of 500,000 iterations, thus yielding a total of 1,000 estimates per parameter per model. Models

were run longer when needed to keep autocorrelation low (<0.1). Models were validated by checking the normality of residuals and lack of bias. Quantitative genetic parameters (clonal variance and population variance), and experimental Best Linear Unbiased Predictors for clones and populations were derived from these models for their use in subsequent analysis.

We computed pairwise cross-environment correlations among trait estimates for clone and population mean data to estimate Genotype $\times$ Environment (G $\times$ E) interactions. Within-population broad-sense heritability (H^2 , Falconer and Mackay 1996) for survival and height was estimated as:

$$H^2 = (\sigma^2_{pop.clone}) / (\sigma^2_{pop.clone} + \sigma^2_e)$$

where $\sigma^2_{pop.clone}$ is the variance among clones within populations and σ^2_e the error variance. For the estimation of broad-sense heritability of survival, we included an extra term in the denominator ($\pi^2/3$) to account for implicit logit link function variance (Nakagawa and Schielzeth 2010).

Phenotypic plasticity for survival and height was assessed by computing the ratio of the mean trait estimate at the most favourable site minus the mean trait estimate at the least favourable site, divided by the mean for the same trait in the most favourable site (Valladares *et al.* 2006). We estimated the heritability of plasticity as the ratio of the G $\times$ E interaction variance to the total phenotypic variance (Scheiner and Lyman 1989). To estimate the variance component of the interaction, we ran a multi-site two step analysis. First, we extracted error variance estimates from previous individual site models (see above). Next, we used an extended factor restricted maximum likelihood (REML) model implemented in AsReml 3.0 (Gilmour *et al.* 2009), with specified heterogeneous error variances and covariances (drawn from previous models) and heterogeneous variances for random factors, to estimate the variances of the site by clone interaction.

Genetic differentiation for quantitative traits, Q_{ST} (Spitze 1993), was estimated as:

$$Q_{ST} = \frac{\sigma^2_{Bpop}}{\sigma^2_{Bpop} + 2\sigma^2_{Wpop}}$$

where σ^2_{Bpop} and σ^2_{Wpop} are the genetic variance between populations and within populations, respectively. Even though our experimental design did not include families, we were able to compute a broad sense Q_{ST} proxy (Q_{ST}^*), approaching genetic variance from clonal variance estimates. Genetic variance is mainly additive in complex quantitative traits (Hill *et al.* 2008), as reported in maritime pine (Cotterill *et al.* 1987), and thus we expect negligible biases with our approach. Moreover, Q_{ST}^* was computed from common garden data with low environmental heterogeneity, and independently in each location, thus minimizing potential

biases due to $G \times E$ interactions. MCMC models provided 1,000 independent Q_{ST}^* estimates. To disentangle the effects of genetic drift from those of selection, Q_{ST}^* probability distributions were compared to those of neutral F_{ST} , using a nonparametric and free distribution two-sample test for equality of the 2.5 and the 97.5 quantiles with Bonferroni correction for multiple comparisons (Lamy *et al.* 2011). Both Q_{ST}^* and F_{ST} were computed at the species level and within neutral gene pools.

Identification of conservation groups

We considered gene pools revealed with molecular markers (Figure 4.1) as a baseline for conservation needs because they reflect the demographic history of the species. However, we found high quantitative genetic differentiation (Q_{ST}^*) for adaptive traits within those neutral gene pools (see Results). Thus, to identify adaptive population groups that are relevant for genetic conservation within neutral gene pools, we conducted a hierarchical clustering analysis considering nine variables: survival and height at the four common gardens and the plasticity index for height. For this analysis, we excluded populations with five clones or less (Madisouka, Carbonero and Cmpeta) and Rodoiros population, which proved to be non-native, judging by cluster analyses (see Figure S4.1 in Supplementary Information) and historical records. Uncertainty of hierarchical clusters was assessed via multiscale bootstrap resampling, using *pvclust* package (Suzuki and Shimodaira 2006) in R (R Core Team 2015). We run 10,000 bootstraps using the average agglomerative method and correlation as distance measure. The most significant factors for population clustering were identified with the Random Forest package (Liaw and Wiener 2002) attending to mean decrease accuracy and mean decrease in Gini index. These estimates indicate, respectively, how important are individual variables for the correct estimation of the tree, and how much they contribute to the homogeneity of the nodes and leaves in the resulting random forest. We considered all variables randomly sampled as candidates at each split, and grew 10^6 trees. Height in Asturias (i.e. the site with the highest survivorship and best growth rates) was by far the most influential variable (Supplementary Information Figure S4.3) providing the highest among-population differentiation. Then, we also carried out subgroup identification using just height data from the Asturias site. These subgroup definitions within neutral gene pools were based on pairwise post-hoc population comparisons at the $\alpha=0.1$ level, as phenotypic variation within populations is typically large.

Characterization of conservation groups

For each of the population groups identified, we computed heritability and evolvability (the clonal variance to mean ratio for each trait), two key parameters related to short-term evolutionary potential. Genetic variance components for each trait and site were obtained from MCMC models, where population groups were included as a random factor. We also computed Q_{ST}^* , F_{ST} and height plasticity for each conservation-relevant population group as described above. The effective population size (N_e) of each group was assessed for both nuSSRs and SNPs using the heterozygote-excess method of Pudovkin (Pudovkin *et al.* 2009), as implemented in NeEstimator V2 (Do *et al.* 2014). To investigate levels of inbreeding, we also computed a multilocus estimate of selfing for each group using RMES software (David *et al.* 2007).

Climatic zonation

The pan-European genetic conservation strategy for forest trees relies mainly on the climatic classification of Europe (Metzger *et al.* 2005, 2013). We wanted to test the usefulness of broad scale climatic boundaries as a proxy for genetic differentiation in maritime pine. To do so, we overlapped the climatic stratification of Europe (as defined in Lefèvre *et al.* 2014, modified from the original map in Metzger *et al.* 2005), the distribution of maritime pine neutral gene pools, the location of the existing 42 conservation units (see <http://portal.eufgis.org>) and the conservation-relevant population groups proposed here. Additionally, we tested for correlations among height and continuous environmental variables. We retrieved summary climate data for the period 1950–2000 from *Worldclim* (Hijmans *et al.* 2005) and a regional climatic model (Gonzalo-Jiménez 2010) for the 11 non-Spanish and the 24 Spanish maritime pine populations, respectively. Pearson correlation coefficients were estimated between height estimates at each site and nine geographic and environmental variables related to temperature and rainfall (summarized in Supplementary Information Table S4.6). We also run a Principal Component Analysis (PCA) for those geographic and environmental variables, based upon the correlation matrix, with *varimax* rotation. The analysis was implemented by the *psych* package (Revelle 2011) on the R platform (R Core Team 2015). PCA loadings of the three most important PCs were later used for correlation analysis with height.

RESULTS

Molecular markers

Both nuSSRs and SNPs revealed similar genetic differentiation to previous markers (e.g. chloroplast SSRs, Bucci *et al.* 2007), leading to the definition of six extant gene pools (optimal $K=6$): French Atlantic coast, Iberian Atlantic coast, central Spain, southern Spain, Eastern Mediterranean (including Corsica) and Morocco (Supplementary Information Figure S4.1). We found no population structure (optimal $K=1$) within each specific gene pool (data not shown). Maritime pine's H_e for nuSSRs and SNPs was 0.65 ± 0.01 and 0.30 ± 0.00 , respectively. H_e estimates were consistently above 0.23 for nuSSRs and 0.5 for SNP markers, for all tested populations (Table S4.1 and Fig. S4.2 in Supplementary Information) and for the six distinct neutral gene pools (Table 1). Allelic richness ranged from 2.20 to 3.19 for nuSSRs and from 1.39 to 1.63 for SNP markers in the studied populations (for details see Table S4.1 in Supplementary Information), with no significant differences among gene pools (Table 1). F_{ST} estimates at the species level were moderate for both nuSSRs (0.099, CI: 0.085-0.117) and SNP markers (0.130, CI: 0.118-0.142). Among gene pool F_{ST} explained most of the neutral population differentiation (0.070 for nuSSRs and 0.116 for SNP markers). As expected, we detected a reduced level of molecular marker genetic differentiation among populations within all neutral gene pools, except for Morocco (Table 4.1).

Table 4.1. Genetic diversity (allelic richness and Nei's expected heterozygosity, H_e ; Nei, 1978) and population genetic differentiation (F_{ST}) based on 12 nuSSRs and 266 SNPs, for the six extant maritime pine gene pools. SE in brackets.

Gene pool	Allelic richness		Nei's H_e		F_{ST}			
	nuSSRs	SNPs	nuSSRs	SNPs	nuSSRs		SNPs	
French Atlantic	2.55	1.53	0.588	0.262	0.008	(0.004)	0.006	(0.001)
Iberian Atlantic	2.49	1.52	0.581	0.257	0.023	(0.090)	0.025	(0.002)
Central Spain	2.63	1.59	0.603	0.290	0.024	(0.005)	0.022	(0.002)
Southern Spain	2.83	1.60	0.677	0.301	0.045	(0.015)	0.045	(0.006)
Eastern Europe	2.28	1.51	0.520	0.258	0.002	(0.012)	0.016	(0.007)
Morocco	2.36	1.40	0.568	0.192	0.161	(0.094)	0.155	(0.036)
Species	2.52	1.53	0.647	0.303	0.099	(0.000)	0.130	(0.006)

Quantitative traits

Survival ranged from 3% to 98% across the four common gardens (98% CI: 0.97-0.99 in Asturias, 66% CI: 0.60-0.70 in Portugal, 21% CI: 0.18-0.26 in Madrid, and 3% CI: 0.20-0.40 in Cáceres). Mean height per site ranged from 27.7 cm (CI: 25.9-28.8) in the most favourable environment (Asturias) to 17.0 (CI: 14.8-18.9) in the harshest (Cáceres; for details see Table S4.2 in Supplementary Information). Both traits evidenced highly contrasted levels of environmental stress among trial sites. Phenotypic variance for survival was almost entirely driven by environmental variability as suggested by low differentiation across clones. Height thus turned out to be a more suitable adaptive trait than survival to study population differentiation in our study. We identified highly significant site-to-site correlations for both traits among sites with similar environmental conditions (e.g. correlation for height in sites under Atlantic climate: $r = 0.49$, $p < 10^{-10}$, and under Mediterranean climate: $r = 0.24$, $p < 10^{-3}$). Overall, cross-environment correlations were stronger for height, with all correlations being significant ($P < 0.01$; Supplementary Information Table S4.3).

Broad-sense heritability (H^2) varied depending on the trait and trial site. Survival H^2 was very low, ranging from 0.02 (CI: 0.01-0.02) in Portugal to 0.05 (CI: 0.04-0.06) in Asturias. Height H^2 was substantially higher (range of 0.15-0.28), particularly in more stressful environments (e.g. 0.28 CI: 0.13-0.41 in Cáceres). Mediterranean neutral gene pools showed higher H^2 compared to Atlantic ones (see Table S4.4 in Supplementary Information). Quantitative genetic differentiation (Q_{ST}^*) ranged from 0.11 (CI: 0.08-0.16) in Asturias to 0.39 (CI: 0.30-0.48) in Madrid for survival, and from 0.31 (CI: 0.23-0.50) in Portugal to 0.45 (CI: 0.31-0.58) in Asturias for height. In the four sites, overall Q_{ST}^* estimates were significantly higher than F_{ST} estimates at the species level (Figure 4.2). Quantitative population differentiation was still high within neutral gene pools, suggesting the existence of distinct phenotypic groups ($Q_{ST}^* = 0.10$ to 0.87 for height, Figure 4.2; and $Q_{ST}^* = 0.01$ to 0.35 for survival; Supplementary Information Table S4.4).

Figure 4.2. Molecular (F_{ST}) and quantitative (Q_{ST}^*) genetic differentiation among range-wide maritime pine populations for height. Results are given at the species level for the four trial sites and within neutral gene pools in Asturias trial site.

Maritime pine showed high levels of phenotypic plasticity, although variable among neutral gene pools. Phenotypic plasticity for height varied between Atlantic and Mediterranean populations, decreasing from northern to southern gene pools, being the French Atlantic and Iberian Atlantic those with the highest plasticity index (0.96 and 0.93 respectively) and the Moroccan populations those with the least (0.75). Heritability of height plasticity was low (estimate for the combined analysis of 0.06), albeit higher in more stressful environments (in decreasing order of harshness: 0.11 ± 0.04 in Cáceres, 0.08 ± 0.03 in Madrid, 0.07 ± 0.02 in Portugal and 0.05 ± 0.02 in Asturias).

Identification and characterization of conservation groups

Hierarchical cluster analysis within neutral gene pools allowed their further subdivision based on quantitative traits. Among the nine variables used for clustering, height measured in Asturias was the most important (mean decrease accuracy=-113.61, mean decrease in Gini index=1.29, Supplementary Information Figure S4.3). Pairwise post-hoc population comparisons within gene pools for this variable led to very similar population clustering than

with all nine variables. For the three neutral gene pools with more than three populations (French Atlantic, Iberian Atlantic and Central Spain), clustering based on quantitative traits is shown in Figure 4.3. The remaining neutral gene pools (Southern Spain, Eastern Mediterranean and Morocco) exhibited significant differentiation among all populations within the gene pool. Accordingly, the six initial neutral gene pools led to the identification of ten conservation-relevant population groups, subsequently considered as the minimum that should be depicted in a preliminary maritime pine genetic conservation program.

Figure 4.3. Proposed conservation-relevant population groups for maritime pine based on genetic distinctiveness and geographic boundaries of neutral gene pools (a). Clustering based on height and survival as measured in four common gardens with contrasting environments and mean height distribution assessed at Asturias site, for French Atlantic (b), Central Spain (c) and Iberian Atlantic (d) gene pools. Each inset shows a hierarchical clustering dendrogram on the left and a height distribution plot at the population level on the right, with populations represented with broken and solid lines corresponding to different conservation groups.

All conservation-relevant population groups showed high levels of neutral genetic diversity for both nuSSRs (minimum of 0.517 in Pinia) and SNPs (minimum of 0.187 in Tamrabta), reduced levels of inbreeding (Figure 4.4) and high effective population sizes (NeEstimatorv2 gave *infinite* N_e for all groups, i.e. no evidence for variation caused by genetic drift due to a finite number of parents; Do *et al.* 2014). They also showed high evolvability for height and survival in the four trials, and high levels of phenotypic plasticity (Figure 4.4), two key parameters determining their evolutionary potential. As expected, genetic differentiation among populations for neutral markers (F_{ST}) and quantitative traits (Q_{ST}^*) were markedly lower within the proposed conservation-relevant population groups than within the six initial neutral gene pools (Supplementary Information Table S4.5).

Figure 4.4. Characterization of conservation-relevant population groups in maritime pine: Nei's genetic diversity (H_e) and inbreeding coefficient (s) based on 12 nuSSRs and 266 SNPs, plasticity index for height and evolvability in the four common garden sites (Asturias, Portugal, Cáceres and Madrid).

Usefulness of climatic zonation

The entire maritime pine natural range is comprised within six climatic zones (Figure 4.1). The existing genetic conservation network (42 EUFGIS conservation units) comprises populations from the six climatic regions where the species can be found, but do not follow an even pattern within climatic regions (Figure S4.4 in Supplementary Information), showing gaps in the implementation of the pan-European strategy. The Mediterranean North climate is clearly overrepresented (52% of conservation units belong to this region), while the Atlantic Central, Lusitanian and Mediterranean South climates are underrepresented (7, 10 and 10% of total, respectively). Moreover, it only includes populations from four out of the six neutral gene pools identified within the species, with the Eastern Mediterranean group clearly overrepresented, while populations from Iberian Atlantic or Moroccan gene pools are lacking. Indeed, some populations within the same climatic zone belong to different neutral gene pools and, conversely, single neutral gene pools are distributed along distinct climatic zones (Figure 4.1). Climatic boundaries were neither able to distinguish among population groups genetically distinct for adaptive traits within neutral gene pools (i.e. conservation-relevant population groups, as defined here). However, we found significant correlations between height and geographic and environmental variables in trials under Atlantic environments (e.g. $r=0.55$, $p=6.2 \times 10^{-4}$ with annual rainfall in Asturias; $r=-0.37$, $p=2.9 \times 10^{-2}$, with mean temperature of the warmest month in Portugal, Supplementary Information Table S4.6). No significant correlations were found in trials under Mediterranean climates (Madrid and Cáceres), with lower survival. Principal Component Analyses showed that population mean height was highly correlated with PC1 (54% variance explained) in trials with higher trait variability ($r=0.63$ in Asturias, $p=6.5 \times 10^{-5}$; $r=0.49$ in Portugal, $p=3.3 \times 10^{-3}$; see details in Tables S4.7 and S4.8). PC1 was mainly contributed by the continentality index and altitude, thus summarizing environmental harshness.

DISCUSSION

In this study, we used an integrated approach to identify a minimum set of genetically distinct population groups in maritime pine. We departed from gene pools defined using neutral molecular markers and, after incorporating information on genetic variation and adaptive potential in two key quantitative traits, we were able to distinguish ten conservation-relevant population groups. Molecular marker and quantitative trait statistics suggested high levels of genetic diversity, plasticity and evolutionary potential in these population groups. Their representation in a genetic conservation network would increase the chances for long-term maintenance of the species' genetic diversity. Our approach represents a step forward in setting priorities for upcoming conservation efforts, stressing the effectiveness of gradual

integration of genetic information at different scales and time frames as it becomes available for the target species.

Intraspecific conservation in forest trees

Most widespread European forest trees, including maritime pine, are included within the IUCN category of 'least concern' or are still to be evaluated (Farjon 2013). However, the need for intra-specific diversity conservation in forest trees is widely recognized by the research community (Ledig 1986; Kapeller *et al.* 2012 for Norway spruce, Leites *et al.* 2012 for Douglas-fir). Some particular maritime pine populations (e.g. Tramuntana in Balearic Islands or those of subspecies *P. pinaster* ssp. *renoui* in North Africa) are endangered as a result of overexploitation and habitat degradation (Farjon 2013). Differential levels of forest decline have also recently been anticipated for distinct maritime pine gene pools (Jaramillo-Correa *et al.* 2015). In the same line, environmental niche modelling foretell that maritime pine will be able to withstand upcoming climate change, but with contrasted population responses (Benito *et al.* 2011). Population genetic structure analyses based on the studied molecular markers led to the definition of six extant gene pools (Figure 4.1 and Supplementary Information Figure S4.1). This is consistent with the widely known phylogeographic history of the species (Santos-del-Blanco *et al.* 2012, Jaramillo-Correa *et al.* 2015). We found high levels of genetic variation and phenotypic plasticity at the species, gene pool and conservation group levels (Figure 4.4), suggesting potential for evolutionary change to cope with new selection pressures. However, our consistently higher Q_{ST}^* than F_{ST} estimates (Figure 4.2), for both traits in the four common garden sites, point to widespread divergent selection too. Thus, contrasted evolutionary changes are expected in different parts of the maritime pine range, in particular under harsh environmental conditions. All these appeal for the need to further develop genetic conservation strategies at the intraspecific level for maritime pine, as for most other widespread forest trees (e.g. see CONFORGEN in Canada and EUFORGEN in Europe).

Usefulness of molecular markers and quantitative traits in conservation

Thirty years after the first definition of Evolutionary Significant Units by Ryder (1986), with scarce guidelines for operational application, it is still debated how conservation units should be defined and the relative importance that should be given to neutral and adaptive components of genetic variation (Crandall *et al.* 2000; Fraser and Bernatchez 2001). Our proposal is a cascade decision approach, using climatic boundaries when no other information is available, but gradually incorporating molecular marker and phenotypic data as they become available for the species. Extensive studies using molecular markers are now

available, or could easily be in the near future, for many forest trees (e.g. *Eucalyptus*, *Pinus*, *Populus*, *Quercus*; Neale and Kremer 2011). Incorporating such information into conservation planning would already represent a qualitative leap forward from using traditional regions of provenance (Alía *et al.* 1996) or seed zones (Hamann *et al.* 2005). Molecular markers effectively define major cohesive groups with limited gene flow between them (Fraser and Bernatchez 2001). Nonetheless, they are often unsuitable to quantify adaptive variation or evolutionary potential (Reed and Frankham 2001; Ouborg *et al.* 2006), as the ability of species to evolve is primarily determined by quantitative genetic variation (Frankham 2010). Thus, when data are available, phenotypic traits should be used in a second step to differentiate subgroups of populations under different selective pressures, improving previous proposals based on demographic observations or genetic differentiation estimated only with neutral molecular markers (e.g. Petit *et al.* 1998). Sagnard *et al.* (2002) provided one of the few examples combining quantitative traits and marker data in a conservation study in *Abies alba*. Extensive quantitative trait data from traditional common garden experiments (i.e. provenance trials) are already available for many economically and ecologically important tree species (see below and Pâques 2013). This information could be used in a third step to identify worth-protecting populations, but we are unaware of any conservation program for forest trees using this information so far. The described methodology can be applied to any forest tree species, as data gradually become available. However, lack of genetic information could hinder a broad scale implementation of the described approach for species with less economic and ecological importance.

Once conservation-relevant population groups are identified, assessing to which extent they are able to maintain the species' evolutionary potential remains a complex issue. This can also be addressed by integrating molecular marker and quantitative trait data (Aravanopoulos 2011; Lefèvre *et al.* 2014). In maritime pine, the proposed conservation-relevant population groups display high levels of genetic diversity and effective population sizes, and negligible inbreeding (Figure 4.4), suggesting high resilience to new environmental conditions. Low quantitative genetic differentiation among populations within these groups facilitates further selection of specific populations within them as conservations units, while maintaining genetic variation and evolutionary potential. The high phenotypic plasticity found in all conservation-relevant groups supports the available evidence on fast and strong plastic responses of Mediterranean plants to environmental change, with plasticity highly differing, as in our case study, among populations within a species (Matesanz and Valladares 2014). Maritime pine Atlantic populations are more plastic, and height plasticity decreases gradually as we move South within the species range (Figure 4.4). For populations with higher plasticity, the demographic impact of strong selection would be

smooth, allowing more time for evolutionary adaptations to take place and reducing the amount of evolutionary change necessary to track the moving environmental optimum (Chevin *et al.* 2010).

Implications for maritime pine conservation

The Mediterranean basin has been recognized as a biodiversity hotspot and a prime target for conservation efforts (Myers *et al.* 2000; Fady-Welterlen 2005). However, Lefèvre *et al.* (2013) recently detected gaps in genetic conservation efforts for forest trees in the Mediterranean climatic zones. The pan-European genetic conservation strategy for forest trees uses climatic zones (Metzger *et al.* 2005, 2013) as a surrogate to account for adaptive processes. Still, the usefulness of broad scale climatic boundaries as a proxy for genetic differentiation to prioritize populations for genetic conservation remains questionable, even for species exhibiting phenotypic clines along environmental gradients. Our results corroborate the existence of a clinal variation along temperature and rainfall gradients in the species (Jaramillo-Correa *et al.* 2015; Santos-del-Blanco *et al.* 2012), and the important role of the environment in maritime pine local adaptation (Serra-Varela *et al.* 2015). However, neither neutral nor adaptive genetic variability are captured by the climatic zoning of Metzger *et al.* (2005), most likely because of its large spatial scale or inadequate selection of classification variables for maritime pine. Large scale environmental classifications constitute a good starting point, but should be considered with caution when addressing conservation priorities, especially when extensive genetic data is available (e.g. for some of the main forestry species in Europe, including Scots pine, Aleppo pine, common ash, European beech, wild cherry and Norway spruce, among others; see Pâques 2013).

Once conservation-relevant population groups have been identified, as we did in this study, selection of particular populations within these groups should meet the minimum requirements for genetic conservation units developed by EUFORGEN (Koskela *et al.* 2013), as well as national and local use and socio-economic criteria. This highlights the difficulty of descending from conservation-relevant population groups at a regional scale (i.e. conservation planning) to practical conservation at the national or local level (i.e. choosing particular populations and protecting them). We present, nevertheless, a feasible approach to ease the challenge and improve practical conservation by using available genetic information to produce candidate groups of populations to be represented in conservation networks.

The ten conservation-relevant population groups presented here (Figure 4.3) are the result of the integration of all genetic information currently available in maritime pine, but they still constitute the minimum to be considered. Collection of genetic data from unexplored regions

and the inclusion of a wider range of adaptive and fitness-related traits - e.g. pest and disease resistance, germination and seedling establishment ability, and adult survival and reproduction traits - could increase the number of population groups to be considered for effective maritime pine conservation. Although it would be ideal to count on as many phenotypic traits as possible, in this species height has proved to be a highly integrative trait, closely related to resistance to biotic and abiotic factors (Santos-del-Blanco *et al.* 2012; Alía *et al.* 2014; Jaramillo-Correa *et al.* 2015). This made the classification based on height a reliable approximation.

Conservation efforts should additionally consider establishing genetic conservation units in marginal populations growing at distribution edges or under extreme ecological conditions (Fady *et al.* 2015), in particular those that have already shown genetic divergence from nearby core populations (e.g. Fuencaliente in Ciudad Real, Spain). The maritime pine range was traditionally divided into regions of provenance (e.g. Alía *et al.* 1996 for Spain, CEMAGREF 2003 for France) that correspond to areas with uniform ecological conditions and seed sources with putatively (i.e. not experimentally tested) similar phenotypes or genetic features (as evaluated by quantitative genetics, and terpene and isozyme studies) for forest management and afforestation purposes. These regions of provenance could be used to establish geographical boundaries for the defined conservation-relevant population groups. This strategy would work well, for example, in Spain, where a large number of disconnected provenance regions has been established (Alía *et al.* 1996; Supplementary Information Figure S4.5). However, our case study also shows that the Landes and Corsican provenance regions (CEMAGREF 2003) need to be subdivided for conservation purposes (Figure 4.3). This highlights the need for standardization of criteria for selection of conservation units across countries.

DISCUSSION CONCLUDING REMARKS

Quantitative genetics, population genetics and genotype-phenotype association methods made it possible to advance in the understanding of the genetic basis of adaptive traits in maritime and Aleppo pines. Firstly, results in *Chapter 1* led to dismiss a global relationship between heterozygosity and fitness in maritime pine, and suggested that not enough variance in inbreeding depression together with the species demographic history may have decoupled heterozygosity and inbreeding. Throughout the four chapters, I found evidence of intraspecific genetic variation and local adaptation processes in both species. Adaptive traits exhibited a moderate genetic control and were structured across populations. Among-population quantitative genetic differentiation was consistently above that estimated with molecular markers, across traits and testing environments, at different hierarchical levels, suggesting local adaptation (*Chapters 3 and 4*). I also identified putatively adaptive *loci*, underlying important fitness traits in both species, although further studies would be needed to confirm if they are located in functional genes (*Chapters 2 and 3*). Results point to many *loci* of small effect underlying the studied traits, with local adaptation driven by small allele frequency changes. Environmental factors influenced genetic variability and the detected genetic architecture of the studied adaptive traits in both pine species (*Chapters 2, 3 and 4*).

Results are broadly discussed at the end of each chapter, integrating different spatial scales and testing environments with previous knowledge on the species' biology, evolutionary and demographic history. Here I just present a brief concluding remark to sum up what has been discussed previously in the thesis chapters.

Local effects drive heterozygosity-fitness correlations in maritime pine

The study in *Chapter 1* suggests that genome-wide heterozygosity has little impact on maritime pine fitness, even though variation in inbreeding seems to be associated with survival in the species. However, a novel application of a genetic association test allowed to detect heterozygosity at some putatively adaptive markers to be associated with survival, revealing that local fitness effects may drive HFCs in maritime pine. Nevertheless, because the SNP assay was enriched for polymorphisms from candidate genes, this assertion must be taken with caution. HFCs were stable across distinct maritime gene pools pointing to broad, species-wide effects on fitness. Moreover, significant HFCs were found in common gardens under harsh Mediterranean climatic conditions, suggesting that heterozygosity at these *loci* is more important for survival under increased selection pressures. Thus, heterozygosity in specific candidate genes may hold a relevant role for survival of forest tree populations under warmer and drier climates, as those expected under current climate

change. This study highlights the utmost importance of integrating knowledge on species evolutionary and demographic history, across different spatial scales and testing environments, in the estimation and interpretation of HFCs, although commonly overlooked in previous studies.

Genetic basis of adaptive traits in Aleppo pine

Results in *Chapter 2* revealed a number of interesting correlations among traits related to growth, drought resistance, reproduction and defense at individual and population levels in Aleppo pine. I also tested the relationship between ecogeographical variables and the different adaptive phenotypes. Overall, I found supporting evidence for intraspecific local adaptation and that allocation to adaptive traits related to growth, reproduction and defense trade-off against each other, matching theoretical predictions from allocation and life-history theories in the species.

In addition, at the molecular marker level, I unveiled several associations that provide insights on the genetic basis of the studied adaptive traits. I found fifteen SNPs significantly associated with growth traits, wood extractives, lignin content and flowering, explaining a moderate variation in phenotypic traits. Results show that integrated phenotypes in genotype-phenotype association tests can slightly enhance the proportion of phenotypic variance explained, as well as giving the chance of detecting possible pleiotropic effects. Association tests using multi-trait phenotypes corresponding to the main PCs from a PCA, revealed one SNP to be significantly associated with a PC mainly influenced by wood extractive traits. The same SNP was associated with dichloromethane, ethanol and total extractives in single-locus and multi-locus association tests. This result pointing to pleiotropy, must be however considered with caution, as the method cannot confirm that the marker locus affects all components of the integrated phenotype directly.

Phenotypic trade-offs are commonly overlooked in association studies considering multiple traits. Nevertheless, if previous quantitative analysis reveal high covariance among traits, it is advisable to prioritize increasing the sampling effort in association studies, instead of the number of traits phenotyped, or, if possible, directly use integrated phenotypes. New insights in discovering the genetic architecture of adaptive traits may come from jointly testing for associations in multiple traits that are correlated due to shared (but hidden) genetic or non-genetic factors in a single multivariate association analyses. Integration of quantitative genetic knowledge and novel genotype-phenotype association methods is essential to unravel the genetic basis of adaptive traits in long-lived tree species, which in turn will determine forest's capacity to cope with upcoming environmental conditions derived from human-driven climate change.

Polygenic adaptation in maritime pine

Individual SNPs underlying adaptive traits are difficult to identify, but even when only a few significantly associated loci were detected in genotype-phenotype association tests in maritime pine, we were able to identify strong signals of polygenic adaptation for fitness traits. I combined population and association genetics, and tested for excess variance in among-population differentiation and signals of excessive correlations with environmental variables. Results show an excess of variance among populations for height and survival, consistent with directional selection, mainly driven by covariance among *loci* (i.e. linkage disequilibrium). *Chapter 3* supports the polygenic view of adaptation in forest trees, which most likely happens through many allele shifts of small effect, individually undetectable (Rockman 2012; Le Corre and Kremer 2012). The study highlights the advantages of a polygenic perspective, and advocates for multiple complementary approaches when looking for signals of selection (as recently suggested by Keller *et al.* 2012; Tiffin and Ross-Ibarra 2014). Seeing adaptation through a polygenic lens represents a shift from previous local adaptation studies in trees and can be the key for a better understanding of the genetic basis of past, ongoing and future adaptive processes in our forests.

Maritime pine conservation genetics

The study in *Chapter 4* is the first study that considers jointly molecular marker, quantitative trait and environmental variation to address forest tree conservation. I introduce key molecular and quantitative trait statistics to assess population differentiation and identify sound population groups for dynamic conservation purposes. Results bear out that environmental variation is a good starting point for *in-situ* gene conservation strategies, but neither a purely climatic nor a neutral molecular marker classification would capture all variability that should be considered in conservation planning. Current conservation strategies fail to integrate available genetic information, despite its proving effective to identify populations worth protecting in tree species with complex evolutionary histories. Further work is needed to select particular populations for the establishment of genetic conservation units in maritime pine, but dissemination of first-step case studies, as the one presented here in *Chapter 4*, represents a step forward in the urgent need to bridge the gap between basic research in ecology and evolution and the practical needs of conservation managers (Mace and Purvis 2008; Shafer *et al.* 2015). Incorporation of quantitative genetic information in conservation programs for long-lived organisms such as forest trees is challenging and complex, but precise and integrated demographic, neutral and adaptive genetic diversity data are essential to design effective tree dynamic conservation strategies to minimize biodiversity losses in the face of climate change.

CONCLUSIONS

1. The combined study of phenotypes and genotypes, using both well-established methods in quantitative and population genetics and new genotype-phenotype association tests, provides key insights to understand past adaptive processes in forest trees. It also helps to disentangle genetic responses from plasticity, as well as to identify *loci* associated with adaptive traits and/or environmental variables driving selection, for a better prediction of likely future forest evolutionary processes, which can be incorporated into conservation planning (*Chapters 1 to 4*).
2. Genome-wide heterozygosity is not correlated with survival in *Pinus pinaster* at any spatial level (i.e. species, gene pool or population). However, heterozygosity at specific candidate genes may play a significant role for increasing fitness species-wide, especially under increased selection pressures such as the ones derived from climate change (*Chapter 1*).
3. Considering the species evolutionary and demographic history and different spatial scales and testing environments, as well as an adequately high number of molecular markers is of key importance when assessing and interpreting heterozygosity-fitness correlations (*Chapter 1*).
4. In *P. halepensis*, local adaptation is suggested by the correlation between ecogeographical variables and phenotypes integrated by multiple adaptive traits. Allocation to adaptive traits related to growth, reproduction and defence trade-off against each other, matching theoretical predictions from allocation and life history theories (*Chapter 2*).
5. Integrated phenotypes including different traits related to fitness (i.e. growth, stress response and reproduction) can be useful to identify pleiotropy by using genotype-phenotype association tests, although significant spurious associations due to (non-functional) phenotypic correlations among traits cannot be discarded (*Chapter 2*).
6. A number of significant molecular markers underlying key adaptive traits such as survival and tree growth were found in *P. pinaster* and *P. halepensis* using candidate-gene association tests, showing that at least part of the variance for fitness traits can be explained by genes with detectable effects (*Chapters 2 and 3*).

7. Although candidate-gene association tests proved effective to identify some of the main *loci* underlying adaptive traits in *Pinus pinaster* and *Pinus halepensis*, the explained phenotypic variance remained low, supporting a polygenic quantitative model (*Chapters 2 and 3*).
8. In *Pinus pinaster*, there is a polygenic adaptation signal for survival and height, stable across contrasted testing environments and successive years, driven mainly by linkage disequilibrium. Polygenic adaptation is thus most likely than selective sweeps to drive local adaptation in the species (*Chapter 3*).
9. Environmental data, neutral molecular marker differentiation and standing phenotypic variation in adaptive traits should be jointly considered and integrated in managers' decision criteria to choose target populations for dynamic conservation, especially in species with population structure and complex evolutionary histories such as maritime pine (*Chapter 4*).
10. The gradual integration of molecular marker and phenotypic data, as they become available for a given species, accounting for both key neutral and adaptive evolutionary components, would ensure conservation of intraspecific genetic diversity and could improve current forest tree conservation strategies and forest sustainability under the current climate change scenario (*Chapter 4*).

REFERENCES

- Abrahamsson S. 2011 Genetic dissection of quantitative traits in Scots pine. PhD Thesis, Swedish University of Agricultural Sciences, Umea, Sweden.
- Agudo R, Carrete M, Alcaide M, Rico C, Hiraldo F, Donazar JA. 2012. Genetic diversity at neutral and adaptive loci determines individual fitness in a long-lived territorial bird. *Proceedings of the Royal Society B: Biological Sciences* 279: 3241–3249.
- Ahuja M, Neale DB. 2005. Evolution of genome size in conifers. *Silvae genetica* 3: 126–137.
- 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.
- Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R, *et al.* 2013. Potential for evolutionary responses to climate change - evidence from tree populations. *Global Change Biology* 19: 1645 – 1661.
- Alho JS, Välimäki K, Merilä J. 2010. *Rhh*: an R extension for estimating multilocus heterozygosity and heterozygosity-heterozygosity correlation. *Molecular Ecology Resources* 10: 720–722.
- Alía R, Chambel R, Notivol E, Climent J, González-Martínez SC. 2014. Environment-dependent microevolution in a Mediterranean pine (*Pinus pinaster* Aiton). *BMC Evolutionary Biology* 14: 200.
- Alía R, Majada J. 2013. Phenotyping for the future and the future of phenotyping. *In*: Novel Tree Breeding. Lee S and Woolliams J Ed., INIA, Madrid, Spain.
- Alía R, Martín S, De-Miguel J, Galera R, Agúndez D, Gordo J, Salvador L, Catalán G, Gil L. 1996. Las regiones de procedencia de *Pinus pinaster* Aiton. Organismo Autónomo de Parques Nacionales, Madrid, Spain.
- Alía R, Martín S. 2003. EUFORGEN: Technical guidelines for genetic conservation and use for Maritime pine (*Pinus pinaster*). International Plant Genetic Resources Institute, Rome, Italy.
- Alía R, Moro J, Denis JB. 1997. Performance of *Pinus pinaster* provenances in Spain: interpretation of the genotype by environment interaction. *Canadian Journal of Forest Research* 27: 1548–1559.
- Alley RB. 2003. Abrupt Climate Change. *Science* 299: 2005–2010.
- Alvarez JM, Cortizo M, Bueno N, Rodríguez A, Ordás RJ. 2013. CLAVATA1-LIKE, a leucine-rich-repeat protein receptor kinase gene differentially expressed during adventitious caulogenesis in *Pinus pinaster* and *Pinus pinea*. *Plant Cell, Tissue and Organ Culture* 112: 331–342.
- Alves A, Santos A, Rozenberg P, Pâques LE, Charpentier J-P, Schwanninger M, Rodrigues J. 2010. A common near infrared—based partial least squares regression model for the prediction of wood density of *Pinus pinaster* and *Larix x eurolepis*. *Wood Science and Technology* 46: 157–175.

- Amos W, Acevedo-Whitehouse K. 2009. A new test for genotype-fitness associations reveals a single microsatellite allele that strongly predicts the nature of tuberculosis infections in wild boar. *Molecular Ecology Resources* 9: 1102–1111.
- Andersen SB. 2005. Haploids in the improvement of woody species. *In* Haploids in crop improvement II. Palmer WA, Keller W & Kasha K. Springer, Berlin, Germany.
- Anderson JT, Lee CR, Rushworth CA, Colautti RI, Mitchell-Olds T. 2013. Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology* 22: 699–708.
- Aravanopoulos FA. 2011. Genetic monitoring in natural perennial plant populations. *Botany* 89: 75–81.
- Armbruster P, Reed DH. 2005. Inbreeding depression in benign and stressful environments. *Heredity* 95: 235–242.
- Armbruster WS, Pélabon C, Bolstad GH, Hansen TF. 2014. Integrated phenotypes: understanding trait covariation in plants and animals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 369: 20130245.
- Aschard H, Vilhjálmsson BJ, Greliche N, Morange PE, Trégouët DA, Kraft P. 2014. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *American Journal of Human Genetics* 94: 662–676.
- Atwell S, Huang YS, Vilhjálmsson 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 *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.
- Balding DJ. 2006. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* 7: 781–91.
- Ballesta P, Mora F, Ruiz E, Contreras-Soto R. 2014. Marker-trait associations for survival, growth, and flowering components in *Eucalyptus cladocalyx* under arid conditions. *Biologia Plantarum* 59: 389–392.
- Balloux F, Amos W, Coulson T. 2004. Does heterozygosity estimate inbreeding in real populations? *Molecular ecology* 13: 3021–3031.
- Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nature Review Genetics* 12: 767–780.
- Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67: 1–48.
- Beaulieu J, Doerksen T, Boyle B, Clément S, Deslauriers M, Beauseigle S, Blais S, Poulin P-L, Lenz P, Caron S, *et al.* 2011. Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. *Genetics* 188: 197–214.
- Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13: 969–980.
- Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F. 2012. Impacts of climate change on the future of biodiversity. *Ecology Letters* 15: 365–377.

- Benito M, Alía R, Robson TM, Zavala MA. 2011. Intra-specific variability and plasticity influence potential tree species distributions under climate change. *Global Ecology and Biogeography* 20: 766–778.
- Berg JJ, Coop G. 2014. A Population Genetic Signal of Polygenic Adaptation. *PLoS genetics* 10: e1004412.
- Bloom JS, Ehrenreich IM, Loo WT, Lite TLV, Kruglyak L. 2013. Finding the sources of missing heritability in a yeast cross. *Nature* 494: 234–237.
- Bouffier L, Raffin A, Alia R. 2013. Maritime pine – *Pinus pinaster* Ait. In *Best Practice for Tree Breeding in Europe*. Mullin TJ & Lee SJ, ed. Skogforsk, Sweden: 65–75.
- 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.
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences U.S.A.* 101: 15255–15260.
- Bucci G, Anzidei M, Madaghiele A, Vendramin GG. 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, Plomion C, Ribeiro MM, Sebastiani F, Alía R, Vendramin GG. 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular ecology* 16: 2137–2153.
- Budde KB, Heuertz M, Hernández-Serrano A, Pausas JG, Vendramin GG, Verdú M, González-Martínez SC. 2014. In situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). *New Phytologist* 201: 230–241.
- Burban C, Petit RJ. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* 12: 1487–1495.
- Bush RM, Smouse PE, Ledig FT. 1987. The fitness consequences of multiple-locus heterozygosity: the relationship between heterozygosity and growth rate in pitch pine (*Pinus rigida* Mill.). *Evolution* 41: 787–798.
- Byers DL, Waller DM. 1999. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. *Annual Review of Ecology, Evolution, and Systematics* 30: 479–513.
- Cardon LR, Bell JL. 2001. Association study designs for complex diseases. *Nature reviews. Genetics* 2: 91–99.
- CEMAGREF. 2003. Conseils d'utilisation des matériels forestiers de reproduction. Régions de provenance et variétés améliorées. CEMAGREF, Paris, France.
- Chambel MR, Climent J, Alía R. 2007. Divergence among species and populations of Mediterranean pines in biomass allocation of seedlings grown under two watering regimes. *Annals of Forest Science* 64: 87–97.

- Chambel MR, Climent J, Pichot C, Ducci F. 2013. Chapter 5: Mediterranean Pines (*Pinus halepensis* Mill. and *brutia* Ten.). In *Forest Tree Breeding in Europe: Current State-of-the-Art and Perspectives*. Springer, Dordrecht, Netherlands: 229–265.
- Chancerel E, Lamy J-B, Lesur I, Noirot C, Klopp C, Ehrenmann F, Boury C, Le Provost G, Label P, Lalanne C, *et al.* 2013. High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC biology* 11: 50.
- Chapman JR, Nakagawa S, Coltman D, Slate J, Sheldon B. 2009. A quantitative review of heterozygosity-fitness correlations in animal populations. *Molecular Ecology* 18: 2746–2765.
- Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* 10: 783–796.
- Chen J, Zhang D, Zhang C, Xia X, Yin W, Tian Q. 2015. A Putative PP2C-encoding gene negatively regulates ABA signaling in *Populus euphratica*. *PLoS ONE* 10: 1–13.
- Chevin L-M, Collins S, Lefèvre F. 2012. Phenotypic plasticity and evolutionary demographic responses to climate change: taking theory out to the field (A Hoffmann, Ed.). *Functional Ecology*.
- Chevin LM, Hospital F. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* 180: 1645–1660.
- Chevin LM, Lande R, Mace GM. 2010. Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. *PLoS Biology* 8:e1000357.
- Climent J, Calama R, Chambel MR, Sanchez de Ron D, Alía R. 2008. To grow or to seed: ecotypic variation in reproductive allocation and cone production by young female Aleppo pine (*Pinus halepensis*, pinaceae). *American Journal of Botany* 95: 833–842.
- Coltman DW, Slate J. 2003. Microsatellite measures of inbreeding: a meta-analysis. *Evolution* 57: 971–983.
- Comeault A, Soria-Carrasco V, Gompert Z, Farkas TE, Buerkle CA, Parchman TL, Nosil P. 2014. Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *The American naturalist* 183: 711–727.
- Conner JK, Hartl DL. 2004. What is Ecological Genetics? In *A primer of ecological genetics*. Sinauer Associates Inc, Sunderland, Massachusetts: 1–8.
- Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. *PLOS Biology* 13: e1002112.
- Corlett RT, Westcott DA. 2013. Will plant movements keep up with climate change? *Trends in Ecology and Evolution* 28: 482–488.
- Correia I, Almeida MH, Aguiar A, Alía R, David TS, Pereira JS. 2008. Variations in growth, survival and carbon isotope composition ($\delta^{13}\text{C}$) among *Pinus pinaster* populations of different geographic origins. *Tree Physiology* 28: 1545–1552.
- Cosgrove DJ. 2005. Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology* 6: 850–861.

- Cotterill PP, Dean CA, van Wyk G. 1987. Additive and dominance effects in *Pinus pinaster*, *P. radiata* and *P. elliotii* and some implications for breeding strategi. *Silvae Genetica* 36: 221–232.
- Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK. 2000. Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution* 15: 290–295.
- Cregg BM, Zhang JW. 2001. Physiology and morphology of *Pinus sylvestris* seedlings from diverse sources under cyclic drought stress. *Forest Ecology and Management* 154: 131–139.
- 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.
- Da Silva A, Gaillard JM, Yoccoz NG, Hewison JM, Galan M, Coulson T, Allainé D, Vial L, Delorme D, Van Laere G, *et al.* 2009. Heterozygosity-fitness correlations revealed by neutral and candidate gene markers in roe deer from a long-term study. *Evolution* 63: 403–417.
- Dabney A, Storey JD. 2012. Bioconductor's qvalue package: Q-value estimation for false discovery rate control. <http://stat.R/CRAN/doc/packages/qvalue.pdf>
- David P, Pujol B, Viard F, Castella V, Goudet J. 2007. Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology* 16: 2474–2487.
- David P. 1998. Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* 80: 531–537.
- Davis MB, Shaw RG. 2001. Range shifts and adaptive responses to quaternary climate change. *Science* 292: 673–679.
- De La Mata R, Merlo E, Zas R. 2014. Among-population variation and plasticity to drought of atlantic, mediterranean, and interprovenance hybrid populations of maritime pine. *Tree Genetics and Genomes* 10: 1191–1203.
- De la Torre AR, Birol I, Bousquet J, Ingvarsson PK, Jansson S, Jones SJM, Keeling CI, MacKay J, Nilsson O, Ritland K, *et al.* 2014. Insights into Conifer Giga-Genomes. *Plant Physiology* 166: 1724–1732.
- De Lucas AI, González-Martínez SC, Hidalgo H, Bravo F, Heuertz M. 2009. Admixture, one-source colonization or long-term persistence of maritime pine in the Castilian Plateau? Insights from nuclear microsatellite markers. *Investigación Agraria: Sistemas y Recursos Forestales* 18: 3–12.
- De Lucas AI, Robledo-Arnuncio JJ, Hidalgo E, González-Martínez SC. 2008. Mating system and pollen gene flow in Mediterranean maritime pine. *Heredity* 100: 390–399.
- De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* 22: 1383–1399.
- De Woody YD, De Woody JA. 2005. On the estimation of genome-wide heterozygosity using molecular markers. *Heredity* 96: 85–88.

- Dillon SK, Nolan M, Li W, Bell C, Wu HX, Southerton SG. 2010. Allelic variation in cell wall candidate genes affecting solid wood properties in natural populations and land races of *Pinus radiata*. *Genetics* 185: 1477–1487.
- Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. 2014. *NeEstimator* v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources* 14: 209–214.
- Durel CE, Bertin P, Kremer A. 1996. Relationship between inbreeding depression and inbreeding coefficient in maritime pine (*Pinus pinaster*). *Theoretical and Applied Genetics* 92: 347–356.
- Dutkowski G, Costa e Silva J, Gilmour AR, Wellendorf H, Aguiar A. 2006. Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Canadian Journal of Forest Research* 36: 1851–1870.
- Earl DA, Von Holdt BM. 2011. *STRUCTURE HARVESTER*: a website and program for visualizing *STRUCTURE* output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.
- Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky K V, 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, Wegrzyn JL, Cumbie WP, Goldfarb B, Huber D a, Tolstikov V, Fiehn O, Neale DB. 2012. Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. *New Phytologist* 193: 890–902.
- Eriksson G, Namkoong G, Roberds JH. 1993. Dynamic gene conservation for uncertain futures. *Forest Ecology and Management* 62: 15–37.
- Esteban LG, Martín JA, de Palacios P, Fernández FG, López R. 2010. Adaptive anatomy of *Pinus halepensis* trees from different Mediterranean environments in Spain. *Trees Structure and Function* 24: 19–30.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.
- Eveno E, Collada C, Guevara MA, Léger V, Soto A, Díaz 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.
- Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103: 285–298.
- Fady B, Cottrell J, Ackzell L, Alía R, Muys B, Prada A, González-Martínez SC. 2015. Forests and global change: what can genetics contribute to the major forest management and policy challenges of the twenty-first century? *Regional Environmental Change* 16: 927–939.
- Fady B, Semerci H, Vendramin GG. 2003. EUFORGEN Technical guidelines for genetic conservation and use for Aleppo pine (*Pinus halepensis*) and Brutia pine (*Pinus brutia*). International Plant Genetic Resources Institute, Rome, Italy.

- Fady-Welterlen B. 2005. Is there really more biodiversity in Mediterranean forest ecosystems? *Taxon* 54: 905–910.
- Falconer DS, Mackay TFC. 1996. Introduction to Quantitative Genetics. Longman, Essex, UK.
- Falconer DS. 1989. Introduction to quantitative genetics. Longman Scientific and Technical, Essex, UK.
- Falster DS, Westoby M. 2003. Plant height and evolutionary games. *Trends in Ecology & Evolution* 18: 337–343.
- FAO. 2015. Global Forest Resources Assessment 2015. Rome, Italy.
- Farjon A. 2013. *Pinus pinaster*. The IUCN Red List of Threatened Species. Version 2014.2. <http://www.iucnredlist.org/details/summary/42390/1>
- Farquhar G, O'Leary M, Berry J. 1982. On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentration in leaves. *Australian Journal of Plant Physiology* 9: 121.
- Ferrio JP, Florit A, Vega A, Serrano L, Voltas J. 2003. $\Delta^{13}\text{C}$ and tree-ring width reflect different drought responses in *Quercus ilex* and *Pinus halepensis*. *Oecologia* 137: 512–518.
- Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180: 977–993.
- 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.
- Fournier-Level A, Wilczek AM, Cooper MD, Roe JL, Anderson J, Eaton D, Moyers BT, Petipas RH, Schaeffer RN, Pieper B, et al. 2013. Paths to selection on life history loci in different natural environments across the native range of *Arabidopsis thaliana*. *Molecular Ecology* 22: 3552–3566.
- Fox CW, Reed DH. 2011. Inbreeding depression increases with environmental stress: an experimental study and meta-analysis. *Evolution* 65: 246–258.
- Frankham R. 2010. Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation* 143: 1919–1927.
- Fraser DJ, Bernatchez L. 2001. Adaptive evolutionary conservation: towards a unified concept for defining coservation units. *Molecular ecology* 10: 2741–2752.
- 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.
- Furlotte NA, Eskin E. 2015. Efficient multiple trait association and estimation of genetic correlation using the matrix-variate linear mixed-model. *Genetics* 200: 59–68.
- García-Gutiérrez A, Cantón FR, Gallardo F, Sánchez-Jiménez F, Cánovas FM. 1995. Expression of ferredoxin-dependent glutamate synthase in dark-grown pine seedlings. *Plant Molecular Biology* 27: 115–128.

- García-Navas V, Cáliz-Campal C, Ferrer ES, Sanz JJ, Ortego J. 2014. Heterozygosity at a single locus explains a large proportion of variation in two fitness-related traits in great tits: a general or a local effect? *Journal of Evolutionary Biology* 27: 2807–2819.
- Gauli A, Vaillancourt RE, Bailey TG, Steane DA, Potts BM. 2015. Evidence for local climate adaptation in early-life traits of Tasmanian populations of *Eucalyptus pauciflora*. *Tree Genetics and Genomes* 11: 104.
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R. 2009. ASReml User Guide Release 3.0. VSN International Ltd., Hemel Hempstead, UK. www.vsnl.co.uk
- González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, Moreno V. 2007. *SNPassoc*: an R package to perform whole genome association studies. *Bioinformatics* 23: 644–645.
- González-Martínez SC, Alía R, Gil L. 2002. Population genetic structure in a Mediterranean pine (*Pinus pinaster* Ait.): a comparison of allozyme markers and quantitative traits. *Heredity* 89: 199–206.
- González-Martínez SC, Burczyk J, Nathan R, Nanos N, Gil L, Alía R. 2006. Effective gene dispersal and female reproductive success in Mediterranean maritime pine (*Pinus pinaster* Aiton). *Molecular Ecology* 15: 4577–4588.
- González-Martínez SC, Ersoz ES, Brown GR, Wheeler NC, Neale DB. 2006. DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172: 1915–1926.
- González-Martínez SC, Huber D, Ersoz ES, Davis JM, Neale DB. 2008. Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* 101: 19–26.
- González-Martínez SC, Mariette S, Ribeiro MM, Burbank C, Raffin A, Chambel MR, Ribeiro CAM, Aguiar A, Plomion C, Alía R, *et al.* 2004. Genetic resources in maritime pine (*Pinus pinaster* Aiton): molecular and quantitative measures of genetic variation and differentiation among maternal lineages. *Forest Ecology and Management* 197: 103–115.
- González-Martínez SC, Wheeler NC, Ersoz ES, Nelson CD, Neale DB. 2007. Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175: 399–409.
- Gonzalo-Jiménez J. 2010. Diagnóstico fitoclimático de la España Peninsular: hacia un modelo de clasificación funcional de la vegetación y de los ecosistemas peninsulares españoles. Organismo Autónomo de Parques Nacionales, Madrid, Spain.
- Goudet J. 2005. *HIERFSTAT*, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes* 5: 184–186.
- Govindaraju DR, Larson MG, Yin X, Benjamin EJ, Rao MB, Vasan RS. 2009. Association between SNP heterozygosity and quantitative traits in the Framingham Heart Study. *Annals of Human Genetics* 73: 465–473.
- Grattapaglia D, Plomion C, Kirst M, Sederoff RR. 2009. Genomics of growth traits in forest trees. *Current Opinion in Plant Biology* 12: 148–156.
- Graudal L, Aravanopoulos F, Bennadji Z, Changtragoon S, Fady B, Kjær ED, Loo J, Ramamonjisoa L, Vendramin GG. 2014. Global to local genetic diversity indicators of evolutionary potential in tree species within and outside forests. *Forest Ecology and Management* 333: 35–51.

- Grime JP. 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *The American Naturalist* 111: 1169–1194.
- Grivet D, Climent J, Zabal-Aguirre M, Neale DB, Vendramin GG, González-Martínez SC. 2013. Adaptive evolution of Mediterranean pines. *Molecular Phylogenetics and Evolution* 68: 555–566.
- 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 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.
- Groover AT. 2007. Will genomics guide a greener forest biotech? *Trends in Plant Science* 12: 234–238.
- Grueber CE, Waters JM, Jamieson IG. 2011. The imprecision of heterozygosity-fitness correlations hinders the detection of inbreeding and inbreeding depression in a threatened species. *Molecular Ecology* 20: 67–79.
- Guan Y, Stephens M. 2011. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics* 5: 1780–1815.
- Guevara MA, Soto A, Collada C, Plomion C, Savolainen O, Neale DB, González-Martínez SC, Cervera MT. 2005. Genomics applied to the study of adaptation in pine species. *Investigación Agraria: Sistemas y Recursos Forestales* 14: 292–306.
- Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. *Genetics* 195: 205–20.
- Guy RD, Holowachuk DL. 2001. Population differences in stable carbon isotope ratio of *Pinus contorta* Dougl. ex Loud.: relationship to environment, climate of origin, and growth potential. *Canadian Journal of Botany* 79: 274–283.
- Hadfield J, Wilson AJ, Garant D, Sheldon BC, Kruuk LEB. 2010. The misuse of BLUP in ecology and evolution. *The American naturalist* 175: 116–125.
- Hadfield J. 2010. MCMC methods for multi-response generalized linear mixed models: the *MCMCglmm* R package. *Journal of Statistical Software* 33: 1–22.
- Hamann A, Smets P, Yanchuk AD, Aitken SN. 2005. An ecogeographic framework for *in situ* conservation of forest trees in British Columbia. *Canadian Journal of Forest Research* 35: 2553–2561.
- Hampe A, Petit RJ. 2005. Conserving biodiversity under climate change: the rear edge matters. *Ecology letters* 8: 461–467.
- Hamrick JL, Godt MJW, Sherman-Broyles SL. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests* 6: 95–124.

- Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, Sukernik R, Utermann G, Pritchard J, Coop G, *et al.* 2010. Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the National Academy of Sciences* 107: 8924–8930.
- Hansson B, Westerberg L. 2002. On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* 11: 2467–2474.
- 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.
- Hayton JC, Allen DG, Scarpello V. 2004. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organizational Research Methods* 7: 191–205.
- Hedrick PW. 2012. What is the evidence for heterozygote advantage selection? *Trends in Ecology and Evolution* 27: 698–704.
- Heuertz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N. 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174: 2095–2105.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
- Hill WG, Goddard ME, Visscher PM. 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* 4: e1000008.
- Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6: 95–108.
- Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, Lacy RC, Dasmahapatra KK. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. *Proceeding of the National Academy of Sciences U.S.A.* 111: 3775–3780.
- 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.
- Houle D, Govindaraju DR, Omholt S. 2010. Phenomics: the next challenge. *Nature Reviews Genetics* 11: 855–866.
- Houle D. 1991. Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters. *Evolution in Health and Disease* 45: 630–648.
- Howe G, Aitken S. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Journal of Botany* 1266: 1247–1266.
- Hwang EY, Song Q, Jia G, Specht JE, Hyten DL, Costa J, Cregan PB. 2014. A genome-wide association study of seed protein and oil content in soybean. *BMC genomics* 15: 1.

- Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S. 2008. Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 Locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* 178: 2217–2226.
- Ingvarsson PK, Street N. 2011. Association genetics of complex traits in plants. *New Phytologist* 189: 909–922.
- Ishizaki T, Ohsumi C, Totsuka K, Igarashi D. 2010. Analysis of glutamate homeostasis by overexpression of Fd-GOGAT gene in *Arabidopsis thaliana*. *Amino Acids* 38: 943–950.
- Isik F, Mora CR, Schimleck LR. 2011. Genetic variation in *Pinus taeda* wood properties predicted using non-destructive techniques. *Annals of Forest Science* 68: 283–293.
- Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806.
- Jaramillo-Correa JP, Prunier J, Vázquez-Lobo A, Keller SR, Moreno-Letelier A. 2015. Molecular signatures of adaptation and selection in forest trees. *Advances in Botanical research* 74: 265–306.
- Jaramillo-Correa JP, Rodríguez-Quilón I, Grivet D, Lepoittevin C, Sebastiani F, Heuertz M, Garnier-Géré P, Alía R, Plomion C, Vendramin GG, *et al.* 2015. Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics* 199: 793–807.
- 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.
- Kang HM, Sul JJHJ, Zaitlen N, Kong S-Y, Service SK, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* 42: 348–354.
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. 2008. Efficient control of population structure in model organism association mapping. *Genetics* 178: 1709–1723.
- Kapeller S, Lexer MJ, Geburek T, Hiebl J, Schueler S. 2012. Intraspecific variation in climate response of Norway spruce in the eastern Alpine range: Selecting appropriate provenances for future climate. *Forest Ecology and Management* 271:46–57.
- Kass RE, Raftery AE. 1995. Bayes Factors. *Journal of the American Statistical Association* 90: 773–795.
- 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.
- Keller SR, Levensen N, Olson MS, Tiffin P. 2012. Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution* 29: 3143–3152.

- Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. 2014. Selection for complex traits leaves little or no classic signatures of selection. *BMC genomics* 15: 246.
- Koskela J, Buck A, Tessier du Cros E. 2007. Climate change and forest genetic diversity: Implications for sustainable forest management in Europe. Bioversity International, Rome, Italy.
- Koskela J, Lefèvre F, Schueler S, Kraigher H, Olrik DC, Hubert J, Longauer R, Bozzano M, Yrjänä L, Alizoti P, *et al.* 2013. Translating conservation genetics into management: Pan-European minimum requirements for dynamic conservation units of forest tree genetic diversity. *Biological Conservation* 157: 39–49.
- Krauskopf E, Harris PJ, Putterill J. 2005. The cellulose synthase gene *PrCESA10* is involved in cellulose biosynthesis in developing tracheids of the gymnosperm *Pinus radiata*. *Gene* 350: 107–116.
- Laikre L, Allendorf FW, Aroner LC, Baker CS, Gregovich DP, Hansen MM, Jackson JA, Kendall KC, McKelvey K, Neel MC, *et al.* 2010. Neglect of genetic diversity in implementation of the convention on biological diversity. *Conservation Biology* 24: 86–88.
- Lambert AB, Bauer F, Don D. 1803. A description of the genus *Pinus*. J. White Ed., London UK.
- Lamy JB, Bouffier L, Burlett R, Plomion C, Cochard H, Delzon S. 2011. Uniform selection as a primary force reducing population genetic differentiation of cavitation resistance across a species range. *PloS one* 6: e23476.
- Latta RG. 2003. Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist* 161: 51–58.
- 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.
- Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular ecology* 21: 1548–1566.
- Ledig FT. 1986. Conservation strategies for forest gene resources. *Forest Ecology and Management* 14: 77–90.
- Lefèvre F, Boivin T, Bontemps A, Courbet F, Davi H, Durand-Gillmann M, Fady B, Gauzere J, Gidoïn C, Karam M-J, *et al.* 2014. Considering evolutionary processes in adaptive forestry. *Annals of Forest Science* 71: 723–739.
- Lefèvre F, Koskela J, Hubert J, Kraigher H, Longauer R, Olrik DC, Schüler S, Bozzano M, Alizoti P, Bakys R, *et al.* 2013. Dynamic conservation of forest genetic resources in 33 European countries. *Conservation Biology* 27: 373–384.
- Leimu R, Fischer M. 2008. A meta-analysis of local adaptation in plants. *PloS one* 3: e4010.
- Leimu R, Mutikainen P, Koricheva J, Fischer M. 2006. How general are positive relationships between plant population size, fitness and genetic variation? *Journal of Ecology* 94: 942–952.

- Leinonen T, McCairns RJS, O'Hara RB, Merilä J. 2013. Q_{ST} - F_{ST} comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics* 14: 179–190.
- Leites LP, Robinson AP, Rehfeldt GE, Marshall JD, Crookston NL. 2012. Height-growth response to climatic changes differs among populations of Douglas-fir: a novel analysis of historic data. *Ecological Applications* 22: 154–165.
- Leone V, Lovreglio R. 2004. Conservation of Mediterranean pine woodlands: scenarios and legislative tools. *Plant Ecology* 171: 221–235.
- Lepoittevin C, Frigerio JM, Garnier-Géré P, Salin F, Cervera MT, Vornam B, Harvengt L, Plomion C. 2010. *In Vitro* vs *in Silico* detected SNPs for the development of a genotyping array: What can we learn from a non-model species? *PLoS ONE* 5: e11034.
- 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.
- Lesbarrères D, Primmer CR, Laurila A, Merilä J. 2005. Environmental and population dependency of genetic variability-fitness correlations in *Rana temporaria*. *Molecular Ecology* 14: 311–323.
- Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, *et al.* 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature genetics* 45: 43–50.
- Li Y, Jones L, McQueen-Mason S. 2003. Expansins and cell growth. *Current Opinion Plant Biology* 6: 603–610.
- Li YF, Costello JC, Holloway AK, Hahn MW. 2008. 'Reverse ecology' and the power of population genomics. *Evolution* 62: 2984–2994.
- Li Z, Gopal V, Li X, Davis JM, Casella G. 2012. Simultaneous SNP identification in association studies with missing data. *Annals of Applied Statistics* 6: 432–456.
- Liaw A, Wiener M. 2002. Classification and regression by Random Forest. *R news* 2: 18–22.
- Lieutenant-Gosselin M, Bernatchez L. 2006. Local heterozygosity-fitness correlations with global positive effects on fitness in threespine stickleback. *Evolution* 60: 1658–1668.
- Lindner M, Maroschek M, Netherer S, Kremer A, Barbati A, Garcia-Gonzalo J, Seidl R, Delzon S, Corona P, Kolström M, *et al.* 2010. Climate change impacts adaptive capacity and vulnerability of European forest ecosystems. *Forest Ecology and Management* 259: 698–709.
- Lipka AE, Kandianis CB, Hudson ME, Yu J, Drnevich J, Bradbury PJ, Gore MA. 2015. From association to prediction: statistical methods for the dissection and selection of complex traits in plants. *Current Opinion in Plant Biology* 24: 110–118.
- Lippert C, Xiang J, Horta D, Widmer C, Kadie C, Heckerman D, Listgarten J. 2014. Greater power and computational efficiency for kernel-based association testing of sets of genetic variants. *Bioinformatics* 30: 3206–3214.
- Loehle CF. 1988. Tree life history strategies: the role of defenses. *Canadian Journal of Forest Research* 18: 209–222.

- 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.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4: 981–994.
- Luikart G, Pilgrim K, Vistry J, Ezenwa VO, Schwartz MK. 2008. Candidate gene microsatellite variation is associated with parasitism in wild bighorn sheep. *Biology Letters* 4: 228–231.
- Luquet E, David P, Lena JP, Joly P, Konecny L, Dufresnes C, Perrin N, Plenet S. 2011. Heterozygosity-fitness correlations among wild populations of European tree frog (*Hyla arborea*) detect fixation load. *Molecular Ecology* 20: 1877–1887.
- Lynch M, Walsh B. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc., Sunderland, USA.
- Mace GM, Purvis A. 2008. Evolutionary biology and practical conservation: bridging a widening gap. *Molecular Ecology* 17: 9–19.
- Mackay TF, Langley CH. 1990. Molecular and phenotypic variation in the *achaete-scute* region of *Drosophila melanogaster*. *Nature* 348: 64–66.
- Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10: 565–577.
- Majada J, Martínez-Alonso C, Feito I, Kidelman A, Aranda I, Alía R. 2011. Mini-cuttings: an effective technique for the propagation of *Pinus pinaster* Ait. *New Forests* 41: 399–412.
- Makowsky R, Pajewski NM, Klimentidis YC, Vazquez AI, Duarte CW, Allison DB, de los Campos G. 2011. Beyond missing heritability: prediction of complex traits. *PLoS Genetics* 7: e1002051.
- Marguerit E, Bouffier L, Chancerel E, Costa P, Lagane F, -Marc Guehl J, Plomion C, Brendel O. 2014. The genetics of water-use efficiency and its relation to growth in maritime pine. *Journal of Experimental Botany* 65: 4757–4768.
- Marjoram P, Zubair A, Nuzhdin S V. 2014. Post-GWAS: where next? More samples, more SNPs or more biology? *Heredity* 112: 79–88.
- Matesanz S, Gianoli E, Valladares F. 2010. Global change and the evolution of phenotypic plasticity in plants. *Annals of the New York Academy of Sciences* 1206: 35–55.
- Matesanz S, Valladares F. 2014. Ecological and evolutionary responses of Mediterranean plants to global change. *Environmental and Experimental Botany* 103: 53–67.
- McKay JK, Latta RG. 2002. Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution* 17: 285–291.
- Merilä J, Sheldon BC. 1999. Genetic architecture of fitness and non-fitness traits: Empirical patterns and development of ideas. *Heredity* 83: 103–109.
- Metzger MJ, Bunce RGH, Jongman RHG, Mùcher CA, Watkins JW. 2005. A climatic stratification of the environment of Europe. *Global Ecology and Biogeography* 14: 549–563.

- Metzger MJ, Bunce RGH, Jongman RHG, Sayre R, Trabucco A, Zomer R. 2013. A high-resolution bioclimate map of the world: A unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography* 22: 630–638.
- Milad M, Schaich H, Bürgi M, Konold W. 2011. Climate change and nature conservation in Central European forests: A review of consequences, concepts and challenges. *Forest Ecology and Management* 261: 829–843.
- Miller JM, Coltman DW. 2014. Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: a meta-analysis. *Molecular Ecology* 23: 1899–1909.
- Miller JM, Malenfant RM, David P, Davis CS, Poissant J, Hogg JT, Festa-Bianchet M, Coltman DW. 2014. Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity* 112: 240–247.
- Mittell EA, Nakagawa S, Hadfield JD. 2015. Are molecular markers useful predictors of adaptive potential? *Ecology Letters* 18: 772–778.
- Mosseler A, Major JE, Rajora OP. 2003. Old-growth red spruce forests as reservoirs of genetic diversity and reproductive fitness. *Theoretical and Applied Genetics* 106: 931–937.
- Mueller JC, Hermisson J, Olano-Marin J, Hansson B, Kempenaers B. 2011. Linking genetic mechanisms of heterozygosity-fitness correlations to footprints of selection at single loci. *Evolutionary Ecology* 25: 1–11.
- Muller B, Bourdais G, Reidy B, Bencivenni C, Massonneau A, Condamine P, Rolland G, Conéjéro G, Rogowsky P, Tardieu F. 2007. Association of specific expansins with growth in maize leaves is maintained under environmental, genetic, and developmental sources of variation. *Plant Physiology* 143: 278–290.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.
- Nakagawa S, Schielzeth H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society* 85: 935–956.
- Namkoong G. 1997. A gene conservation plan for loblolly pine. *Canadian Journal of Forest Research* 27: 433–437.
- Narum SR, Hess JE. 2011. Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* 11: 184–194.
- Naydenov KD, Alexandrov A, Matevski V, Vasilevski K, Naydenov MK, Gyuleva V, Carcaillet C, Wahid N, Kamary S. 2014. Range-wide genetic structure of maritime pine predates the last glacial maximum: Evidence from nuclear DNA. *Heredity* 151: 1–13.
- 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, Kremer A. 2011. Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12: 111–122.

- Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. *Trends in Plant Science* 9: 325–330.
- Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
- Nordborg M, Tavaré S. 2002. Linkage disequilibrium: What history has to tell us. *Trends in Genetics* 18: 83–90.
- Obeso JR. 2002. The costs of reproduction in plants. *New Phytologist* 155: 321–348.
- Olano-Marin J, Mueller JC, Kempenaers B. 2011. Heterozygosity and survival in blue tits (*Cyanistes caeruleus*): contrasting effects of presumably functional and neutral loci. *Molecular Ecology* 20: 4028–4041.
- Ouborg NJ, Vergeer P, Mix C. 2006. The rough edges of the conservation genetics paradigm for plants. *Journal of Ecology* 94: 1233–1248.
- Pâques LE. 2013. Forest tree breeding in Europe. Current state-of-the-art and perspectives. In: Managing forest ecosystems. Volume 25. Springer, Dordrecht, Netherlands.
- Parchman TL, Gompert 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.
- Paux E, Carocha V, Marques C, De Sousa AM, Borralho N, Sivadon P, Grima-Pettenati J. 2005. Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New Phytologist* 167: 89–100.
- Petit RJ, El Mousadik A, Pons O. 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12: 844–855.
- Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics* 37: 187–214.
- Petit RJ, Hu FS, Dick CW. 2008. Forests of the past: a window to future changes. *Science* 320: 1450–2.
- Pinheiro JC, Bates DM. 2000. Mixed Effects Models in S and S-Plus. Springer-Verlag, New York, USA.
- Pinosio S, González-Martínez SC, Bagnoli F, Cattonaro F, Grivet D, Marroni F, Lorenzo Z, Pausas JG, Verdú M, Vendramin GG. 2014. First insights into the transcriptome and development of new genomic tools of a widespread circum-Mediterranean tree species, *Pinus halepensis* Mill. *Molecular Ecology Resources* 14: 846–856.
- Plomion C, Bartholomé J, Lesur I, Boury C, Rodríguez-Quilón I, Lagrœulet H, Ehrenmann F, Bouffier L, Gion JM, Grivet D, *et al.* 2016. High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*). *Molecular Ecology Resources* 16: 574–587.
- Plomion C, Chancerel E, Endelman J, Lamy J-B, Mandrou E, Lesur I, Ehrenmann F, Isik F, Bink MC, van Heerwaarden J, *et al.* 2014. Genome-wide distribution of genetic diversity and linkage disequilibrium in a mass-selected population of maritime pine. *BMC Genomics* 15: 171.

- Pritchard JK, Di Rienzo A. 2010. Adaptation - not by sweeps alone. *Nature reviews Genetics* 11: 665–667.
- Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology* 20: R208–R215.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–59.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959. .
- Prunier J, Pelgas B, Gagnon F, Despons M, Isabel N, Beaulieu J, Bousquet J. 2013. The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. *BMC Genomics* 14: 368.
- Pudovkin AI, Zhdanova OL, Hedgecock D. 2009. Sampling properties of the heterozygote-excess estimator of the effective number of breeders. *Conservation Genetics* 11: 759–771.
- 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. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org>
- Rader RB, Belk MC, Shiozawa DK, Crandall KA. 2005. Empirical tests for ecological exchangeability. *Animal Conservation* 8: 239–247.
- Reed DH, Frankham R. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55: 1095–1103.
- Reed DH, Frankham R. 2003. Correlation between fitness and genetic diversity. *Conservation Biology* 17: 230–237.
- Revelle W. 2011. *psych*: Procedures for personality and psychological research. Northwestern University, Evanston (R package, version 1.1.12).
- Richardson DS, Komdeur J, Burke T. 2004. Inbreeding in the Seychelles warbler: environment-dependent maternal effects. *Evolution* 58: 2037–2048.
- Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014. Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology and Evolution* 29: 165–176.
- Rockman MV. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66: 1–17.
- Rodríguez-Quilón I, Santos-del-Blanco L, Grivet D, Jaramillo-Correa JP, Majada J, Vendramin GG, Alía R, González-Martínez SC. 2015. Local effects drive heterozygosity–fitness correlations in an outcrossing long-lived tree. *Proceedings of the Royal Society B: Biological Sciences* 282: 20152230.
- Roff DA, Fairbairn DJ. 2007. The evolution of trade-offs: Where are we? *Journal of Evolutionary Biology* 20: 433–447.

- Rogers LA, Dubos C, Surman C, Willment J, Cullis IF, Mansfield SD, Campbell MM. 2005. Comparison of lignin deposition in three ectopic lignification mutants. *New Phytologist* 168: 123–140.
- Rueda-López M, Crespillo R, Cánovas FM, Ávila C. 2008. Differential regulation of two glutamine synthetase genes by a single *Dof* transcription factor. *Plant Journal* 56: 73–85.
- Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A. 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. *The Plant Cell* 19: 2370–2390.
- Ryan MG. 2010. Temperature and tree growth. *Tree Physiology* 30: 667–668.
- Ryder, OA. 1986. Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology & Evolution* 1: 9–10.
- Sagnard F, Barberot C, Fady B. 2002. Structure of genetic diversity in *Abies alba* Mill. from southwestern Alps: Multivariate analysis of adaptive and non-adaptive traits for conservation in France. *Forest Ecology and Management* 157: 175–189.
- Sampedro J, Sieiro C, Revilla G, Gonzalez-Villa T, Zarra I. 2001. Cloning and expression pattern of a gene encoding an α -xylosidase active against xyloglucan oligosaccharides from *Arabidopsis*. *Plant Physiology* 126: 910–920.
- Sánchez M, Gianzo C, Sampedro J, Revilla G, Zarra I. 2003. Changes in α -xylosidase during intact and auxin-induced growth of pine hypocotyls. *Plant and Cell Physiology* 44: 132–138.
- Santos-del-Blanco L, Alía R, González-Martínez SC, Sampedro L, Lario F, Climent J. 2015. Correlated genetic effects on reproduction define a domestication syndrome in a forest tree. *Evolutionary Applications* 8: 403–410.
- Santos-del-Blanco L, Bonser SP, Valladares F, Chambel MR, Climent J. 2013. Plasticity in reproduction and growth among 52 range-wide populations of a Mediterranean conifer: adaptive responses to environmental stress. *Journal of Evolutionary Biology* 26: 1912–1924.
- 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.
- Santos-del-Blanco L. 2014. Evolutionary ecology of reproduction in two Mediterranean pine species: *Pinus pinaster* Ait. and *Pinus halepensis* Mill. PhD Thesis, Universidad de Valladolid, Spain.
- Sardell RJ, Arcese P, Keller LF, Reid JM. 2011. Sex-specific differential survival of extra-pair and within-pair offspring in song sparrows, *Melospiza melodia*. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 278: 3251–3259.
- Savolainen O, Hedrick P. 1995. Heterozygosity and fitness: no association in Scots pine. *Genetics* 140: 755–766.
- Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature reviews Genetics* 14: 807–820.

- Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* 38: 595–619.
- Scheiner S, Lyman R. 1989. The genetics of phenotypic plasticity I. Heritability. *Journal of Evolutionary Biology* 107: 95–107.
- Schimleck LR. 2007. NIR: A rapid, non-destructive method for measuring wood properties and its application to tree breeding. *New Zealand Journal of Forestry Science* 38: 1–35.
- Scotti I, González-Martínez SC, Budde KB, Lalagüe H. 2015. Fifty years of genetic studies: what to make of the large amounts of variation found within populations? *Annals of Forest Science* 73: 69–75.
- 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: 825–830.
- Sellis D, Callahan BJ, Petrov DA, Messer PW. 2011. Heterozygote advantage as a natural consequence of adaptation in diploids. *Proceedings of the National Academy of Sciences U.S.A.* 108: 20666–20671.
- Serra-Varela MJ, Grivet D, Vincenot L, Broennimann O, Gonzalo-Jiménez J, Zimmermann NE. 2015. Does phylogeographical structure relate to climatic niche divergence? A test using maritime pine (*Pinus pinaster* Ait.). *Global Ecology and Biogeography* 24: 1302–1313.
- Sgrò CM, Lowe AJ, Hoffmann AA. 2011. Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications* 4: 326–337.
- Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G, Dalén L, De Meester L, Ekblom R, *et al.* 2015. Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* 30: 78–87.
- Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE, McEwan JC. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* 93: 255–265.
- Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. 2013. Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics and Genomes* 9: 901–911.
- Soto A, Robledo-Arnuncio JJ, González-Martínez SC, Smouse PE, Alía R. 2010. Climatic niche and neutral genetic diversity of the six Iberian pine species: A retrospective and prospective view. *Molecular Ecology* 19: 1396–1409.
- Spitze K. 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* 135: 367–374.
- Stearns S. 1989. Trade-offs in life-history evolution. *Functional Ecology* 3: 259–268.
- Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100: 158–170.
- Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences U.S.A.* 100: 9440–9445.

- Suo C, Touloupoulou T, Bramon E, Walshe M, Picchioni M, Murray R, Ott J. 2013. Analysis of multiple phenotypes in genome-wide genetic mapping studies. *BMC Bioinformatics* 14: 151.
- Suzuki R, Shimodaira H. 2006. *Pvclust*: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22: 1540–1542.
- Szulkin M, Bierne N, David P. 2010. Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* 64: 1202–1217.
- Szulkin M, David P. 2011. Negative heterozygosity-fitness correlations observed with microsatellites located in functional areas of the genome. *Molecular Ecology* 20: 3949–3952.
- 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.
- Temple SJ, Vance CP, Gantt JS. 1998. Glutamate synthase and nitrogen. *Trends in Plant Science* 1385: 51–56.
- Tester M, Langridge P. 2010. Breeding technologies to increase crop production in a changing world. *Science* 327: 818–822.
- Thomas CD, Thomas CD, Cameron A, Cameron A, Green RE, Green RE, Bakkenes M, Bakkenes M, Beaumont LJ, Beaumont LJ, *et al.* 2004. Extinction risk from climate change. *Nature* 427: 145–148.
- Thuiller W. 2007. Climate change and the ecologist. *Nature* 448: 550–552.
- Thumma BR, Nolan MF, Evans R, Moran GF. 2005. Polymorphisms in *cinnamoyl* CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171: 1257–1265.
- Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology and Evolution* 29: 673–680.
- Tognetti R, Michelozzi M, Giovannelli A. 1997. Geographical variation in water relations, hydraulic architecture and terpene composition of Aleppo pine seedlings from Italian provenances. *Tree Physiology* 17: 241–250.
- Tuberosa R. 2012. Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology* 3: 1–26.
- Turchin MC, Chiang CWK, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics* 44: 1015–1019.
- Valladares F, Matesanz S, Guilhaumon F, Araújo MB, Balaguer L, Benito-Garzón M, Cornwell W, Gianoli E, van Kleunen M, Naya DE, *et al.* 2014. The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters* 17: 1351–1364.

- Valladares F, Sanchez-Gomez D, Zavala MA. 2006. Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology* 94: 1103–1116.
- Vasemagi A. 2006. The adaptive hypothesis of clinal variation revisited: Single-locus clines as a result of spatially restricted gene flow. *Genetics* 173: 2411–2414.
- Vendramin GG, Anzidei M, Madaghiele A, Bucci G. 1998. Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *Theoretical and Applied Genetics* 97: 456–463.
- Voegeli B, Saladin V, Wegmann M, Richner H. 2013. Heterozygosity is linked to the costs of immunity in nestling great tits (*Parus major*). *Ecology and Evolution* 3: 4815–4827.
- Voltas J, Chambel MR, Prada MA, Ferreiro JP. 2008. Climate-related variability in carbon and oxygen stable isotopes among populations of Aleppo pine grown in common-garden tests. *Trees* 22: 759–769.
- Voltas J, Lucabaugh D, Chambel MR, Ferrio JP. 2015. Intraspecific variation in the use of water sources by the circum-Mediterranean conifer *Pinus halepensis*. *New Phytologist* 208: 1031–1041.
- Vries S, Alan M, M Bozzano, Burianek V, Collin E, Cottrell J, Ivankovic M, Keeling CI, Koskela J, Rotach P, *et al.* 2014. Pan-European strategy for genetic conservation of forest trees. Bioversity International, Rome, Italy.
- Waller DM, Dole J, Bersch AJ. 2008. Effects of stress and phenotypic variation on inbreeding depression in *Brassica rapa*. *Evolution* 62: 917–931.
- Wang TL, O'Neill GA, Aitken SN. 2010. Integrating environmental and genetic effects to predict responses of tree populations to climate. *Ecological Applications* 20: 153–163.
- Warren C, McGrath J, Adams M. 2001. Water availability and carbon isotope discrimination in conifers. *Oecologia* 127: 476–486.
- Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai C-J, Neale DB. 2010. Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. *New Phytologist* 188: 515–532.
- Weir B, Cockerham C. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.
- White TL, Adams WT, Neale DB. 2007. Forest genetics. Cabi Ed., Cambridge, UK.
- Whitlock MC, Guillaume F. 2009. Testing for spatially divergent selection: comparing Q_{ST} to F_{ST} . *Genetics* 183: 1055–1063.
- Yanagisawa S, Akiyama A, Kisaka H, Uchimiya H, Miwa T. 2004. Metabolic engineering with *Dof1* transcription factor in plants: Improved nitrogen assimilation and growth under low-nitrogen conditions. *Proceedings of the National Academy of Sciences U.S.A.* 101: 7833–7838.
- Yang J, Benyamin B, McEvoy B. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature* 42: 565–569.

- Yang Q, Wang Y. 2012. Methods for analyzing multivariate phenotypes in genetic association studies. *Journal of Probability and Statistics* 2012: 1–13.
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, 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.
- Yu Q, Li B, Nelson C, McKeand S, Batista V, Mullin T. 2006a. Association of the *cad-n1* allele with increased stem growth and wood density in full-sib families of loblolly pine. *Tree Genetics and Genomes* 2: 98–108.
- Zhong R, McCarthy RL, Haghighat M, Ye ZH. 2013. The poplar MYB master switches bind to the SMRE site and activate the secondary wall biosynthetic program during wood formation. *PLoS one* 8: 18–21.
- Zhu C, Gore M, Buckler ES, Yu J. 2008. Status and prospects of association mapping in plants. *The Plant Genome* 1: 5–20.

SUPPLEMENTARY INFORMATION

Supplementary Information Chapter 1

Table S1.1. Average assignment probability for each of the optimal K=6 clusters (C1 through C6) representing six geographical genetic groups (i.e. 'gene pools') found in *Pinus pinaster*. Assignment probabilities were assessed using the Bayesian clustering method implemented in STRUCTURE 2.2 (Pritchard *et al.* 2000), running an admixture model with correlated allele frequencies. Ten runs were performed for each number of clusters from K=2 to K=10, with a burn-in length of 100,000 and a run length of 1,000,000 iterations. The optimal number of population clusters, K, was calculated following Evanno *et al.* (Evanno *et al.* 2005). The ten independent runs were averaged using CLUMP (Jakobsson and Rosenberg 2007) to generate the Q matrix of individual posterior cluster probabilities for the optimal K=6.

Genotype	C1	C2	C3	C4	C5	C6	Gene pool
ALT1	0.001	0.000	0.001	0.027	0.971	0.000	Iberian Atlantic
ALT10	0.003	0.000	0.386	0.062	0.548	0.001	Iberian Atlantic
ALT2	0.001	0.000	0.001	0.135	0.863	0.000	Iberian Atlantic
ALT3	0.014	0.001	0.004	0.113	0.868	0.000	Iberian Atlantic
ALT4	0.002	0.001	0.008	0.178	0.811	0.000	Iberian Atlantic
ALT5	0.002	0.000	0.089	0.127	0.762	0.020	Iberian Atlantic
ALT6	0.004	0.000	0.492	0.003	0.501	0.000	Iberian Atlantic
ALT8	0.001	0.000	0.002	0.178	0.819	0.000	Iberian Atlantic
ALT9	0.002	0.000	0.000	0.001	0.997	0.000	Iberian Atlantic
ARM1	0.001	0.000	0.001	0.001	0.995	0.002	Iberian Atlantic
ARM10	0.002	0.020	0.001	0.001	0.975	0.001	Iberian Atlantic
ARM3	0.001	0.000	0.012	0.001	0.985	0.001	Iberian Atlantic
ARM4	0.001	0.000	0.001	0.005	0.993	0.000	Iberian Atlantic
ARM5	0.001	0.000	0.057	0.016	0.926	0.000	Iberian Atlantic
ARM6	0.002	0.004	0.001	0.001	0.992	0.000	Iberian Atlantic
ARM7	0.001	0.028	0.003	0.000	0.967	0.001	Iberian Atlantic
ARM9	0.033	0.002	0.097	0.021	0.841	0.006	Iberian Atlantic
ARN10	0.019	0.001	0.895	0.072	0.008	0.005	Central Spain
ARN11	0.007	0.001	0.937	0.001	0.011	0.043	Central Spain
ARN12	0.001	0.005	0.936	0.003	0.027	0.028	Central Spain

...

(The complete table with 509 genotypes can be found in

<http://rspb.royalsocietypublishing.org/content/282/1820/20152230.figures-only>)

Table S1.2. Average survival for 34 *Pinus pinaster* populations (BLUP $\pm$ SE) evaluated in four common gardens with contrasted environments (Asturias, Portugal, Madrid and Cáceres). Values are in logit scale and represent deviations from the site mean estimate, which can be found in the last row. These population mean estimates are exclusively used to standardize individual genotype estimates when merging individuals from different populations to compute HFCs.

Gene pool	Population	Survival			
		Asturias	Portugal	Madrid	Cáceres
French Atlantic coast	Hourtin	-0.03 $\pm$ 0.20	0.10 $\pm$ 0.14	0.20 $\pm$ 0.16	0.20 $\pm$ 0.21
	Leverdon	-0.02 $\pm$ 0.20	0.04 $\pm$ 0.14	0.15 $\pm$ 0.16	0.06 $\pm$ 0.21
	Mimizan	0.02 $\pm$ 0.20	-0.25 $\pm$ 0.13	-0.03 $\pm$ 0.15	-0.19 $\pm$ 0.21
	Olonne sur Mer	-0.27 $\pm$ 0.20	-0.14 $\pm$ 0.14	-0.08 $\pm$ 0.17	0.11 $\pm$ 0.21
	Petrocq	-0.18 $\pm$ 0.20	-0.02 $\pm$ 0.15	0.15 $\pm$ 0.16	0.22 $\pm$ 0.21
	Pleucadec	-0.01 $\pm$ 0.20	-0.24 $\pm$ 0.14	-0.27 $\pm$ 0.17	-0.39 $\pm$ 0.23
	St-Jean des Monts	-0.29 $\pm$ 0.19	0.15 $\pm$ 0.14	0.24 $\pm$ 0.15	0.54 $\pm$ 0.19
Iberian Atlantic coast	Alto de la Llama	0.08 $\pm$ 0.23	0.09 $\pm$ 0.19	0.07 $\pm$ 0.22	-0.21 $\pm$ 0.28
	Armayán	0.25 $\pm$ 0.22	-0.01 $\pm$ 0.16	0.24 $\pm$ 0.18	-0.02 $\pm$ 0.24
	Cadavedo	-0.09 $\pm$ 0.20	-0.21 $\pm$ 0.14	0.02 $\pm$ 0.16	-0.24 $\pm$ 0.23
	Castropol	0.00 $\pm$ 0.23	0.04 $\pm$ 0.19	-0.28 $\pm$ 0.22	-0.32 $\pm$ 0.28
	Lamuño	-0.02 $\pm$ 0.23	0.00 $\pm$ 0.19	-0.15 $\pm$ 0.22	-0.13 $\pm$ 0.28
	Leiria	0.00 $\pm$ 0.21	0.05 $\pm$ 0.15	-0.25 $\pm$ 0.17	0.29 $\pm$ 0.21
	Puerto de Vega	-0.04 $\pm$ 0.21	0.12 $\pm$ 0.15	-0.15 $\pm$ 0.17	-0.05 $\pm$ 0.22
	San Cipriano	-0.16 $\pm$ 0.20	-0.20 $\pm$ 0.14	0.07 $\pm$ 0.16	0.24 $\pm$ 0.20
	Sergude	-0.03 $\pm$ 0.21	-0.24 $\pm$ 0.15	-0.27 $\pm$ 0.18	0.01 $\pm$ 0.23
	Sierra de Barcia	0.17 $\pm$ 0.24	-0.22 $\pm$ 0.19	0.33 $\pm$ 0.22	-0.02 $\pm$ 0.28
Central Spain	Arenas de S.Pedro	0.14 $\pm$ 0.21	0.01 $\pm$ 0.14	-0.10 $\pm$ 0.16	-0.24 $\pm$ 0.22
	Bayubas de Abajo	-0.11 $\pm$ 0.21	-0.12 $\pm$ 0.16	-0.14 $\pm$ 0.18	-0.02 $\pm$ 0.23
	Boniches	-0.07 $\pm$ 0.23	0.30 $\pm$ 0.20	0.23 $\pm$ 0.22	0.28 $\pm$ 0.26
	Carbonero	0.13 $\pm$ 0.24	-0.18 $\pm$ 0.21	-0.02 $\pm$ 0.24	0.01 $\pm$ 0.29
	Cenicientos	-0.02 $\pm$ 0.23	0.23 $\pm$ 0.19	0.15 $\pm$ 0.22	0.01 $\pm$ 0.27
	Coca	0.07 $\pm$ 0.21	-0.04 $\pm$ 0.14	-0.41 $\pm$ 0.17	-0.39 $\pm$ 0.23
	Cuellar	0.12 $\pm$ 0.20	-0.19 $\pm$ 0.14	-0.43 $\pm$ 0.17	-0.22 $\pm$ 0.22
	Olba	0.29 $\pm$ 0.22	0.49 $\pm$ 0.16	0.41 $\pm$ 0.16	0.51 $\pm$ 0.20
	Quatretonda	0.08 $\pm$ 0.22	0.36 $\pm$ 0.17	0.49 $\pm$ 0.18	0.10 $\pm$ 0.23
	San Leonardo	0.02 $\pm$ 0.22	-0.50 $\pm$ 0.17	-0.06 $\pm$ 0.20	-0.07 $\pm$ 0.25
	Valdemaqueda	0.03 $\pm$ 0.23	0.10 $\pm$ 0.18	0.08 $\pm$ 0.20	0.07 $\pm$ 0.26
Corsica	Pinia	0.00 $\pm$ 0.23	0.11 $\pm$ 0.19	0.10 $\pm$ 0.21	-0.10 $\pm$ 0.27
	Pineta	-0.01 $\pm$ 0.22	0.00 $\pm$ 0.16	0.09 $\pm$ 0.18	0.34 $\pm$ 0.22
Southern Spain	Cómpeta	-0.14 $\pm$ 0.24	0.10 $\pm$ 0.22	0.17 $\pm$ 0.26	0.12 $\pm$ 0.30
	Oria	0.22 $\pm$ 0.20	0.11 $\pm$ 0.13	-0.12 $\pm$ 0.15	-0.25 $\pm$ 0.21
Morocco	Madisouka	0.02 $\pm$ 0.25	0.03 $\pm$ 0.25	0.08 $\pm$ 0.30	-0.07 $\pm$ 0.33
	Tamrabta	-0.22 $\pm$ 0.19	-0.26 $\pm$ 0.14	-0.77 $\pm$ 0.18	-0.41 $\pm$ 0.23
Species		3.47 $\pm$ 0.23	0.52 $\pm$ 0.09	-1.08 $\pm$ 0.11	-3.12 $\pm$ 0.20

Table S1.3. Site to site correlations for survival BLUPs, used as fitness proxy in a multi-site *Pinus pinaster* common garden. Below diagonal, correlations for individual (genotype) estimates. Above diagonal, correlations among sites for population mean estimates. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

	Asturias	Portugal	Madrid	Cáceres
Asturias		0.207	0.186	-0.199
Portugal	0.020		0.525***	0.419**
Madrid	-0.066	0.108		0.642***
Cáceres	-0.037	0.136	0.183	

Table S1.4. Correlations between individual (genotype) multilocus observed heterozygosity (H_o) based on 6,100 SNPs and survival estimates in the four common gardens. Correlations are made using all individuals (i.e. species-wide) or only subsets of them by population, gene pool or climatic region (Atlantic vs Mediterranean). When merging individual (genotype) data from different populations, heterozygosity and survival were standardized by the population means. Asterisks indicate significant (uncorrected for multiple testing) correlations, $P < 0.05$. No significant correlations were observed after Bonferroni correction. NA: All genotypes for Sierra de Barcia population survived in the Asturias common garden, so correlations for this population in this site could not be computed.

Gene pool	Population	Asturias	Portugal	Madrid	Cáceres
French Atlantic coast	Hourtin	-0.233	0.104	0.240	0.131
	Leverdon	0.329	0.121	0.201	0.147
	Mimizan	-0.131	0.342	0.134	0.175
	Olonne sur Mer	-0.082	-0.034	0.219	0.177
	Petrocq	0.687*	-0.113	0.011	0.042
	Pleucadec	-0.026	-0.034	0.087	0.267
	St-Jean des Monts	-0.093	0.098	-0.025	-0.282
<i>Overall</i>		<i>0.108</i>	<i>0.042</i>	<i>0.108</i>	<i>0.056</i>
Iberian Atlantic coast	Alto de la Llama	-0.170	0.568	0.459	0.667*
	Armayán	-0.536	-0.132	-0.002	0.110
	Cadavedo	-0.329	0.377	0.503	0.350
	Castropol	-0.269	0.710*	-0.314	0.163
	Lamuño	0.354	-0.065	-0.186	-0.375
	Leiria	-0.115	0.193	0.079	0.121
	Puerto de Vega	-0.136	-0.240	-0.512	-0.073
	San Cipriano	-0.156	-0.457	-0.631	0.321
	Sergude	0.180	0.239	0.185	0.084
	Sierra de Barcia	-	0.063	0.376	-0.033
<i>Overall</i>		<i>-0.092</i>	<i>0.136</i>	<i>0.044</i>	<i>0.076</i>
Atlantic region		0.049	0.071	0.088	0.061
Central Spain	Arenas de S. Pedro	-0.312	-0.107	0.019	0.430
	Bayubas de Abajo	0.289	-0.211	-0.514*	-0.252
	Boniches	-0.330	0.354	0.166	0.358
	Cenicientos	-0.041	-0.224	-0.250	-0.321
	Coca	0.507*	0.096	0.198	0.047
	Cuellar	0.188	0.019	0.028	0.276
	Olba	-0.127	-0.171	0.048	0.292
	Quatretonda	-0.139	0.242	0.430	0.387
	San Leonardo	0.394	-0.097	-0.235	0.228
	Valdemaqueda	-0.442	-0.188	0.239	-0.250
<i>Overall</i>		<i>0.071</i>	<i>0.066</i>	<i>-0.011</i>	<i>0.174*</i>
Corsica	Pinia	0.033	-0.275	-0.261	-0.022
	Pineta	-0.588	-0.060	-0.592	-0.095
<i>Overall</i>		<i>-0.262</i>	<i>-0.085</i>	<i>-0.337</i>	<i>-0.131</i>
Southern Spain	Oria	-0.175	-0.247	0.070	-0.138
Morocco	Tamrabta	0.404	-0.027	0.168	-0.137
Mediterranean region		0.047	0.027	-0.026	0.120
Species		0.047	0.050	0.038	0.087

Table S1.5. Correlations between individual (genotype) multilocus observed heterozygosity (H_o) based on 2,661 control SNPs and survival estimates in the four common gardens. Correlations are made using all individuals (i.e. species-wide) or only subsets of them by population, gene pool or climatic region (Atlantic vs Mediterranean). When merging individual (genotype) data from different populations, heterozygosity and survival were standardized by the population means. Asterisks indicate significant (uncorrected for multiple testing) correlations, $P < 0.05$. No significant correlations were observed after Bonferroni correction. NA: All genotypes for Sierra de Barcia population survived in the Asturias common garden, so correlations for this population in this site could not be computed.

Gene pool	Population	Asturias	Portugal	Madrid	Cáceres
French Atlantic coast	Hourtin	-0.031	-0.050	0.019	0.070
	Leverdon	0.316	0.106	0.215	0.155
	Mimizan	-0.186	0.268	0.101	0.169
	Olonne sur Mer	-0.062	-0.045	0.238	0.172
	Petrocq	0.627*	-0.169	-0.017	0.062
	Pleucadec	0.217	-0.058	0.275	0.230
	St-Jean des Monts	-0.119	0.108	-0.012	-0.280
<i>Overall</i>		<i>0.114</i>	<i>0.011</i>	<i>0.107</i>	<i>0.052</i>
Iberian Atlantic coast	Alto de la Llama	-0.052	0.599	0.433	0.558
	Armayán	-0.277	-0.187	-0.407	0.162
	Cadavedo	-0.345	0.295	0.567	0.280
	Castropol	-0.331	0.703*	-0.363	0.166
	Lamuño	0.267	-0.045	-0.250	-0.698*
	Leiria	-0.098	0.137	-0.200	0.040
	Puerto de Vega	-0.245	-0.199	-0.593	-0.091
	San Cipriano	0.113	-0.587	-0.348	0.336
	Sergude	0.123	0.287	0.225	0.207
	Sierra de Barcia	-	0.021	0.317	-0.120
<i>Overall</i>		<i>-0.109</i>	<i>0.118</i>	<i>0.001</i>	<i>0.065</i>
Atlantic region		0.051	0.042	0.075	0.055
Central Spain	Arenas de S. Pedro	-0.179	-0.068	0.123	0.499*
	Bayubas de Abajo	0.318	-0.245	-0.461*	-0.308
	Boniches	-0.259	0.299	0.214	0.295
	Cenicientos	-0.053	-0.294	-0.435	-0.364
	Coca	0.492*	0.170	-0.008	0.195
	Cuellar	0.197	0.029	0.036	-0.029
	Olba	-0.110	-0.101	0.025	0.299
	Quatretonda	-0.140	0.267	0.449	0.435
	San Leonardo	0.348	-0.154	-0.144	0.277
	Valdemaqueda	-0.523	-0.159	0.077	-0.314
<i>Overall</i>		<i>0.029</i>	<i>0.015</i>	<i>0.051</i>	<i>0.159*</i>
Corsica	Pinia	-0.087	-0.467	-0.265	-0.323
	Pineta	-0.523	-0.149	-0.605	-0.084
<i>Overall</i>		<i>0.430</i>	<i>-0.442*</i>	<i>0.335</i>	<i>0.396</i>
Southern Spain	Oria	0.078	-0.303	0.050	-0.112
Morocco	Tamrabta	0.300	0.041	0.125	-0.188
Mediterranean region		0.144*	-0.083	0.094	0.152*
Species		0.086	-0.009	0.082	0.093*

Figure S1.1. Correlation between multilocus observed heterozygosity (H_o) based on 6,100 SNPs and estimated inbreeding coefficients $s(g2)$ in *Pinus pinaster*; $s(g2)$ values were not computed for the three populations with less than 5 genotypes, and thus they are not shown in the plot. Symbols represent the distinct gene pools found in the species.

Supplementary Information Chapter 2

Table S2.1. Principal component (PC) loadings and variance explained for the PCA of 13 phenotypic traits measured in *P. halepensis*, using individual tree data.

Trait	PC1	PC2	PC3
Dichloromethane	0.85	-0.06	0.14
Ethanol	0.62	-0.07	-0.13
Water	0.64	0.04	-0.12
Total extract	0.96	0.05	0.03
Lignin	-0.48	-0.11	0.85
Ratio HG	-0.12	-0.12	0.93
Carbon δ^{13}	0.01	0.39	0.02
Total Carbon	0.05	0.01	0.36
Height	-0.08	0.92	0.05
Diameter	-0.01	0.95	-0.03
Stem volume	0.01	0.93	0.02
Female flowering	-0.02	-0.08	-0.14
Male flowering	-0.11	0.03	-0.10
% Variance Explained	0.23	0.22	0.12
Cumulative Proportion	0.23	0.45	0.58

Table S2.2. Principal component (PC) loadings and variance explained for the PCA of 13 phenotypic traits using population mean data of 28 *P. halepensis* populations.

Trait	PC1	PC2	PC3
Dichloromethane	0.71	-0.28	0.06
Ethanol	0.53	-0.47	-0.09
Water	0.66	0.14	0.39
Total extract	0.84	-0.03	0.18
Lignin	-0.92	0.04	0.2
Ratio.HG	-0.55	-0.21	0.32
Carbon δ^{13}	0.07	0.13	-0.77
Total Carbon	-0.05	0.57	0.38
Height	-0.02	0.91	-0.3
Diameter	0.04	0.95	-0.19
Stem volume	0.01	0.95	-0.15
Female flowering	-0.07	-0.15	0.90
Male flowering	0.3	-0.47	0.42
% Variance Explained	0.32	0.24	0.14
Cumulative Proportion	0.32	0.56	0.70

Table S2.3. Multi-locus results for the association test between 294 SNPs and 13 traits in *P. halepensis*: proportion of variance explained (PVE), conditional prior probability of a single-nucleotide polymorphism (SNP) being in the model (p_{SNP}), number of SNPs in the model (N_{SNP}), and average effect of a SNP on the phenotype (σ_{AVE}).

Trait	PVE	p_{SNP}	N_{SNP}	σ_{AVE}
Dichloromethane	0.041	0.032	7	0.131
Ethanol	0.080	0.025	9	0.179
Water	0.017	0.033	9	0.086
Total extract	0.108	0.083	18	0.220
Lignin	0.059	0.059	11	0.144
Ratio.HG	0.043	0.019	4	0.354
Carbon δ^{13}	0.058	0.069	15	0.116
Total Carbon	0.078	0.077	22	0.089
Height	0.097	0.116	22	0.159
Diameter	0.046	0.030	9	0.096
Stem volume	0.060	0.030	6	0.412
Female flowering	0.043	0.036	5	0.215
Male flowering	0.065	0.013	7	0.194

Table S2.4. Parameter estimates and 95% credible intervals (equal-tail probability intervals) for multi-locus association parameters in Table S2.3.

Trait	PVE low	PVE up	p_{SNP} low	p_{SNP} up	N_{SNP} low	N_{SNP} up	σ_{AVE} low	σ_{AVE} up
Dichloromethane	0.041	0.041	0.029	0.036	6	8	0.119	0.143
Ethanol	0.080	0.080	0.023	0.026	9	9	0.163	0.195
Water	0.017	0.017	0.031	0.034	9	9	0.070	0.102
Total extract	0.108	0.108	0.083	0.083	18	18	0.201	0.239
Lignin	0.059	0.059	0.057	0.062	10	12	0.131	0.157
Ratio.HG	0.043	0.043	0.014	0.024	3	5	0.340	0.368
Carbon δ^{13}	0.058	0.058	0.069	0.070	15	15	0.100	0.132
Total Carbon	0.078	0.078	0.073	0.082	21	23	0.077	0.101
Height	0.097	0.097	0.114	0.118	21	23	0.143	0.175
Diameter	0.046	0.046	0.029	0.031	9	9	0.083	0.110
Stem volume	0.060	0.060	0.028	0.032	5	7	0.357	0.467
Female flowering	0.043	0.043	0.032	0.040	4	6	0.202	0.228
Male flowering	0.065	0.066	0.007	0.019	5	9	0.183	0.205

Table S2.5. Multi-locus association for 294 SNPs and 13 traits in *P. halepensis*: Estimates of the posterior inclusion probability (PIP) and the magnitude of effect on phenotype (β) for the three single-nucleotide polymorphisms identified with the highest posterior inclusion probability (PIP) for each trait. The SNPs that are present in multiple association models are given in color.

Trait	SNP 1	PIP 1	β 1	SNP 2	PIP 2	β 2	SNP 3	PIP 3	β 3
Dichloromethane	snp326	0.362	-0.019	snp264	0.335	0.019	snp246	0.298	0.010
Ethanol	snp38	0.189	-0.032	snp264	0.185	0.030	snp97	0.120	0.010
Water	snp309	0.111	0.012	snp273	0.095	0.015	snp30	0.053	0.004
Total extract	snp264	0.565	0.171	snp22	0.121	-0.044	snp8	0.066	-0.009
Lignin	snp227	0.237	0.023	snp29	0.134	0.016	snp22	0.102	0.008
Ratio.HG	snp227	0.123	0.018	snp273	0.109	-0.023	snp225	0.053	-0.008
Carbon δ^{13}	snp227	0.390	0.017	snp126	0.368	-0.012	snp2	0.322	0.007
Total Carbon	snp242	0.063	0.009	snp273	0.038	0.006	snp349	0.038	0.004
Height	snp259	0.337	0.014	snp128	0.284	0.009	snp98	0.281	0.008
Diameter	snp267	0.204	0.019	snp227	0.154	-0.013	snp140	0.109	-0.010
Stem volume	snp227	0.223	-0.031	snp267	0.095	0.009	snp142	0.073	0.007
Female flowering	snp308	0.358	-0.310	snp227	0.232	-0.027	snp136	0.193	-0.030
Male flowering	snp163	0.175	0.009	snp259	0.165	-0.006	snp43	0.162	0.005

Figure S2.1. Boxplots showing variation in wood quality traits within each *P. halepensis* population. The red dotted line corresponds to the trait mean estimate across populations.

Figure S2.2. Boxplots showing variation in growth traits within each *P. halepensis* population. The red dotted line corresponds to trait mean estimate across populations.

Figure S2.3. Boxplots showing variation in female flowering (number of cones) and male flowering (1-4 scale) within each *P. halepensis* population. The red dotted line corresponds to trait mean estimate across populations.

Figure S2.4. Graphical pairwise correlations, correlation coefficients and significance of the correlation for 13 phenotypic traits at the population level (28 provenance BLUPs) for *P. halepensis*. Red boxes pool together traits of the same group (from top left to bottom right: wood quality, drought resistance, growth and reproductive traits). Below diagonal, graphical pairwise relationships. Above diagonal, pairwise Pearson's correlation coefficients and level of significance of the correlation (** $P < 0.001$, * $P < 0.01$, $P < 0.05$, $P < 0.1$).

Figure S2.5. Pairwise relationship among the 9 geographic and environmental variables in the 28 *Pinus halepensis* source populations. Below diagonal, graphical pairwise relationships. Above diagonal, pairwise Pearson's correlation coefficients and level of significance of the correlation (** $P < 0.001$, * $P < 0.01$, $P < 0.05$, $P < 0.1$).

Figure S2.6. Scree plot for the PCA of the nine geographic and environmental variables.

Figure S2.7. Boxplots (genotypic effects) of SNPs showing significant associations with wood quality and growth traits using a Bayesian Variable Selection Regression model as implemented in piMASS (Guan & Stephens, 2011). In 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. Trait values are transformed (linear regression on structure Q's + qqnorm transformed).

Supplementary Information Chapter 3

Table S3.1. Maritime pine tested populations, gene pool membership and number of clones (N) per population. * Indicates the population not included in the Bordeaux common garden.

Population	Gene pool	N
Hourtin	French Atlantic	26
Leverdon	French Atlantic	27
Mimizan	French Atlantic	18
Olonne sur Mer	French Atlantic	24
Petrocq	French Atlantic	24
Pleucadec	French Atlantic	20
St-Jean des Monts	French Atlantic	28
Alto de la Llama	Iberian Atlantic	9
Armayán	Iberian Atlantic	8
Cadavedo	Iberian Atlantic	10
Castropol	Iberian Atlantic	10
Lamuño	Iberian Atlantic	9
Leiria	Iberian Atlantic	23
Puerto de Vega	Iberian Atlantic	8
San Cipriano	Iberian Atlantic	9
Sergude	Iberian Atlantic	21
Sierra de Barcia	Iberian Atlantic	8
Arenas de San Pedro*	Central Spain	17
Bayubas de Abajo	Central Spain	18
Boniches	Central Spain	9
Carbonero el Mayor	Central Spain	6
Cenicientos	Central Spain	9
Coca	Central Spain	18
Cuellar	Central Spain	28
Olba	Central Spain	22
Quatretonda	Central Spain	17
San Leonardo	Central Spain	14
Valdemaqueda	Central Spain	12
Pinia	Corsica	16
Pineta	Corsica	9
Oria	Southern Spain	26
Tamrabta	Morocco	19
TOTAL		522

Table S3.2. Variance partition among hierarchical levels (metapopulation, population and clone) for height and survival measured on *Pinus pinaster* trees grown at five contrasting common gardens. Variance components and 95% credible intervals were obtained from Bayesian (G)MLM with block as fixed factor and each hierarchical level as random factors.

Trait	Common garden	Year	Metap	lo.ci.Metap	up.ci.Metap	Pop	lo.ci.Pop	up.ci.Pop	Clone	lo.ci.Clone	up.ci.Clone
Height	Asturias	2011	4.88	1.31	48.99	4.41	2.21	9.77	10.57	8.48	12.96
		2012	24.48	5.33	195.96	25.79	12.40	45.32	38.73	30.71	48.22
		2014	86.45	13.59	517.01	62.90	24.56	121.52	103.05	73.75	134.76
	Portugal	2011	306.54	0.10	2011.19	489.72	190.57	900.81	835.20	643.60	1038.74
		2012	1262.72	172.38	8104.25	666.27	358.03	1881.72	1801.09	1108.04	2580.74
		2013	2363.79	0.00	15138.86	1511.70	699.90	3705.31	3146.32	1964.96	4229.99
	Bordeaux	2013	60.88	12.60	402.15	23.10	7.86	45.69	50.56	40.90	66.65
	Madrid	2011	390.00	59.17	1947.44	268.84	94.13	604.88	779.48	581.75	1151.58
	Cáceres	2011	3.81	0.13	34.09	1.69	0.25	5.39	6.09	3.19	10.47
Survival	Asturias	2011	0.033	0.024	0.040	0.002	0.001	0.002	0.117	0.085	0.144
		2012	0.030	0.023	0.040	0.002	0.001	0.002	0.104	0.085	0.144
		2014	2.40E-05	1.79E-05	3.15E-05	0.034	0.026	0.045	0.673	0.524	0.877
	Portugal	2011	1.20E-09	9.19E-10	1.57E-09	0.088	0.067	0.113	0.046	0.035	0.062
		2012	0.005	0.004	0.006	0.115	0.090	0.150	0.136	0.101	0.177
		2013	5.62E-08	4.16E-08	7.40E-08	0.128	0.098	0.163	0.104	0.080	0.135
	Bordeaux	2012	6.27E-09	4.58E-09	7.63E-09	0.000	0.000	0.000	1.230	0.920	1.523
		2013	7.41E-05	5.51E-05	9.93E-05	0.001	0.001	0.001	0.986	0.826	1.361
		2014	0.036	0.028	0.048	0.038	0.029	0.049	0.492	0.366	0.632
	Madrid	2011	3.84E-08	3.20E-08	5.37E-08	0.105	0.077	0.133	0.082	0.067	0.116
	Cáceres	2011	1.58E-08	1.14E-08	1.96E-08	0.089	0.070	0.121	0.140	0.104	0.176

Table S3.3. Q_X statistics for height and survival measured on *Pinus pinaster* trees grown at five contrasting common gardens and partition into its two components (F_{ST} -like component and LD-like component).

Trait	Location	Year	Q_X	Q_X p -val	F_{ST} - comp	F_{ST} -comp p -val	LD-comp	LD-comp p -val
Height	Asturias	2011	78.43	0	31.16	0.5082	47.27	0
		2012	83.08	0	31.73	0.4182	51.35	0
		2014	74.14	0	32.72	0.2984	41.42	0
	Bordeaux	2013	68.43	2E-04	32.25	0.3433	36.19	1.00E-04
	Portugal	2011	87.32	0	31.56	0.4467	55.76	0
		2012	76.61	0	31.59	0.4228	45.01	0
		2013	77.03	0	31.87	0.3862	45.16	0
	Madrid	2011	94.10	0	30.90	0.5642	63.20	0
	Cáceres	2011	122.62	0	39.65	0.0194	82.98	0
Survival	Asturias	2011	33.55	0.507	25.91	0.978	7.64	2.59E-01
		2012	48.54	0.0448	27.64	0.8406	20.90	0.0102
		2014	39.30	0.324	26.65	0.9246	12.65	0.1573
	Bordeaux	2012	58.26	0.0027	31.00	0.5164	27.27	0.0016
		2013	62.99	0.0016	30.19	0.7081	32.80	0.0007
		2014	58.75	0.007	30.98	0.6093	27.77	0.0047
	Portugal	2011	57.51	0.0115	27.42	0.9954	30.09	0.003
		2012	38.30	0.5201	28.38	0.9403	9.92	0.4046
		2013	44.31	0.4758	26.66	0.9859	17.65	0.312
	Madrid	2011	41.94	0.0905	25.94	0.9995	16.00	0.0254
	Cáceres	2011	46.22	0.0411	26.38	0.9974	19.84	0.0106

Figure S3.1. Barplot of individual assignment probability to each of the optimal K=6 clusters representing six neutral gene pools in *Pinus pinaster*, as produced by STRUCTURE 2.3.4 (Pritchard *et al.* 2000) using 6,100 SNPs. We performed ten runs for each number of clusters, K=1 to K=8, with a burn-in length of 10000 and a run length of 100000 iterations, using an admixture model with correlated allele frequencies. Then, we used Structure Harvester (Earl and vonHoldt 2011) to select the optimal number of clusters and found K=6 to be the optimal number of clusters for the studied individuals. 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.

Figure S3.2. Number of SNPs retained at each step in the procedure to detect polygenic adaptation in Maritime pine, departing from a set of 6100 SNP markers.

Supplementary Information Chapter 4

Table S4.1. Sample size, allelic richness and Nei's heterozygosity (H_e) estimated both with 12 nuclear microsatellite and 266 SNP markers for each of the 35 maritime pine tested populations, structured in six different gene pools.

Population	Gene Pool	N° Clones	Allelic richness		Nei's H_e	
			nuSSRs	SNPs	nuSSRs	SNPs
Hourtin	French Atlantic	26	2.43	1.52	0.55	0.25
Leverdon	French Atlantic	26	2.53	1.53	0.57	0.26
Mimizan	French Atlantic	18	2.49	1.53	0.58	0.26
Olonne sur Mer	French Atlantic	23	2.52	1.52	0.57	0.25
Petrocq	French Atlantic	22	2.39	1.52	0.55	0.26
Pleucadec	French Atlantic	20	2.56	1.54	0.61	0.27
St-Jean des Monts	French Atlantic	26	2.58	1.53	0.60	0.26
Alto de la Llama	Iberian Atlantic	9	2.43	1.50	0.66	0.24
Armayán	Iberian Atlantic	8	2.23	1.45	0.56	0.23
Cadavedo	Iberian Atlantic	10	2.22	1.48	0.55	0.24
Castropol	Iberian Atlantic	10	2.31	1.50	0.54	0.24
Lamuño	Iberian Atlantic	9	2.20	1.47	0.50	0.23
Leiria	Iberian Atlantic	23	2.62	1.53	0.62	0.26
Puerto de Vega	Iberian Atlantic	8	2.39	1.48	0.54	0.23
Rodoiros	Iberian Atlantic	7	2.49	1.57	0.57	0.28
San Cipriano	Iberian Atlantic	8	2.52	1.51	0.59	0.24
Sergude	Iberian Atlantic	21	2.43	1.51	0.56	0.25
Sierra de Barcia	Iberian Atlantic	8	2.43	1.51	0.56	0.26
Arenas de San Pedro	Central Spain	17	2.58	1.57	0.58	0.29
Bayubas de Abajo	Central Spain	19	2.53	1.58	0.58	0.28
Boniches	Central Spain	9	2.53	1.63	0.57	0.30
Carbonero el Mayor	Central Spain	5	2.37	1.55	0.65	0.27
Cenicientos	Central Spain	9	2.47	1.55	0.56	0.27
Coca	Central Spain	17	2.56	1.58	0.60	0.29
Cuellar	Central Spain	28	2.72	1.57	0.63	0.28
Olba	Central Spain	21	2.53	1.59	0.58	0.29
Quatretonda	Central Spain	17	2.75	1.58	0.64	0.29
San Leonardo	Central Spain	14	2.56	1.56	0.58	0.28
Valdemaqueda	Central Spain	12	2.47	1.56	0.55	0.28
Pinia	Corsica	9	2.26	1.52	0.52	0.25
Pineta	Corsica	12	2.26	1.50	0.52	0.25
Cómpeta	Southern Spain	4	3.19	1.60	0.76	0.30
Oria	Southern Spain	26	2.69	1.58	0.64	0.29
Madisouka	Morocco	1	-	-	-	-
Tamrabta	Morocco	15	2.36	1.39	0.55	0.19

Figure S4.1. Barplot of individual assignment probability to each of the optimal $K=6$ clusters representing six neutral gene pools in *Pinus pinaster*, as produced by STRUCTURE 2.3.4 software using 266 SNPs. 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. Data for the unique Madisouka genotype is displayed in the first line segment of the Morocco neutral gene pool.

Figure S4.2. Nei's genetic diversity for 266 SNPs and inbreeding estimates (indicated by crosses) in *Pinus pinaster* populations. Conservation group membership as stated in the key. Data not shown for Carbonero, Cmpeta and Madisouka populations, as the reduced number of clones would have led to biased estimates.

Table S4.2. Survival (%) and height (cm) corrected means and 95% credible intervals, overall and for each of the 35 maritime pine tested populations, evaluated in four common gardens under Atlantic (Asturias, Portugal) and Mediterranean climates (Madrid, Cáceres).

Population	Gene Pool	Height				Survival			
		Asturias	Portugal	Madrid	Cáceres	Asturias	Portugal	Madrid	Cáceres
Hourtin	French Atlantic	31.8 (30.3-33.8)	27.5 (26.0-29.0)	22.8 (21.1-24.7)	19.4 (18.0-22.2)	98.2 (97.2-98.7)	68.2 (61.2-75.3)	24.5 (19.5-33.0)	3.7 (2.6-6.2)
Leverdón	French Atlantic	31.3 (29.8-33.4)	25.7 (24.4-27.5)	23.5 (22.0-25.5)	20.8 (17.9-22.2)	98.1 (97.1-98.7)	67.5 (59.4-73.2)	25.6 (18.7-31.2)	3.2 (2.1-5.2)
Mimizan	French Atlantic	26.9 (24.9-28.1)	20.9 (19.6-22.5)	18.0 (16.4-20.0)	16.0 (14.0-18.5)	98.1 (97.4-98.7)	60.1 (51.6-65.6)	19.3 (15.5-25.8)	2.8 (1.6-4.0)
Olonne sur Mer	French Atlantic	31.1 (29.2-32.8)	27.5 (25.9-29.2)	22.7 (20.9-24.7)	20.8 (18.6-22.9)	98.1 (96.1-98.3)	61.8 (54.9-69.4)	19.3 (15.1-27.5)	4.2 (2.3-5.3)
Petrocq	French Atlantic	30.9 (29.2-32.8)	24.8 (23.3-26.5)	23.6 (21.7-25.4)	19.6 (17.2-21.6)	98.1 (96.5-98.4)	66.4 (57.9-72.4)	24.8 (18.0-31.7)	3.8 (2.5-5.9)
Pleucadeuc	French Atlantic	26.4 (24.7-28.1)	22.3 (20.6-23.6)	18.9 (17.4-21.4)	18.2 (15.9-21.5)	98.1 (97.1-98.6)	60.1 (49.7-65.8)	18.5 (11.6-22.2)	2.1 (1.2-3.4)
St-Jean des Monts	French Atlantic	31.3 (29.4-32.8)	26.9 (25.9-28.9)	24.4 (22.8-26.3)	22.9 (20.5-24.3)	98.1 (96.0-98.3)	69.5 (63.7-76.5)	25.6 (20.7-34.2)	6.0 (3.6-8.4)
Alto de la Llama	Iberian Atlantic	28.9 (27.0-31.4)	25.7 (23.9-28.1)	18.5 (15.8-20.7)	17.0 (12.9-20.6)	98.1 (97.1-98.8)	69.9 (57.1-76.8)	22.8 (14.7-32.7)	2.2 (1.4-4.4)
Armayán	Iberian Atlantic	32.1 (30.2-34.2)	24.4 (22.7-26.3)	17.4 (15.2-19.1)	15.8 (13.0-18.2)	98.2 (97.7-99.1)	66.3 (57.2-74.2)	27.0 (19.9-36.4)	3.2 (1.8-5.0)
Cadavedo	Iberian Atlantic	28.3 (26.2-29.7)	22.5 (20.8-24.2)	16.3 (14.6-18.3)	14.8 (11.3-16.7)	98.1 (96.9-98.6)	60.3 (52.1-67.5)	21.7 (16.2-28.8)	2.7 (1.5-3.7)
Castropol	Iberian Atlantic	28.7 (26.9-31.4)	23.0 (21.4-25.5)	16.8 (14.0-19.7)	16.2 (11.7-19.9)	98.2 (96.9-98.7)	67.2 (57.2-75.6)	18.3 (10.4-23.6)	2.3 (1.1-3.8)
Lamuño	Iberian Atlantic	29.6 (27.1-31.6)	22.1 (19.8-24.1)	18.5 (15.5-21.1)	16.8 (13.1-20.4)	98.2 (97.0-98.8)	66.4 (54.9-74.2)	18.3 (11.8-26.6)	2.9 (1.4-4.6)
Leiria	Iberian Atlantic	30.1 (28.2-31.7)	22.9 (21.0-24.3)	18.4 (16.6-20.7)	16.6 (14.3-18.3)	98.1 (97.2-98.7)	66.1 (59.9-74.2)	16.7 (11.7-22.9)	4.6 (2.8-6.8)
Puerto de Vega	Iberian Atlantic	30.3 (28.3-31.8)	23.4 (22.0-25.1)	15.8 (13.9-17.9)	13.4 (10.9-15.8)	98.1 (96.9-98.6)	71.5 (61.6-75.8)	19.0 (13.3-24.4)	2.8 (1.8-4.6)
Rodoiros	Iberian Atlantic	24.9 (21.9-27.0)	18.9 (16.6-21.0)	20.0 (16.9-22.4)	15.7 (12.6-20.3)	98.2 (97.2-98.9)	73.5 (64.3-83.4)	21.9 (14.5-31.9)	3.1 (1.3-4.7)
San Cipriano	Iberian Atlantic	30.5 (28.5-32.0)	23.6 (22.0-25.1)	17.3 (16.0-19.5)	17.1 (15.1-19.2)	98.1 (96.6-98.4)	58.9 (52.3-67.4)	25.5 (17.4-30.5)	4.1 (2.7-6.1)
Sergude	Iberian Atlantic	27.4 (25.6-29.3)	20.9 (19.2-22.4)	16.1 (13.7-18.1)	14.6 (12.5-17.5)	98.2 (97.0-98.7)	61.1 (50.8-66.7)	16.2 (11.7-22.9)	3.3 (1.8-4.8)
Sierra de Barcia	Iberian Atlantic	27.8 (25.5-30.2)	20.6 (18.2-23.0)	18.6 (16.3-21.1)	15.8 (12.9-19.7)	98.2 (97.4-99.1)	57.0 (47.6-69.3)	26.4 (19.6-39.8)	3.0 (1.7-5.4)
Arenas de San Pedro	Central Spain	24.7 (22.7-26.1)	20.3 (18.8-22.0)	17.7 (16.0-19.7)	15.6 (12.9-18.2)	98.2 (97.6-98.9)	65.2 (59.1-73.5)	19.7 (14.4-25.8)	2.4 (1.4-3.8)
Bayubas de Abajo	Central Spain	22.1 (20.7-24.4)	18.9 (17.7-21.2)	18.7 (16.3-20.4)	14.2 (11.7-16.7)	98.1 (96.7-98.5)	62.7 (54.4-70.3)	20.4 (13.0-25.5)	2.8 (1.9-5.0)
Boniches	Central Spain	26.7 (24.6-29.3)	19.7 (18.0-22.0)	21.6 (19.0-23.8)	16.8 (14.2-19.6)	98.1 (96.7-98.7)	74.4 (64.4-82.5)	24.6 (17.8-36.5)	4.5 (2.3-7.1)
Carbonero el Mayor	Central Spain	23.9 (20.1-25.4)	20.1 (17.9-23.2)	18.2 (14.6-20.5)	15.2 (10.6-18.4)	98.2 (97.3-99.1)	61.6 (48.8-71.7)	20.7 (12.7-30.8)	3.5 (1.7-5.9)
Cenicentos	Central Spain	27.0 (24.3-29.0)	23.8 (21.6-25.8)	18.2 (15.2-20.0)	14.9 (11.5-17.9)	98.1 (96.8-98.7)	70.9 (61.7-79.9)	24.1 (16.5-34.6)	3.6 (1.8-5.8)
Coca	Central Spain	20.7 (18.9-22.5)	19.0 (17.6-20.7)	17.5 (15.4-19.8)	15.3 (12.9-18.6)	98.1 (97.5-98.9)	66.0 (57.0-72.0)	14.8 (9.77-19.0)	2.2 (1.1-3.2)
Cuellar	Central Spain	21.3 (19.8-23.3)	19.6 (17.9-21.0)	17.1 (15.5-19.5)	15.6 (13.5-18.6)	98.2 (97.5-98.9)	59.4 (53.4-68.0)	14.3 (9.39-18.3)	2.7 (1.4-3.6)
Olba	Central Spain	27.8 (26.0-29.4)	24.3 (22.5-25.6)	24.4 (22.7-26.2)	21.0 (18.9-22.8)	98.3 (97.7-99.1)	79.2 (70.2-83.4)	29.7 (23.2-39.2)	5.6 (3.3-8.2)
Quatretonda	Central Spain	27.2 (24.9-28.7)	23.1 (21.2-24.5)	20.3 (18.7-22.3)	17.4 (14.8-20.0)	98.1 (97.4-98.9)	73.7 (66.9-81.5)	33.1 (24.8-42.1)	3.4 (2.1-5.7)
San Leonardo	Central Spain	22.9 (20.8-24.7)	18.0 (16.0-20.1)	18.6 (16.1-20.6)	14.3 (11.7-17.8)	98.3 (97.1-98.8)	51.6 (40.6-60.7)	20.6 (14.1-27.6)	2.9 (1.6-4.8)
Valdemaqueda	Central Spain	24.5 (22.8-26.9)	21.1 (18.8-22.5)	17.6 (15.0-19.4)	16.4 (14.2-19.4)	98.2 (97.3-98.9)	69.9 (58.0-76.1)	24.2 (16.0-32.3)	3.4 (1.9-5.7)
Pinia	Eastern Medit.	30.1 (28.0-31.7)	22.7 (20.9-24.3)	18.8 (16.7-21.0)	17.5 (15.2-19.5)	98.1 (96.9-98.6)	68.2 (56.9-73.5)	24.4 (16.6-31.9)	4.0 (2.8-7.1)
Pineta	Eastern Medit.	23.4 (21.2-25.7)	20.4 (18.3-22.3)	17.0 (14.8-19.4)	14.4 (11.5-18.2)	98.1 (97.1-98.8)	69.3 (58.7-77.2)	23.6 (15.8-32.8)	2.6 (1.4-4.8)
Cómpeta	Southern Spain	32.0 (28.8-35.0)	25.3 (21.8-27.3)	24.3 (20.8-27.0)	17.2 (13.3-20.4)	98.1 (96.2-98.7)	67.4 (55.0-77.8)	25.7 (15.6-37.2)	3.7 (1.9-6.8)
Oria	Southern Spain	22.2 (20.7-24.2)	18.8 (17.2-20.0)	17.8 (15.9-19.4)	17.0 (14.2-18.6)	98.2 (97.8-99.0)	69.9 (62.3-75.0)	18.8 (14.4-24.6)	2.5 (1.5-3.7)
Madisouka	Morocco	28.6 (24.6-34.1)	27.1 (22.1-30.8)	24.9 (20.5-30.1)	NA	98.2 (96.7-98.9)	66.9 (53.0-78.1)	22.0 (12.8-36.6)	3.1 (1.3-6.0)
Tamrabta	Morocco	21.1 (19.4-22.9)	17.5 (15.7-18.9)	19.4 (17.0-21.7)	13.9 (10.8-16.3)	98.1 (96.3-98.3)	60.3 (51.4-66.5)	10.1 (6.49-13.6)	2.2 (1.2-3.3)
SPECIES		27.7 (25.9-28.8)	22.6 (21.2-23.7)	19.4 (18.4-21.1)	16.8 (14.8-18.8)	98.1 (97.0-98.7)	66.1 (60.3-70.0)	21.4 (17.7-25.6)	3.1 (1.7-3.9)

Table S4.3. Site to site correlations for survival (below diagonal) and height (above diagonal) clone BLUPs for maritime pine trees grown in four contrasting environments. In diagonal, correlation among traits in the same environment. Significance levels: * $P < 0.01$, ** $P < 0.001$, *** $P < 10^{-10}$

		HEIGHT			
SITE		Asturias	Portugal	Madrid	Cáceres
SURVIVAL	Asturias	0.001	0.485***	0.303***	0.248**
	Portugal	0.011	0.135**	0.381***	0.244**
	Madrid	-0.046	0.102*	0.086	0.237**
	Cáceres	-0.008	0.139**	0.155**	0.055

Table S4.4. Quantitative genetic parameter estimates for neutral gene pools of maritime pine. 95% credible intervals are given in brackets. Estimates derived from five year-old trees measured in a common garden replicated at four contrasting sites, with variance components obtained from Bayesian (G)MLM.

Trait	Site	Neutral gene pool						Species	
		French Atlantic	Iberian Atlantic	Central Spain	Southern Spain	Eastern Med.	Morocco		
H^2	Height	Asturias	0.129 (0.089-0.187)	0.098 (0.069-0.152)	0.187 (0.152-0.258)	0.243 (0.162-0.348)	0.291 (0.195-0.456)	0.179 (0.113-0.317)	0.154 (0.124-0.180)
		Portugal	0.099 (0.042-0.155)	0.122 (0.083-0.184)	0.176 (0.125-0.259)	0.122 (0.069-0.212)	0.399 (0.269-0.549)	0.291 (0.163-0.447)	0.148 (0.117-0.181)
		Madrid	0.156 (0.089-0.265)	0.205 (0.076-0.324)	0.387 (0.224-0.483)	0.135 (0.056-0.262)	0.336 (0.151-0.497)	NE(a)	0.253 (0.172-0.309)
		Caceres	0.171 (0.088-0.338)	0.150 (0.077-0.353)	0.556 (0.321-0.695)	NE	NE(a)	NE	0.282 (0.136-0.407)
	Survival	Asturias	NE (b)	NE(a)	NE(a)	NE(a)	NE (c)	NE (c)	0.045 (0.035-0.057)
		Portugal	0.022 (0.017-0.030)	0.013 (0.010-0.019)	NE(a)	0.038 (0.031-0.054)	0.059 (0.049-0.082)	NE (c)	0.015 (0.011-0.020)
		Madrid	0.038 (0.030-0.049)	0.014 (0.011-0.019)	0.037 (0.029-0.051)	NE(a)	NE (c)	NE (b)	0.017 (0.015-0.025)
		Caceres	0.026 (0.020-0.034)	0.028 (0.022-0.037)	0.069 (0.053-0.088)	NE(a)	NE(a)	NE (c)	0.022 (0.017-0.029)
Q_{ST}^*	Height	Asturias	0.218 (0.052-0.557)	0.097 (0.037-0.193)	0.177 (0.057-0.471)	0.431 (0.264-0.724)	0.403 (0.190-0.637)	0.855 (0.633-0.953)	0.449 (0.312-0.578)
		Portugal	0.495 (0.154-0.745)	0.141 (0.075-0.294)	0.099 (0.025-0.379)	0.420 (0.215-0.676)	NE (b)	0.874 (0.679-0.958)	0.313 (0.229-0.500)
		Madrid	0.265 (0.117-0.467)	NE (b)	0.206 (0.050-0.450)	0.723 (0.427-0.893)	NE (b)	NE (a)	0.361 (0.234-0.552)
		Caceres	0.237 (0.083-0.451)	0.225 (0.094-0.488)	0.127 (0.060-0.285)	NE	NE (a)	NE	0.326 (0.170-0.581)
	Survival	Asturias	NE (b)	NE (b)	NE(a)	NE (b)	NE (c)	NE (c)	0.106 (0.075-0.157)
		Portugal	0.072 (0.053-0.111)	0.192 (0.133-0.258)	NE(a)	0.012 (0.007-0.016)	NE (b)	NE (c)	0.357 (0.273-0.450)
		Madrid	0.036 (0.023-0.048)	0.214 (0.145-0.270)	0.244 (0.172-0.318)	NE(a)	NE (c)	NE (b)	0.388 (0.303-0.472)
		Caceres	0.347 (0.247-0.419)	0.150 (0.103-0.201)	0.164 (0.121-0.232)	NE(a)	NE(a)	NE (c)	0.347 (0.279-0.457)

NE (=Not Estimable) cause:

NE (without letter)=model does not converge, not possible to estimable variance components;

NE(a)= Negligible clone within population variance;

NE(b)= Negligible population variance;

NE(c)= Negligible genetic variance (both population and clone within population).

Table S4.5. Among population molecular marker (F_{ST}) and quantitative trait (Q_{ST}^*) differentiation estimates for the five conservation-relevant population groups with more than two populations proposed for maritime pine. NE: not estimable due to negligible population variance. In brackets standard errors (F_{ST}) or 95% credible intervals (Q_{ST}^*).

	Trait	Site	Conservation-relevant population groups				
			French Atlantic 1	French Atlantic 2	Iberian Atlantic	Central Spain 1	Central Spain 2
F_{ST}	Height	-	0.011 (0.010)	0.006(0.004)	0.023 (0.09)	0.019 (0.004)	0.024 (0.011)
	Survival	-	0.009 (0.003)	0.006 (0.001)	0.025 (0.002)	0.007 (0.002)	0.030 (0.004)
Q_{ST}^*	Height	Asturias	0.000 (0.000-0.000)	0.000 (0.000-0.000)	0.097 (0.037-0.193)	0.011 (0.004-0.026)	0.045 (0.017-0.140)
		Portugal	0.014 (0.005-0.058)	0.018 (0.005-0.058)	0.141 (0.075-0.294)	0.011 (0.004-0.033)	0.012 (0.004-0.084)
		Madrid	0.000 (0.000-0.000)	0.005 (0.002-0.018)	NE	0.000 (0.000-0.000)	0.325 (0.132-0.500)
		Caceres	NE	NE	0.225 (0.094-0.488)	NE	0.118 (0.053-0.307)
	Survival	Asturias	0.000 (0.000-0.000)	0.000 (0.000-0.000)	NE	0.012 (0.009-0.019)	NE
		Portugal	0.000 (0.000-0.000)	0.000 (0.000-0.000)	0.192 (0.133-0.258)	NE	0.082 (0.075-0.086)
		Madrid	0.000(0.000-0.000)	0.000 (0.000-0.000)	0.214 (0.145-0.270)	0.000 (0.000-0.000)	0.085 (0.057-0.122)
		Caceres	0.000 (0.000-0.000)	0.114 (0.078-0.157)	0.150 (0.103-0.201)	0.000 (0.000-0.000)	0.069 (0.048-0.094)

Figure S4.3. Ranking of variable importance for maritime pine population clustering based on quantitative trait data (height or survival at four contrasting sites) assessed by Random Forest.

Table S4.6. Correlations between height BLUPs estimated from maritime pine trees grown in a common garden replicated at four contrasting sites (Asturias, Portugal, Madrid and Cáceres), and nine geographic and environmental variables representing average conditions in source populations. Significant correlations (<0.05) are indicated in bold.

Variable	Common garden site			
	Asturias	Portugal	Madrid	Cáceres
Altitude	-0.61	-0.50	-0.07	-0.44
Latitude	0.35	0.39	0.02	0.43
Longitude	-0.04	0.03	0.25	0.28
Mean temperature of the coldest month	0.44	0.27	-0.18	0.12
Mean temperature of the warmest month	-0.44	-0.37	-0.01	-0.28
Annual temperature mean	0.23	0.06	-0.16	-0.06
Continental index*	-0.60	-0.46	-0.04	-0.26
Annual rainfall	0.55	0.32	-0.13	-0.08
Rainfall warmest quarter**	0.39	0.32	0.08	0.27

*Difference between mean temperature of the warmest month and mean temperature of the coldest month.

**Precipitation in June, July and August.

Table S4.7. Results from Principal Component Analysis applied to environmental and geographical data from 35 maritime pine source populations.

Variable	PC1	PC2	PC3
Altitude	0.416	0.000	-0.128
Latitude	-0.354	0.165	0.446
Longitude	0.000	-0.465	0.600
Mean temperature of the coldest month	-0.355	-0.406	-0.177
Mean temperature of the warmest month	0.414	-0.118	0.000
Annual temperature mean	-0.127	-0.620	-0.312
Continental index*	0.428	0.127	0.000
Rainfall warmest quarter**	-0.349	0.351	0.144
Annual Rainfall	-0.287	0.221	-0.518
Importance of components			
Proportion of variance explained	0.54	0.22	0.14
Cumulative proportion of variance	0.54	0.76	0.90
Standard deviation	2.21	1.40	1.11

*Difference between mean temperature of the warmest month and mean temperature of the coldest month.

**Precipitation in June, July and August.

Table S4.8. Correlations between height BLUPs estimated from *Pinus pinaster* trees grown in a common garden replicated at four contrasting sites (Asturias, Portugal, Madrid and Cáceres) and the first three Principal Components derived from a PCA analysis for nine geographic and environmental variables. Significant correlations (<0.05) are indicated in bold.

Common garden site	PC1	PC2	PC3
Asturias	0.63	-0.17	-0.23
Portugal	0.49	-0.13	-0.02
Madrid	0.02	-0.11	0.21
Cáceres	0.31	-0.11	0.39

Figure S4.4. Location of the 42 current maritime pine conservation units (triangles), climatic stratification of Europe (as defined in Lefèvre *et al.* 2013, modified from the original map in Metzger *et al.* 2005) and distribution of maritime pine (*Pinus pinaster*) neutral gene pools.

Figure S4.5. *Pinus pinaster* regions of provenance in Spain as defined in Alía *et al.* (1996) and location of the five Spanish conservation-relevant population groups.

ANNEX 1

Plomion C, Bartholomé J, Lesur I, Boury C, **Rodríguez-Quilón I**, Lagraulet H, Ehrenmann F, Bouffier L, Gion JM, Grivet D, *et al.* 2016. High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*). *Molecular Ecology Resources* 16: 574-587.

High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*)

C. PLOMION,*† J. BARTHOLOMÉ,*† I. LESUR,*‡ C. BOURY,*† I. RODRÍGUEZ-QUILÓN,§
H. LAGRAULET,*† F. EHRENMANN,*† L. BOUFFIER,*† J. M. GION,*¶ D. GRIVET,§ M. DE MIGUEL,*†
N. DE MARÍA,§ M. T. CERVERA,§ F. BAGNOLI,** F. ISIK,†† G. G. VENDRAMIN** and
S. C. GONZÁLEZ-MARTÍNEZ§

*BIOGECO, UMR 1202, INRA, F-33610 Cestas, France, †BIOGECO, UMR 1202, University of Bordeaux, F-33400 Talence, France, ‡HelixVenture, F-33700 Mérignac, France, §Forest Research Centre, INIA, E-28040 Madrid, Spain, ¶UMR AGAP, CIRAD, F-33612 Cestas, France, **Institute of Biosciences and Bioresources, National Research Council, Sesto Fiorentino (FI), Italy, ††Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA

Abstract

Maritime pine provides essential ecosystem services in the south-western Mediterranean basin, where it covers around 4 million ha. Its scattered distribution over a range of environmental conditions makes it an ideal forest tree species for studies of local adaptation and evolutionary responses to climatic change. Highly multiplexed single nucleotide polymorphism (SNP) genotyping arrays are increasingly used to study genetic variation in living organisms and for practical applications in plant and animal breeding and genetic resource conservation. We developed a 9k Illumina Infinium SNP array and genotyped maritime pine trees from (i) a three-generation inbred (F2) pedigree, (ii) the French breeding population and (iii) natural populations from Portugal and the French Atlantic coast. A large proportion of the exploitable SNPs (2052/8410, i.e. 24.4%) segregated in the mapping population and could be mapped, providing the densest ever gene-based linkage map for this species. Based on 5016 SNPs, natural and breeding populations from the French gene pool exhibited similar level of genetic diversity. Population genetics and structure analyses based on 3981 SNP markers common to the Portuguese and French gene pools revealed high levels of differentiation, leading to the identification of a set of highly differentiated SNPs that could be used for seed provenance certification. Finally, we discuss how the validated SNPs could facilitate the identification of ecologically and economically relevant genes in this species, improving our understanding of the demography and selective forces shaping its natural genetic diversity, and providing support for new breeding strategies.

Keywords: linkage mapping, maritime pine, population genetics, single nucleotide polymorphism

Received 26 February 2015; revision received 28 August 2015; accepted 3 September 2015

Introduction

Maritime pine (*Pinus pinaster* Aiton, Pinaceae) is a long-lived wind-pollinated forest tree species native to the western part of the Mediterranean Basin. Its natural range extends from northern Morocco in the south to French Brittany in the north, and from Portugal in the west to Italy in the east (<http://www.euforgen.org/distribution-maps/>). It is found in various ecological situations, from sea level to an altitude of 2100 m in the High Atlas (Morocco), from regions characterized by heavy annual rainfall in an Atlantic climate to dry regions in the semi-arid Mediterranean climate, and

from calcareous to acidic soils (Alía & Martín 2003). Its scattered distribution has resulted in local adaptations and high levels of genetic differentiation for adaptive traits across its natural range (González-Martínez *et al.* 2002; Lamy *et al.* 2011, 2014; Santos-del-Blanco *et al.* 2012). Considerable genetic differentiation between ecotypes has been reported for various neutral molecular markers, providing clear evidence for a geographic structure of genetic diversity in this species (Bucci *et al.* 2007; Jaramillo-Correa *et al.* 2010, 2015; Santos-del-Blanco *et al.* 2012).

This fast-growing tree, which generally reaches its economically optimum size around 40 years old, has been widely planted in systematic reforestation programmes since the 19th century, to secure coastal (along the Atlantic) and inland (Castilian plateau) sand dune areas, to drain marshes and to create new forests for

Correspondence: Christophe Plomion, Fax: +33557122881; E-mail: plomion@pierroton.inra.fr

resin production. Over the last 50 years, maritime pine has been commercially exploited as a timber resource for the forestry industry (sawmills, wood panels, pulp and paper). In recent years, it has also been used as a source of chemicals for the bio-industry (Jorge *et al.* 2002; Rohdewald 2002; Touriño *et al.* 2005), as bioactive phenolic compounds can be extracted from its bark. There are now 4.2 million ha under maritime pine within its natural range and 200 000 ha outside it (mostly in Australia; Bouffier *et al.* 2013). The breeding of maritime pine began in the 1960s in south-western France, after several species and provenance trials had shown that the local ecotype was the best adapted and fastest growing tree in the Aquitaine soil and climatic conditions (Illy 1966; Harfourche 1995). This programme has now reached its third generation and is one of the most advanced conifer breeding programmes in the world (Mullin *et al.* 2011).

Maritime pine has also been adopted by the forest tree genetics research community as a key model species to study genetic variation or linked mutations underlying phenotypic variability, particularly those selected by the environment and involved in local adaptation (reviewed by Gonzalez-Martinez *et al.* 2011). It is now hoped that the discovery of polymorphisms causing changes in gene expression and/or amino acid sequences will lead to innovations in genetic resource management, for both breeding (Isik 2014) and conservation strategies (Ouborg *et al.* 2010). Such discoveries should also lead to changes in silviculture practices to take into account the evolutionary processes inferred from neutral and selected markers (Lefèvre *et al.* 2014). Major efforts have been devoted to the sequencing and assembly of the maritime pine transcriptome (Canales *et al.* 2014), for studies of the molecular basis of the phenotypic response to biotic and abiotic constraints (Le Provost *et al.* 2013). Moreover, since the pioneering work of Lepoittevin *et al.* (2010), describing the design of the first multiplex single nucleotide polymorphism (SNP) genotyping assay in maritime pine, medium-scale SNP arrays have been developed (Tables S1 and S2, Supporting information). These assays have made it possible to characterize hundreds of trees, at hundreds of loci, for various applications: nucleotide diversity analysis (Plomion *et al.* 2014), QTL detection (de Miguel *et al.* 2014), association mapping (Lepoittevin *et al.* 2012; Budde *et al.* 2014), environmental association (Jaramillo-Correa *et al.* 2015) and linkage map construction (Chancerel *et al.* 2013).

Given the high throughput and reliability of the Infinium platform from Illumina (e.g. Pavy *et al.* 2013, for spruce; Bartholomé *et al.* 2015, for eucalyptus), we used this platform to design a customized genotyping array for maritime pine, including 4712 SNPs from the studies cited above and 4237 SNPs newly identified from RNA-seq data and new amplicon resequencing. The resulting

9k SNP array is the largest genotyping chip ever produced for this species with 7252 workable SNPs in at least one of the four tested populations. We assessed the suitability of this array for linkage mapping, identification of seed sources (Portugal vs. French Atlantic coast) and comparison between natural and breeding populations (within the French Atlantic gene pool). Finally, we discuss the utility of this SNP array for exploring genetic diversity and its contribution to phenotypic variation, genetic inferences about historical demographic events, the past action of natural selection and adaptive evolution, and the implementation of novel tree breeding strategies.

Materials and methods

Design of an Illumina Infinium array for maritime pine

We designed a 9k Illumina Infinium SNP array (8949 SNPs) for maritime pine, including the two subsets of SNPs described below.

Previously available SNPs. Over the past 5 years, several studies have reported the discovery of SNP markers in maritime pine (Tables S1 and S2, Supporting information). We first selected all available SNPs from assays based on VeraCode, GoldenGate or Infinium Illumina technologies (Illumina, San Diego, CA, USA). This number initially amounted to 4997 SNPs. As we knew that some assays had been used on a number of occasions, SNP redundancy was checked using BLASTN by aligning the flanking sequences (ranging from 87 to 419 bp in length) against themselves. We found that 4442 SNPs were used in a single study (94.3%), 255 SNPs (5.4%) in two studies and 15 SNPs (0.3%) in three studies (Fig. S1, Supporting information). Within each of the 270 groups of redundant SNPs, we retained the SNP with the longest flanking sequence. In the case of groups containing strictly identical SNPs (same polymorphic site and same flanking sequences) but with different dbSNP accessions, we arbitrarily selected one of the dbSNP ID. Thus, 285 SNPs were discarded and the final list comprised 4712 SNPs. Then, flanking sequences of each retained SNP were aligned against the most up-to-date maritime pine UniGene (PineV3, Canales *et al.* 2014: http://www.scbi.uma.es/sustainpinedb/home_page) to identify its corresponding contig and position within that contig. Alignments were performed by carrying out BLASTN searches in the BLASTALL version 2.2.26 suite ($E\text{-value} = 10^{-5}$). For only two SNPs, no hit was obtained. A single contig was found for 3494 SNPs (74.2%), whereas several contigs equally aligned (same $E\text{-value}$, alignment length and %identity) with the flanking sequences of 1218 SNPs (25.8%). Applying a more

stringent *E*-value (10^{-10} , 10^{-30}) did not result into a better rate of single hit. Such a high level of multiple assignment was expected considering the relatively high level of redundancy of this resource that comprises 210 513 contigs, while a recent estimate suggests that the pine genome contains about 25 000 genes (J. Wegrzyn & D. Neale, personal communication). For these SNPs, we selected the first PineV3 contig of the BLASTN output, but provided the whole list in Table S3 (Supporting information).

Newly discovered SNPs. This second subset comprised SNPs obtained by the random screening of EST data or specifically detected in candidate gene sequences.

Newly discovered SNPs from 454 sequence reads: A flow-chart describing the steps involved in the identification of SNPs from 454 sequence data is shown in Fig. 1. Three genotypes involved as progenitors of interprovenance hybrids in the framework of the maritime pine breeding programme [accessions 110-4019-1 from Corsica (C), 0284-2 from Landes (L) and 112-4-1 (M) from Morocco] were used. A composite cDNA library was constructed

with the SMART PCR cDNA synthesis kit (Clontech, Laboratories Inc., Mountain View, CA, USA) for each tree. The C and M libraries contained equal proportions of cDNAs from differentiating xylem, swelling buds and young needles, whereas the L library consisted of equal proportions of cDNAs from differentiating xylem, swelling and quiescent buds, and young and mature needles. Pyrosequencing (454 Titanium; Roche, Branford, CT, USA) was performed with the Roche-454 Genome Sequencing platform (FLX Titanium Technology). Sequences (available under accession numbers SRX031589, SRX208012 and SRX031592 from the NCBI short-read archive) were cleaned with the PYROCLEANER tool (Mariette *et al.* 2011), which removes particularly short (<150 bp) and long reads (>600 bp), reads with a percentage of *ns* (ambiguous base calls) >2%, low-complexity regions, and duplicated reads. For each library, the SNPs were identified by aligning each set of sequences against the 198 425 contigs of the second maritime pine UniGene (PineV2) established by Chancerel *et al.* (2013). Alignments were performed with the CLC Genomics Workbench Reference Mapping function of

Fig. 1 Strategy for developing SNPs segregating in three full-sib progenies (C × M, L × C, L × M), from high-throughput sequencing of the parental lines (C, L and M).

Fig. 2 Strategy for developing new SNPs in candidate genes.

CLC GENOMICS WORKBENCH version 6.0 (CLC Bio, Aarhus, Denmark), with the default parameters. SNPs were then detected with the neighbourhood quality standard algorithm (NQS) and the following parameters: minimum coverage = 6, minimum central base quality = 20, minimum neighbourhood quality over a window length of 11 nucleotides = 15, maximum gap and mismatch count = 2 and minimum allele frequency = 20%.

Newly discovered SNPs from candidate genes: We obtained 105 *in vitro* SNPs from two full-length candidate genes with functions relating to phenology (*col1* and *gia*) and 64 amplicons sequenced as part of the Comparative Re-sequencing in European Conifers (CRIEC) project, an EVOLTREE (<http://www.evoltree.eu>) initiative (D. Grivet, unpublished data). An additional set of *in silico* SNPs was then obtained from 281 candidate genes with functions relating to abiotic and biotic (plant defence) stress responses in forest trees selected from the following sources: (i) 66 genes from published (Perdiguerio *et al.* 2013) and unpublished maritime pine sequences available from GenBank, (ii) 149 genes from sequencing studies in other conifers (Wachowiak *et al.* 2009; Kujala & Savolainen 2012), (iii) 53 transcripts displaying differential expression in the presence and absence of pine wood nematode infection (Santos *et al.* 2012), (iv) 10 genes associated with adventitious shoot induction and plant development in pines (Alonso *et al.* 2007; Ordás *et al.* unpublished for *knox* genes) and (v) 3 genes potentially involved in cavitation resistance in beech (Lalagüe *et al.* 2014). These sequences were blasted (BLASTN, $E\text{-value} = 10^{-5}$) against PineV3 UniGene to retrieve the best matching contig and the SNP flanking sequences

from the catalogue of 55 607 available SNPs obtained by Canales *et al.* (2014; Fig. 2). Redundant SNPs were removed.

As described above, the corresponding PineV3 contigs were retrieved for each newly discovered SNP. A single contig was found for 2446 SNPs (58.3%), whereas several contigs equally aligned (same $E\text{-value}$, alignment length and %identity) with the flanking sequences of the remaining SNPs. We selected the first contig of the BLASTN output for the functional annotation, but the whole list is available in Table S3 (Supporting information).

Annotation of synonymous and nonsynonymous substitutions

The respective positions of SNPs were defined in the contigs of the maritime pine UniGene from Canales *et al.* (2014). Within the coding sequences (when characterized), nonsynonymous and synonymous SNPs were annotated by comparing the amino acids translated from the reference codon to the codon containing the SNP. The functional annotation was retrieved from the study by Canales *et al.* (2014). All these items of information are available in Table S3 (Supporting information).

Populations studied and genetic analysis

The SNP assays were tested and validated on the basis of Mendelian segregation in a mapping pedigree (92 genotypes) and genetic diversity analysis in an elite breeding population from France (50 genotypes) and natural pop-

ulations of different origins (French Atlantic coast with 50 genotypes and Portugal with 42 genotypes).

Mapping population and linkage analysis. The mapping population consisted of a three-generation inbred pedigree (F2) obtained by the self-pollination of an interprovenance 'Landes × Corsica' hybrid (accession H12 resulting from the control cross between L146 and C10 genotypes). In total, 638 F2 seeds were planted in a nursery in June 1998 and 626 seedlings were transplanted into the field in March 1999 (4 × 2 m, 0.51 ha; Lacanau de Mios, France). After 15 years, 565 F2 plants were still available for genetic analysis. We used 92 F2 plants (a different set compared to that used by Chancerel *et al.* 2013) to test for the Mendelian segregation of SNP markers and to associate them with a particular linkage map position. The F2 plants with the most recombinant genotypes were selected with MAPPOP software (<http://www.bio.unc.edu/faculty/vision/lab/mappop/>, Vision *et al.* 2000), and a linkage map was established by genotyping 477 F2 plants for 248 SNPs (unpublished) distributed over 12 linkage groups (LG), the haploid chromosome number in pines.

The R package *onemap* version 2.0-3 (Margarido *et al.* 2007; Mollinari *et al.* 2009) was used for linkage mapping. All polymorphic SNPs and individuals were considered in the analysis as they passed the missing data threshold of 5% and 1%, respectively. SNPs were clustered into LGs on the basis of a LOD score >10. The LG names were defined on the basis of previously mapped loci (Chancerel *et al.* 2013). The RECORD algorithm (Os *et al.* 2005) was used to order markers within LGs, with the following parameters: LOD = 3 and max.rf = 0.4. Recombination rates were converted into genetic distances (cM) with the Kosambi mapping function (Kosambi 1943). The goodness of fit of SNP segregations to the expected Mendelian segregation ratio (i.e. 1:2:1 for an F2 population) was assessed in chi-squared tests, with adjustment of the significance threshold for simultaneous multiple tests (Benjamini & Yekutieli 2001) within each LG. The same procedure was also applied to a previous SNP data set genotyped in the same F2 family, but with different genotypes and mapping software (Chancerel *et al.* 2013). The two genetic maps were then combined into a composite linkage map with the R package *LPmerge* (Endelman & Plomion 2014).

Populations of unrelated individuals and genetic diversity estimation. The French Atlantic coast gene pool was represented by two subsets of individuals: (i) 50 trees from two natural populations, Hourtin and Petrocq, sampled from a clonal collection (CLONAPIN) established directly from the source populations Hourtin (45°11' N, 1°09' E) and Petrocq (44°04' N, 1°18' E) are coastal popu-

lations growing at low altitude (<30 m a.s.l.) and under typical maritime climate (annual precipitation of 980–1248 mm and mean annual temperature of 12–13 °C); and (ii) 50 elite trees from a larger set of about 600 genotypes mass-selected in natural forests of the Aquitaine region in the early 1960s to constitute the first generation of the maritime pine breeding program (Illy 1966). These 50 genotypes constitute the founders of a three-generation pedigreed population used to develop proof of concept for genomic selection in maritime pine (Isik *et al.* 2015).

The Portuguese population was also represented by two subsets of genotypes: (i) 19 trees sampled from two provenances in a provenance trial carried out at Mimizan (France: 44°20' N, 1°28' W). These provenances were described by Illy (1966): six trees were from seeds collected at 'Pinhal de Leiria' (an 11 ha coastal forest located at 39°79' N, 8°98' W; 0–50 m a.s.l.; annual precipitation of 700–900 mm) and 13 trees were from seeds collected at 'Trás-os-Montes' (a mountain forest located at 41°57' N, 7°50' W; 1150 m a.s.l., annual precipitation of 1200–1400 mm); and (ii) 23 additional trees from 'Pinhal de Leiria' collected from the CLONAPIN maritime pine collection.

We first tested departure from Hardy–Weinberg equilibrium, using standard chi-squared tests with a nominal significance threshold of $1.05\text{--}1.15 \times 10^{-5}$, corresponding to an experiment-wise type I error of 5% (Bonferroni correction to account for multiple testing). We then estimated three genetic diversity parameters for each SNP: minor allele frequency (MAF), observed (H_O) and expected heterozygosity (H_E , Nei's index of genetic diversity corrected for sample size, Nei 1987). These three parameters were highly correlated (>0.95). We therefore present data for MAF and H_E only. Genetic differentiation (F_{ST}) was assessed for pairs of populations, with a set of 3981 polymorphic SNPs common to both populations, using GDA software (Lewis & Zaykin 2001). Finally, Bayesian clustering analysis (STRUCTURE software, Pritchard *et al.* 2000; Falush *et al.* 2003) was performed to identify different gene pools in maritime pine, as described by Jaramillo-Correa *et al.* (2015).

Folded site-frequency spectrum for the French and Portuguese groups

Sites with missing data were excluded from the samples, to ensure general consistency across loci in terms of the number of sequences analysed for each site. Furthermore, we retained the same SNPs for the French and Portuguese groups, 3513 SNPs in total, corresponding to 88% of the original data set (3981 SNPs). Finally, two individuals were discarded from the French group as they presented 19% (INIA_PET11) and 10% (INIA_PET12)

missing data across the 3981 SNPs. The estimate of folded site-frequency spectrum (SFS) was therefore based on 3513 common SNPs with no missing data, genotyped in 94 French and 42 Portuguese individuals.

SNPs were analysed as unpolarized (i.e. no ancestral state inferred), with the least frequent nucleotide at a given SNP site considered to correspond to the minor allele and thus used for the calculation of MAF. MAF was calculated for the French and Portuguese groups separately. The folded SFS was then plotted from the observed MAF and from the expected SFS. We used equation 6 from the paper by Ganapathy & Uyenoyama (2009) to determine the expected SFS. This equation is widely used to describe the expected SFS for a genomic sample of SNPs, each of which is assumed to correspond to a mutation on an independent gene genealogy.

DNA extraction and genotyping

Developing needles were collected after bud burst for each genotype of the F2, natural and breeding populations. They were stored at -80°C or dried with silica gel before DNA extraction. About 30–40 mg of frozen/dried needles were crushed with a mixer mill (Retsch MM300, Haan, Germany). Genomic DNA was isolated with the Invisorb DNA plants 96 kit from Invitex (GmbH, Berlin, Germany), according to the manufacturer's instructions. All concentrations were determined with a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and a fluorescence assay (Quant-IT kit; Invitrogen, Carlsbad, CA, USA). Samples with DNA concentrations $>50\text{ ng}/\mu\text{L}$ (based on fluorescence measurements) were used for the Infinium assay. Genotyping was carried out at the Genes Diffusion Facility (Douai, France).

Results

SNP detection

The initial SNP genotyping array included 8949 SNPs, but only 8410 were assayed because of the loss of bead types during array manufacture. A total of 4712 (i.e. 52.65%) were already available (166 from Lepoittevin *et al.* 2010; 835 SNPs from Chancerel *et al.* 2011; 3378 SNPs from Chancerel *et al.* 2013, 434 SNPs from de Miguel *et al.* 2014 and 184 unpublished SNPs; Fig. S1, Supporting information). The other 4237 (i.e. 47.35%) corresponded to newly discovered SNPs.

Newly discovered SNPs

In silico SNPs from 454 sequence reads: Pyrosequencing of the C (Corsica), M (Morocco) and L (Landes) libraries

provided 188 494, 611 673 and 458 882 reads, respectively (Fig. 1). After cleaning, we retained 107 089, 484 348 and 225 842 reads for C, L and M, respectively. These reads were aligned on the PineV2 unigene. We identified 2039 SNPs in the C data set, 4937 in the L data set and 4752 in the M data set. We then selected SNPs in a test-cross configuration (1:1 segregation), that is heterozygous in one parent and homozygous in the other, because, in a full-sib family with biallelic codominant markers, pairs of markers presenting this configuration provide the best estimate of recombination rates, particularly if compared with pairs having a test-cross and intercross (1:2:1) configuration (Plomion *et al.* 1997). Thus, we selected 2830, 1695 and 2957 SNPs for the C $\times$ M, L $\times$ C and L $\times$ M crosses, respectively. In total, 784, 2046 and 911 SNPs were informative for the C, M and L parents, respectively. This procedure resulted in the retention of 3741 unique SNPs as particularly suitable for linkage mapping.

In vitro and in silico SNPs from candidate genes: Of the 281 candidate genes, 166 were associated with at least one contig of the maritime pine UniGene (Fig. 2). We retrieved 888 SNPs from these 166 contigs. Finally, after removing duplicated sequences and SNPs located $<100\text{ bp}$ apart, we retained 391 SNPs for SNP array development. In addition, 105 *in vitro* SNPs from 66 resequenced candidate genes were identified and incorporated into the SNP array.

Annotation of synonymous and nonsynonymous substitutions

In total, 4257 (47.6%) SNPs were found in contigs for which a full-length protein was predicted, with 1557 (36.6%) of these SNPs located in noncoding regions. SNPs in coding sequences (2700, i.e. 63.4%) were characterized in terms of their synonymous (S) or nonsynonymous (NS) nature and codon responsible for the NS change (Fig. S2A and Table S3, Supporting information). We identified 567 (21%) synonymous SNPs and found that the point mutation affected the third position in the codon in 87.5% of cases. There were 2133 (79%) nonsynonymous SNPs with the mutation affecting the first codon position in 21.5% of cases and the second codon position in 70.6%. For NS mutations, the change in the protein is indicated in Table S3 (Supporting information). Most of the SNPs were transitions (59.3%). The most common transition was C $\leftrightarrow$ T and the most common transversion G $\leftrightarrow$ T (Fig. S2B, Supporting information).

Linkage mapping

Of the 8410 SNPs used to genotype the 92 trees of the F2 mapping population, 4634 (55.1%) were monomor-

phic (including 2231 SNPs, about half, from the newly developed set), 2052 (from 1672 different contigs) were polymorphic (24.4%) and the remaining 1724 (20.5%) corresponding to failed assays due to poor signal or poor clustering performance as already presented in Chancerel *et al.* (2013). While lower failure rates (<10%) were usually obtained in crop plants where more information are available to optimize the selection of the SNPs (e.g. Delourme *et al.* 2013 in oilseed rape, Sim *et al.* 2012 in tomato, Li *et al.* 2014 in alfalfa), a similar failure rate was reported in oak (19.6%; Lepointevin *et al.* 2015) and Douglas-fir (27.5%; Howe *et al.* 2013) using the iselect Infinium genotyping platform. Besides, as expected, the rate of failed assays was much lower in the set of already validated assays (11.9%) compared to that found in the set of newly developed ones (27.3%). All polymorphic SNPs passed the quality threshold and were therefore used in linkage analysis. The total map length was 1993 cM, spread over 12 LGs corresponding to the haploid number of chromosomes for the maritime pine genome. LG length ranged from 129.2 (LG11) to 198.2 cM (LG5), and the mean number of SNPs per LG was 171 cM (F2_N in Table 1). The mean SNP density was 0.98, but more than half of the markers were grouped into clusters (i.e. groups of markers displaying no recombination). This resulted in 901 unique positions, separated by a mean distance of 2.25 cM. Two factors could account for the clustering of SNPs: (i) the small size of the mapping population (92 genotypes) and (ii) the presence of more than one SNP in 260 of the 1672 mapped contigs. Distorted SNPs were retained in the linkage analysis and accounted for 3.6% of the mapped markers. They were grouped together in five regions located on three LGs (segregation distortion regions, SDRs): two linked SDRs on LG#3 with 10 and 5 markers, one SDR on LG#6 (26 markers) and three SDRs on LG#10 (14, 12 and 2 markers; Fig. 3). The order of SNPs was similar (Spearman $\rho = 0.98$) to that obtained from the linkage map reconstructed with the data set of Chancerel *et al.* (2013) based on different genotypes of the same progeny (F2_O in Table 1). Only 56 of the 1180 markers common to the two genetic maps differed in order between the two maps (Fig. 3). These differences in marker order occurred between tightly linked markers (<5 cM), mostly on LG#1 and LG#10, for which 10 and 15 markers were involved, respectively. Finally, based on the highly conserved marker order between the two maps, a composite map was established (F2_C in Table 1) with LPmerge software. It contained 2353 SNPs (including 1121 SNPs in the same contigs). This composite map included 1661 different loci and is therefore the densest yet gene-based linkage map for this species. The

Table 1 Characteristics of the F2 genetic linkage maps

	LG	1	2	3	4	5	6	7	8	9	10	11	12	Total
F2_N	Size (cM)	157.7	197.2	178.5	180.6	198.2	134.8	161.2	166.9	151.17	184.4	126.2	156	1992.9
	Number of SNPs	157	152	190	176	199	163	181	155	184	148	156	191	2052
	Number of unique positions	65	81	88	77	80	80	76	79	74	66	68	67	901
	Distance between SNPs (cM)	1.01	1.31	0.94	1.03	1	0.83	0.9	1.08	0.83	1.25	0.81	0.82	0.98
	Distance between unique positions (cM)	2.46	2.47	2.05	2.38	2.51	1.71	2.15	2.14	2.07	2.84	1.88	2.36	2.25
F2_O	Distorted SNPs (%)	0	0	7.9	0	0	16	0	0	0	18.9	0	0	3.6
	Size (cM)	121.8	160.1	158.4	141.5	160.5	129.7	144.6	139.7	130.6	149.9	113.2	137.6	1687.6
	Number of SNPs	106	117	147	138	126	113	124	93	134	107	130	146	1481
	Number of unique positions	41	57	60	54	56	58	58	47	45	54	49	50	629
	Distance between SNPs (cM)	1.16	1.38	1.09	1.03	1.28	1.16	1.18	1.5	0.98	1.41	0.88	0.95	1.17
F2_C	Distance between unique positions (cM)	3.04	2.86	2.69	2.67	2.92	2.28	2.54	3.04	2.97	2.83	2.36	2.81	2.75
	Distorted SNPs (%)	0	0	0	0	0	0	0	0	0	33	0	0	2.75
	Size (cM)	124.1	160.1	159.3	142.2	160.5	131.3	144.6	139.7	132.2	155.4	123.7	138.8	1711.7
	Number of SNPs	179	177	221	191	219	188	201	174	207	173	194	229	2353
	Number of unique positions	64	81	95	75	88	88	83	81	75	76	74	75	955
	Distance between SNPs (cM)	0.7	0.91	0.72	0.75	0.74	0.7	0.72	0.81	0.64	0.9	0.64	0.61	0.74
	Distance between unique positions (cM)	1.97	2	1.69	1.92	1.84	1.51	1.76	1.75	1.79	2.07	1.69	1.88	1.82

F2_N: map obtained from the 9 k-array (this study), F2_O: map constructed with data from Chancerel *et al.* (2013) and F2_C: composite map generated from the F2_N and F2_O maps.

Fig. 3 Comparison between the two F2 genetic linkage maps (F2_N in blue and F2_O in white). Segregation distortion regions are indicated in orange.

total map length was 1711.7 cM. There were 955 unique positions, separated by a mean distance of 1.82 cM.

Genetic diversity and population structure analysis

Only a few markers displayed significant departure from Hardy–Weinberg equilibrium in each population: six (0.10%) for the French Atlantic-based breeding population, 16 (0.26%) for the natural French Atlantic population and 18 (0.35%) for the Portuguese population. These markers, together with those for which >10% of the data were missing in any of the three populations, were removed from further analyses, resulting in a set of 5016 SNPs (including 24% monomorphic loci) successfully scored in all populations. We found no differences in genetic diversity, assessed with this common SNP set, between the natural and breeding populations in the French Atlantic region (H_E of 0.336 vs. 0.332, based on polymorphic loci only; Table 2), suggesting that the mass selection in natural forests for constitution of the base breeding population for this ecotype was broad enough

to collect most of the standing genetic variation. This is further supported by the absence of significant genetic differentiation between the natural and mass-selected populations ($F_{ST} = 0.0005$, 95% CI: -0.00002 , 0.00107). Genetic diversity was slightly lower for the Portuguese population (H_E of 0.319) than for both the breeding and natural populations in France. High levels of genetic differentiation were observed between the French Atlantic and Portuguese populations ($F_{ST} = 0.0847$, 95% CI: 0.08087 , 0.08864), with 263 SNPs having F_{ST} above 0.25. These SNP markers are good candidates to replace the biochemical assay (based on terpene content analysis, Baradat & Marpeau-Bezard 1988) currently used to determine the putative origin of adult forest stands in Aquitaine before the collection of seeds and their distribution for commercial purposes in France. Indeed, seed-lots from Portugal were introduced into Aquitaine in the 1950s and the stands they formed suffered frost damage after the exceptionally cold winter of 1985 (Ribeiro *et al.* 2002). Therefore, from 1986 onwards, candidate stands for seed collection in Aquitaine had to be certified as of French origin. The Bayesian clustering pattern observed,

Table 2 Genetic diversity estimates for French Atlantic and Portuguese maritime pine populations

Population	<i>n</i>	% Mono	All SNPs			Common polymorphic SNPs		
			Loci #	MAF	H_E	Loci #	MAF	H_E
French Atlantic								
Breeding population	46	0.242	5016	0.201	0.266	3981	0.254	0.336
Natural population	50	0.241	5016	0.198	0.263	3981	0.250	0.332
Portugal	42	0.227	5016	0.187	0.253	3981	0.236	0.319

Fig. 4 Bayesian clustering analysis for $K = 2$, showing the French Atlantic (red) and Portuguese (green) maritime pine gene pools. The red and green portions of each bar (individual tree) indicate the probability of genetic ancestry (y -axis) for each cluster.

with the identification of only two gene pools ($K = 2$), including the French natural and breeding populations grouped together into a single gene pool significantly separated from the Portuguese population (Fig. 4), is a strong asset in this respect. This result will guide the development of highly informative genotyping multiplexes based on the Sequenom MassARRAY technology that has already been optimized for this species (Chancerel *et al.* 2013).

As expected, the SFS of both the French and Portuguese populations showed a deficit of the observed low-frequency variants (Fig. S3, Supporting information). This deficit is probably due to a combined effect of the approach developed to select most of the SNPs (i.e. in silico analysis of ESTs obtained from small numbers of genotypes, precluding the capture of rare variants), and the sampling scheme (a small population size), hindering the detection of rare variants even if present on the SNP array.

Discussion

In this section, we discuss how the substantial number of polymorphisms detected and validated in this study might help researchers working on maritime pine to address scientific questions in population genomics and quantitative genetics with large samples of individuals from natural and pedigreed populations that can now be surveyed for genetic variation at a few thousand expressed genes across the genome. We also discussed the limitations that this SNP array represents for genetic applications.

Linkage and QTL mapping

High-throughput SNP genotyping has been successfully implemented in several conifers (Pavy *et al.* 2008; Eckert *et al.* 2009a; Chancerel *et al.* 2011, 2013). This approach has generated thousands of markers for genetic mapping, significantly improving our understanding of large and complex conifer genomes (Jermstad *et al.* 2011; Pavy *et al.* 2012; Martínez-García *et al.* 2013; Neves *et al.* 2014). We show here that a combination of highly multiplexed

SNP genotyping and a selective mapping strategy is useful for the cost-efficient mapping of hundreds of genes, as the bin set of 92 highly recombinant F2s represented 95% of the recombination occurring among the initial set of 477 genotypes used to construct the framework linkage map. The genetic map obtained with this strategy was highly reliable and accurate, as revealed by comparison with a previously developed map based on different genotypes from the same progeny (Chancerel *et al.* 2013). A low fraction of molecular markers were distorted, and they did not alter marker order. The use of markers common to the two sets of offspring made it possible to construct a composite genetic linkage map confirming the positions of most of the previously mapped markers. The composite linkage map obtained in this study is the densest gene-based map for maritime pine and is in the range of recently published linkage maps for other conifer species (Pavy *et al.* 2012; Neves *et al.* 2014). However, a considerable fraction of SNPs clustered in the same position. Once high-throughput genotyping of thousands of molecular markers has been achieved, the main limitations for genetic linkage mapping will come from the high number of individuals needed to detect recombination events between closely positioned genes (Bartholomé *et al.* 2015).

Association mapping

In long-lived outbreeding organisms, such as forest trees, association mapping is an attractive alternative to QTL mapping, as it can be used to identify molecular variation underlying phenotypes from multiple genetic backgrounds without the need to produce segregating progenies (Neale & Savolainen 2004). Under certain circumstances (e.g. moderate-high heritability), association mapping can also be applied to natural populations. This approach is useful for studies of adaptive traits in the precise environments in which they evolved (see Parchman *et al.* 2012; Budde *et al.* 2014, for forest trees). Recent years have seen an explosion in the number of association mapping studies in model organisms, such as humans and *Arabidopsis* (see reviews in Stranger *et al.* 2011; Weigel 2012). However, only a few studies, mostly

focusing on technological characters associated with wood quality, have been carried out on forest trees (Gonzalez-Martinez *et al.* 2011 and references therein; Cappa *et al.* 2013; Guerra *et al.* 2013). Progress has been particularly slow in maritime pine, partly due to the lack of a reliable and cost-efficient genotyping platform capable of dealing with the large sample sizes required to detect small- and medium-sized allelic effects. However, promising results were obtained in the first genetic association studies in maritime pine, based on only a few hundred markers (Lepointevin *et al.* 2012; Cabezas *et al.* 2015). Apart from including all SNPs reported to display significant genotype:phenotype associations in previous studies, our new SNP array provides the largest genotyped platform developed to date in maritime pine, with 7252 workable SNPs. Despite its modest size, this resource has the potential to stimulate new association studies for a wide range of adaptive and production traits in this species.

Given the large size of conifer genomes (24.5 Gb for maritime pine, Chagné *et al.* 2002), our SNP array would only covers a small fraction of the potential relevant variation underlying phenotypes. However, this assay is enriched in SNPs from candidate genes, based on all the available information (published or unpublished) concerning genes displaying signatures of natural selection (Eveno *et al.* 2008; Grivet *et al.* 2011), involved in environmental associations (Jaramillo-Correa *et al.* 2015) or displaying differential expression (e.g. for pine nematode resistance, Santos *et al.* 2012; or drought response, Perdiguerro *et al.* 2013) in this species. We therefore expect the use of this SNP genotyping array to generate highly informative data. For example, Westbrook *et al.* (2013) found that a small number of significantly associated SNPs (~20 to 30) had the same predictive power as the full data set (4854 SNPs) in SNP-based models for oleoresin flow in *Pinus taeda* L. Finally, this SNP array will allow unprecedented explorations of the molecular basis of polygenic quantitative traits, through the implementation of multilocus association models (e.g. piMASS, Guan & Stephens 2011). Although largest SNP assays would be desirable, some thousands of well-selected SNPs have been found to be enough to provide relevant insights into polygenic adaptation patterns (Berg & Coop 2014).

Genomic selection

High-throughput genotyping platforms, such as the new SNP array developed here, can guide breeding and selection decisions (Eggen 2012). Genomic selection (GS) is a paradigm shift first introduced into animal breeding in 2008 for the selection of superior individuals in many countries (Goddard 2009). The major difference between

GS and marker-assisted selection (MAS) is the number of markers used. GS makes use of a much larger number of markers to trace all the QTLs with small or large effect (Hayes & Goddard 2010). SNP-based genotyping platforms are reliable and repeatable for the genotyping of large numbers of individuals. There are few missing genotypes, and these genotypes can be handled (dropped or imputed) without markedly decreasing data quality. Thus, SNP arrays have become the choice genotyping platform for animal breeding and human genetics studies, despite advances in the efficiency of DNA sequencing technologies. GS is expected to revolutionize forest tree breeding, by decreasing the need for expensive and time-consuming progeny-testing practices. If successful, GS could halve the long breeding cycles (>15 years) of forest trees and double the genetic gain per unit time (Isik 2014). One probable application of GS in forest trees would involve the use of data from different sources, such as progeny tests and genotyping centres, in a single-step approach to predicting the genetic merit of individuals (Legarra *et al.* 2009). The predicted model for one cycle can then be refined as new data become available. As breeding cycles progress and genotyping/sequencing costs fall, progeny testing will have a lesser effect on selection decisions. Nevertheless, the lack of a reliable and cost-efficient genotyping platform remains the major bottleneck for the routine application of GS in forest trees. The new SNP array presented here constitutes a valuable tool to carry out GS proof of concept in maritime pine population of limited effective size. However, because of the relatively largebreeding effective population size (~135 for the French breeding programme), operative implementation of GS would still need further development of SNP resources. Based on deterministic simulations, Grattapaglia & Resende (2011) suggested a marker density of over 10 SNPs per cM to reach an accuracy of 0.7 with GS. This would translate to at least 15 000 SNP markers for maritime pine, half of which provided in the SNP array developed here.

Population genomics

The high-density SNP array developed in this study provides a powerful tool for the genome-wide genotyping of a large number of populations across the full distribution range of the species. The genotyping of hundreds of individuals across the entire range of maritime pine, with SNPs located both in coding and noncoding regions, would increase our understanding of the role of the evolutionary, demographic and adaptive mechanisms acting on natural populations. The available molecular markers have shown that maritime pine populations are spatially structured into regional gene pools connected by gene flow, particu-

larly in the Iberian Peninsula (Burban & Petit 2003; Bucci *et al.* 2007; Jaramillo-Correa *et al.* 2015). However, the timing of the historical events leading to this spatial separation and the degree of connectivity between gene pools remain to be determined. In addition, maritime pine grows in diverse environmental conditions, resulting in the local adaptation of populations over the range of this species (González-Martínez *et al.* 2002; Jaramillo-Correa *et al.* 2015; Serra-Varela *et al.* 2015), but the molecular mechanisms underlying this adaptation are poorly understood. Our SNP array provides a means of identifying functional variation and the molecular bases of adaptation, through various methods based on the differentiation of allele frequencies between populations (F_{ST} -based methods; see e.g. Prunier *et al.* 2011; Chen *et al.* 2012), SFS statistics (Eckert *et al.* 2009b), correlations with environmental variables (Eckert *et al.* 2010; Jaramillo-Correa *et al.* 2015) or combinations of these approaches. 'Reverse' ecology approaches connecting genomic data with environmental parameters have also proved useful for identifying the major ecological drivers of adaptation (Levy & Borenstein 2012). Ascertainment biases, resulting from the small panel of individuals from which SNPs were obtained (ascertainment width) and from the stringent criteria in terms of minimum allele frequency and read coverage used to retain the SNPs (ascertainment depth), must be taken into account when interpreting the results of future studies of natural populations. A systematic bias would be expected for estimates of nucleotide diversity, inferences about population structure, evolutionary processes based on natural selection and/or historical demographic models, particularly if based on the SFS (Albrechtsen *et al.* 2011; reviewed by Helyar *et al.* 2011). Ascertainment bias has the potential of skewing the SFS towards common alleles, as is the case for our SNP array. A correction, explicitly incorporating SNP ascertainment bias into population genetics models, or the use of methods robust to ascertainment bias are necessary to accurately assess population genetic inferences (Albrechtsen *et al.* 2011; Excoffier *et al.* 2013). Furthermore, most of the SNPs were obtained from independent contigs of the UniGene, precluding the use of linkage disequilibrium (LD) and haplotype diversity as metrics for population genetic inferences. This limitation could therefore potentially reduce the success in identifying loci underlying local adaptation, as this approach – similarly to genome-wide association – exploits LD to indirectly identify adaptive SNPs by relying upon the premise that some markers in LD with a causal SNP should be associated with important ecological variables as well. Despite these limitations, thanks to the methodological advances, our SNP

array should allow to interpret the signatures left in the genome by different evolutionary forces, and to make inferences about complex population processes and geographic patterns of genome-wide genetic variation.

Conclusion and future direction in genomic resource development

In this study, we report the discovery of new SNP markers in maritime pine, from RNA-seq and amplicon resequencing data, and the establishment of an Infinium genotyping array including SNPs that have already been validated. We found this technology highly reliable considering the extremely low level of missing data compared to alternatives such as restriction site-associated DNA sequencing (RAD-Seq, Baird *et al.* 2008), which made unnecessary imputing data. Infinium technology is also unaffected by allele drop out, a major drawback for genetic inferences with RAD-Seq in highly heterozygous species such as maritime pine (Davey *et al.* 2013; Gautier *et al.* 2013; Puritz *et al.* 2014; Mastretta-Yanes *et al.* 2015). The SNP array presented here represents a major step forward for population and conservation genetics and for breeding in maritime pine; however, we also discussed a number of limitations, mostly derived from the large conifer genome sizes and low levels of LD.

In the near future, we will develop new generation genotyping-by-sequencing methods, which should make it possible to decrease the problem of ascertainment bias, because the individuals of interest are sequenced directly. Considering missing data allele drop out concerns, future development of marker technology in this species will be based on sequence capture and direct sequencing (Gnirke *et al.* 2009).

Acknowledgements

This study was carried out with financial support from the European Union Sixth and Seventh Framework Programmes: Evoltree (no. 016322), Procogen (no. 289841), Noveltree (no. 211868), Forger (no. KBBE-289119) and TipTree (BiodivERSA-ERANET); the French National Research Agency: ANR-FLAG (ANR-12ADAP-007-01); the Italian Ministry of Education, University and Scientific Research: Biodiversitalia (RBA-P10A2T4); and the Spanish Ministry of Science and Innovation: AdapCon (CGL2011-30182-C02-01). We thank C García-Barriga for DNA extraction. D Grivet was supported by a 'Ramón y Cajal' Fellowship from the Spanish Ministry of Science and Innovation, I Lesur by ANR-FLAG and C Boury by Procogen. H Lagrault received a PhD fellowship from INRA (EFPA division and ACCAF metaprogramme) and Région Aquitaine (no. 20111203004). J Bartholomé was supported by a postdoctoral fellowship from 'Conseil Général des Landes'. I Rodríguez-Quilón acknowledges a PhD scholarship (FPI-INIA) from the Spanish

National Institute for Agricultural and Food Research and Technology and the Spanish Ministry of Science and Innovation.

References

- Albrechtsen A, Nielsen FC, Nielsen R (2011) Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution*, **27**, 2534–2547.
- Alía R, Martin S (2003) *EUFORGEN Technical Guidelines for Genetic Conservation and Use for Maritime Pine (Pinus pinaster)*, 6 p. IPGRI, Rome, Italy.
- Alonso P, Cortizo M, Cantón FR *et al.* (2007) Identification of genes differentially expressed during adventitious shoot induction in *Pinus pinea* cotyledons by subtractive hybridization and quantitative PCR. *Tree Physiology*, **27**, 1721–1730.
- Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, **3**, e3376.
- Baradat P, Marpeau-Bezard A (1988) *Le pin maritime, Pinus pinaster Ait., biologie et génétique des terpènes pour la connaissance et l'amélioration de l'espèce*. PhD Thesis, University of Bordeaux I.
- Bartholomé J, Mandrou E, Mabila A *et al.* (2015) High-resolution genetic linkage maps of *Eucalyptus* improve BRASUZI reference genome assembly. *New Phytologist*, **206**, 1283–1296.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.
- Berg JJ, Coop G (2014) A population genetic signal of polygenic adaptation. *PLoS Genetics*, **10**, e1004412.
- Bouffier L, Annie Raffin A, Alía R (2013) Maritime pine – *Pinus pinaster* Ait. In: *Best Practice for Tree Breeding in Europe* (eds Mullin TJ, Lee SJ), pp. 65–75. Skogforsk, Uppsala, Sweden.
- 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, Heuertz M, Hernández-Serrano A *et al.* (2014) *In situ* genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster* Aiton). *New Phytologist*, **201**, 230–241.
- Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology*, **12**, 1487–1495.
- Cabezas JA, González-Martínez SC, Collada C *et al.* (2015) Nucleotide polymorphisms in a pine ortholog of the *Arabidopsis* degrading enzyme cellulase KORRIGAN are associated with early growth performance in *Pinus pinaster*. *Tree Physiology*. doi:10.1093/treephys/tpv050.
- Canales J, Bautista R, Label P *et al.* (2014) *De novo* assembly of maritime pine transcriptome: implications for forest breeding and biotechnology. *Plant Biotechnology Journal*, **12**, 286–299.
- Cappa EP, El-Kassaby YA, Garcia MN *et al.* (2013) Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS One*, **8**, e81267.
- Chagné D, Lalanne C, Madur D *et al.* (2002) A high density linkage map of *Pinus pinaster* based on AFLPs. *Annals of Forest Science*, **59**, 627–636.
- 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.
- Chancerel E, Lamy JB, Lesur I *et al.* (2013) High density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biology*, **11**, 50.
- Chen J, Källman T, Ma X *et al.* (2012) Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics*, **191**, 865–881.
- Davey JW, Cezard T, Fuentes-Utrilla P *et al.* (2013) Special features of RAD sequencing data: implications for genotyping. *Molecular Ecology*, **22**, 3151–3164.
- Delourme R, Falentin C, Fomeju BF *et al.* (2013) High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC Genomics*, **14**, 120.
- Eckert AJ, Pande B, Ersoz ES *et al.* (2009a) High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genetics and Genomes*, **5**, 225–234.
- Eckert AJ, Wegrzyn JL, Pande B *et al.* (2009b) Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*). *Genetics*, **183**, 289–298.
- Eckert AJ, van Heerwaarden J, Wegrzyn JL *et al.* (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, **185**, 969–982.
- Eggen A (2012) The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers*, **2**, 10–15.
- Endelman JB, Plomion C (2014) LPMERGE: an R package for merging genetic maps by linear programming. *Bioinformatics*, **30**, 1623–1624.
- Eveno E, Collada C, Guevara MA *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.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**, e1003905.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
- Ganapathy G, Uyenoyama MK (2009) Site frequency spectra from genomic SNP surveys. *Theoretical Population Biology*, **75**, 346–354.
- Gautier M, Gharbi K, Cezard T *et al.* (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, **21**, 3165–3178.
- Gnirke A, Melnikov A, Maguire J *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, **27**, 182–189.
- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, **136**, 245–257.
- Gonzalez-Martinez SC, Dillon S, Garnier-Géré P *et al.* (2011) Patterns of nucleotide diversity and association mapping. In: *Genetics, Genomics and Breeding of Conifer Trees* (eds Plomion C, Bousquet J, Kole C), pp. 239–275. Edenbridge Science Publishers and CRC Press, New York.
- González-Martínez SC, Alía R, Gil L (2002) Population genetic structure in a Mediterranean pine (*Pinus pinaster* Ait.): a comparison of allozyme markers and quantitative traits. *Heredity*, **89**, 199–206.
- Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. *Tree Genetics and Genomes*, **7**, 241–255.
- 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.
- Guan Y, Stephens M (2011) Bayesian variable selection regression for genome-wide association studies, and other large-scale problems. *The Annals of Applied Statistics*, **5**, 1780–1815.
- Guerra FP, Wegrzyn JL, Sykes R *et al.* (2013) Association genetics of chemical wood properties in black poplar (*Populus nigra*). *New Phytologist*, **197**, 162–176.
- Harfourche A (1995) *Variabilité géographique et hybridation interrassiale chez le pin maritime (Pinus pinaster Ait.)*, 153 p. Thèse de Doctorat, Université Henri Poincaré Nancy I, France.
- Hayes B, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. *Genome*, **53**, 876–883.
- Helyar SJ, Hemmer-Hansen J, Bekkevold D *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, **11**, 123–136.
- Howe GT, Yu J, Knaus B *et al.* (2013) A SNP resource for Douglas-fir: *de novo* transcriptome assembly and SNP detection and validation. *BMC Genomics*, **14**, 137.
- Ily G (1966) Recherches sur l'amélioration génétique du pin maritime. *Annal des Sciences Forestières*, **23**, 769–948.

- Isik F (2014) Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests*, **45**, 379–401.
- Isik F, Bartholomé J, Farjat A *et al.* (2015) Genomic selection in maritime pine. *Plant Science*. doi: 10.1016/j.plantsci.2015.08.006
- 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.
- Jaramillo-Correa JP, Grivet D, Lepoittevin C *et al.* (2015) Molecular proxies of climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics*, **199**, 793–807.
- Jermstad KD, Eckert AJ, Wegrzyn JL *et al.* (2011) Comparative mapping in *Pinus*: Sugar pine (*Pinus lambertiana* Dougl.) and loblolly pine (*Pinus taeda* L.). *Tree Genetics and Genomes*, **7**, 457–468.
- Jorge FC, Pascoal Neto C, Irle C *et al.* (2002) Wood adhesives derived from alkaline extracts of maritime pine bark: preparation, physical characteristics and bonding efficacy. *European Journal of Wood and Wood Products*, **60**, 303–310.
- Kosambi DD (1943) The estimation of map distances from recombination values. *Annals of Human Genetics*, **12**, 172–175.
- Kujala S, Savolainen O (2012) Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): signs of clinal adaptation? *Tree Genetics and Genomes*, **8**, 1451–1467.
- Lalagüe H, Csilléry K, Oddou-Muratorio S *et al.* (2014) Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (*Fagus sylvatica* L.) population from southeastern France. *Tree Genetics and Genomes*, **10**, 15–26.
- Lamy JB, Bouffier L, Burlett R, Plomion C, Cochard H, Delzon S (2011) Uniform selection as the primary evolutionary force of cavitation resistance across a species range. *PLoS One*, **6**, e23476.
- Lamy JB, Delzon S, Bouche P *et al.* (2014) Limited genetic variability and phenotypic plasticity for cavitation resistance in a Mediterranean pine. *New Phytologist*, **201**, 874–886.
- Le Provost G, Domergue F, Lalanne C *et al.* (2013) Cuticular wax: an essential component of fast-growing maritime pine saplings to cope with water deficit. *BMC Plant Biology*, **13**, 95.
- Lefèvre F, Boivin T, Bontemps A *et al.* (2014) Considering evolutionary processes in adaptive forestry. *Annals of Forest Science*, **71**, 723–739.
- Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, **92**, 4656–4663.
- Lepoittevin C, Frigerio JM, Garnier-Géré P *et al.* (2010) In vitro vs. in silico detected SNPs for the development of a genotyping array: what can we learn from a non-model species? *PLoS One*, **5**, e11034.
- 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 and Genomes*, **8**, 113–126.
- Lepoittevin C, Bodénès C, Chancerel E *et al.* (2015) Single-nucleotide polymorphism discovery and validation in high density SNP array for genetic analysis in European white oaks. *Molecular Ecology Resources*. doi:10.1111/1755-0998.12407.
- Levy R, Borenstein E (2012) Reverse ecology: from systems to environments and back. *Advances in Experimental Medicine and Biology*, **751**, 329–345.
- Lewis PO, Zaykin D (2001) Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>.
- Li X, Han Y, Wei Y *et al.* (2014) Development of an alfalfa SNP array and its use to evaluate patterns of population structure and linkage disequilibrium. *PLoS One*, **9**, e84329.
- Margarido GRA, Souza AP, Garcia AAF (2007) OneMap: software for genetic mapping in outcrossing species. *Heredity*, **144**, 78–79.
- Mariette J, Noirot C, Klopp C (2011) Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool. *BMC Research Notes*, **4**, 149.
- Martínez-García PJ, Stevens KA, Wegrzyn JL *et al.* (2013) Combination of multipoint maximum likelihood (MML) and regression mapping algorithms to construct a high-density genetic linkage map for loblolly pine (*Pinus taeda* L.). *Tree Genetics and Genomes*, **9**, 1529–1535.
- Mastretta-Yanes A, Arrigo N, Alvarez N *et al.* (2015) Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.
- de Miguel M, Cabezas JA, de María N *et al.* (2014) Genetic control of functional traits related to photosynthesis and water use efficiency in *Pinus pinaster* Ait. drought response: integration of genome annotation, allele association and QTL detection for candidate gene identification. *BMC Genomics*, **15**, 464.
- Mollinari M, Margarido GRA, Vencovsky R, Garcia AAF (2009) Evaluation of algorithms used to order markers on genetic maps. *Heredity*, **103**, 494–502.
- Mullin TJ, Andersson B, Bastien JC *et al.* (2011) Economic importance, breeding objectives and achievements. In: *Genetics, Genomics and Breeding of Conifer Trees* (eds Plomion C, Bousquet J, Kole C), pp. 40–127. Edenbridge Science Publishers and CRC Press, New York.
- Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science*, **9**, 325–330.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Neves LG, Davis JM, Barbazuk WB, Kirst M (2014) A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *Genes, Genomes, Genetics*, **4**, 29–37.
- Os H, Stam P, Visser RF, Eck H (2005) RECORD: a novel method for ordering loci on a genetic linkage map. *Theoretical and Applied Genetics*, **112**, 30–40.
- Ouborg NJ, Pertoldi C, Loeschcke V *et al.* (2010) Conservation genetics in transition to conservation genomics. *Trends in Genetics*, **26**, 177–187.
- 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.
- Pavy N, Pelgas B, Beauseigle S *et al.* (2008) Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics*, **9**, 21.
- Pavy N, Pelgas B, Laroche J *et al.* (2012) A Spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biology*, **10**, 84.
- Pavy N, Gagnon F, Rigault P *et al.* (2013) Development of high-density SNP genotyping arrays for white spruce (*Picea glauca*) and transferability to subtropical and nordic congeners. *Molecular Ecology Resources*, **13**, 324–336.
- Perdiguerro P, Barbero MdelC, Cervera MT *et al.* (2013) Molecular response to water stress in two contrasting Mediterranean pines (*Pinus pinaster* and *Pinus pinea*). *Plant Physiology and Biochemistry*, **67**, 199–208.
- Plomion C, Costa P, Bahrman N (1997) Genetic analysis of needle protein in maritime pine. 1. Mapping dominant and codominant protein markers assayed on diploid tissue, in a haploid-based genetic map. *Silvae Genetica*, **46**, 161–165.
- Plomion C, Chancerel E, Endelman JB *et al.* (2014) Genome-wide distribution of genetic diversity and linkage disequilibrium in a mass-selected population of maritime pine. *BMC Genomics*, **15**, 171.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Prunier J, Laroche J, Beaulieu J, Bousquet J (2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology*, **20**, 1702–1716.
- Puritz JB, Matz MV, Toonen RJ *et al.* (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937–5942.
- Ribeiro MM, Le Provost G, Gerber S *et al.* (2002) Origin identification of maritime pine stands in France using chloroplast single-sequence repeats. *Annals of Forest Science*, **59**, 53–62.
- Rohdewald P (2002) A review of the French maritime pine bark extract (Pycnogenol®), an herbal medication with a diverse clinical

- pharmacology. *International Journal of Clinical Pharmacology and Therapeutics*, **40**, 158–168.
- Santos CS, Pinheiro M, Silva AI *et al.* (2012) Searching for resistance genes to *Bursaphelenchus xylophilus* using high-throughput screening. *BMC Genomics*, **13**, 599.
- 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.
- Serra-Varela MJ, Grivet D, Vincenot L *et al.* (2015) Does phylogeographic structure relate to climatic niche divergence? A test using maritime pine (*Pinus pinaster* Ait.). *Global Ecology and Biogeography*. doi: 10.1111/geb.12369.
- Sim SC, Van Deynze A, Stoffel K *et al.* (2012) High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS One*, **7**, e45520.
- Stranger BE, Stahl EA, Raj T (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, **187**, 367–383.
- Touriño S, Selga A, Jiménez A *et al.* (2005) Procyanidin fractions from pine (*Pinus pinaster*) bark: radical scavenging power in solution, antioxidant activity in emulsion, and antiproliferative effect in melanoma cells. *Journal of Agricultural and Food Chemistry*, **53**, 4728–4735.
- Vision TJ, Brown DG, Shmoys DB *et al.* (2000) Selective mapping: a strategy for optimizing the construction of high-density linkage maps. *Genetics*, **155**, 407–420.
- Wachowiak W, Balk PA, Savolainen O (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genetics and Genomes*, **5**, 117–132.
- Weigel D (2012) Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiology*, **158**, 2–22.
- Westbrook JW, Resende MFR, Munoz P *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.

I.L. and H.L. contributed to in silico SNP detection in mapping populations; I.L. to SNP array design and SNP annotation; C.B., L.B., J.M.G., F.B., G.G.V., D.G., I.R.Q. and S.C.G.M. to logistical aspects of genotyping; C.B., J.B. I.R.Q., S.C.G.M. and D.G. to SNP scoring; C.P., I.L., M.d.M., M.T.C., N.d.M. and F.H. to database compilation from published and unpublished data; J.B. to linkage mapping; S.C.G.M. and D.G. to population genetics analysis; and C.P. to design of the study and overall coordination. All the authors participated in the writing of the paper, and have read and approved the manuscript submitted.

Data accessibility

Roche 454 raw data have been deposited in the short-read archive of NCBI (accessions: SRX031589 for genotype 110-4019-1, SRX208012 for genotype 0284-2 and SRX031592 for genotype 112-4-1). The assembled contigs of the second maritime pine UniGene are available from: http://genotoul-contigbrowser.toulouse.inra.fr:9092/Pinus_pinaster2/index.html. The assembled contigs of the third maritime pine UniGene are available from: http://www.scbi.uma.es/sustainpinedb/home_page. Raw data (SNP genotypes for the breeding population, natural populations from France and Portugal, F2 mapping populations) and SNP accession numbers are available in Table S3 (Supporting information).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Number of SNPs developed over the last 5 years for maritime pine and interdependence of studies (for more details see Table S2).

Fig. S2 Synonymous and nonsynonymous substitutions.

Fig. S3 Distribution of minor allele frequency [MAF from France (A) and Portugal (B)] and folded site-frequency spectrum [SFS from France (C) and Portugal (D)].

Table S1 List of SNP arrays already developed for maritime pine and associated research questions.

Table S2 List of SNP markers already developed for maritime pine.

Table S3 Characteristics of the data sets used in the present study.

ANNEX 2

Jaramillo-Correa JP, **Rodríguez-Quilón I**, Grivet D, Lepoittevin C, Sebastiani F, Heuertz M, Garnier-Géré P, Alía R, Plomion C, Vendramin GG, *et al.* 2015. Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics* 199: 793–807.

Molecular Proxies for Climate Maladaptation in a Long-Lived Tree (*Pinus pinaster* Aiton, Pinaceae)

Juan-Pablo Jaramillo-Correa,^{*,†} Isabel Rodríguez-Quilón,^{*} Delphine Grivet,^{*} Camille Lepoittevin,^{*,§}
Federico Sebastiani,^{**,§} Myriam Heuertz,^{*,§} Pauline H. Garnier-Géré,^{*,§} Ricardo Alía,^{*}
Christophe Plomion,^{*,§} Giovanni G. Vendramin,^{**,§} and Santiago C. González-Martínez^{*,††,1}

^{*}Department of Ecology and Genetics, Forest Research Centre, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, E-28040 Madrid, Spain, [†]Department of Evolutionary Ecology, Institute of Ecology, Universidad Nacional Autónoma de México, AP 70-275 México City, Mexico, [‡]Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1202 (Biodiversité Gènes Ecosystèmes; Biogeco), F-33610 Cestas, France, [§]University of Bordeaux, Unité Mixte de Recherche 1202 Biogeco, F-33170 Talence, France, ^{**}Institute of Biosciences and Bioresources, National Research Council, I-50019 Sesto Fiorentino, Florence, Italy, and ^{††}Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland

ABSTRACT Understanding adaptive genetic responses to climate change is a main challenge for preserving biological diversity. Successful predictive models for climate-driven range shifts of species depend on the integration of information on adaptation, including that derived from genomic studies. Long-lived forest trees can experience substantial environmental change across generations, which results in a much more prominent adaptation lag than in annual species. Here, we show that candidate-gene SNPs (single nucleotide polymorphisms) can be used as predictors of maladaptation to climate in maritime pine (*Pinus pinaster* Aiton), an outcrossing long-lived keystone tree. A set of 18 SNPs potentially associated with climate, 5 of them involving amino acid-changing variants, were retained after performing logistic regression, latent factor mixed models, and Bayesian analyses of SNP–climate correlations. These relationships identified temperature as an important adaptive driver in maritime pine and highlighted that selective forces are operating differentially in geographically discrete gene pools. The frequency of the locally advantageous alleles at these selected loci was strongly correlated with survival in a common garden under extreme (hot and dry) climate conditions, which suggests that candidate-gene SNPs can be used to forecast the likely destiny of natural forest ecosystems under climate change scenarios. Differential levels of forest decline are anticipated for distinct maritime pine gene pools. Geographically defined molecular proxies for climate adaptation will thus critically enhance the predictive power of range-shift models and help establish mitigation measures for long-lived keystone forest trees in the face of impending climate change.

KEYWORDS climate adaptation; environmental associations; genetic lineages; single nucleotide polymorphisms; fitness estimates

PAST and present climate changes are major drivers of species displacements and range-size variation (Hughes 2000; Franks and Hoffmann 2012). Current predictions indicate that the impact of climate change will intensify over the next 20–100 years (Loarie *et al.* 2009; Bellard *et al.* 2012), with concomitant phenotypic and genetic effects on wild populations (Gamache and Payette 2004; Franks and Hoffmann 2012; Alberto *et al.* 2013a). The capability of species to respond to such alterations will rely on phenotypic plasticity,

potential for *in situ* adaptation, and/or migration to more suitable habitats (Aitken *et al.* 2008). While phenotypic plasticity and migration might be insufficient to cope with these changes (McLachlan *et al.* 2005; Malcom *et al.* 2011; Zhu *et al.* 2011), successful *in situ* adaptation will depend on the amount of standing genetic variation and the rate at which new alleles arise, are maintained, and/or get to fixation within populations (Hancock *et al.* 2011). Thus, our ability to detect present adaptive polymorphisms and to integrate them in predictive models of future maladaptation might be decisive to ensure the persistence of natural populations under climate change, particularly for keystone taxa (Franks and Hoffmann 2012; Kremer *et al.* 2012).

“Maladaptation” to climate refers to a decrease of the mean population fitness produced by a mismatch between the optimal and realized mean genotype frequencies, which may result from the inability to adjust to rapidly changing

Copyright © 2015 by the Genetics Society of America

doi: 10.1534/genetics.114.173252

Manuscript received July 2, 2014; accepted for publication December 20, 2014; published Early Online December 29, 2014.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.173252/-DC1>.

Nuclear microsatellite and SNP genotypes from this article have been deposited with the Dryad Digital Repository under DOI 10.5061/dryad.fb436.

¹Corresponding author: Department of Ecology and Genetics, Forest Research Centre, INIA-CIFOR, E-28040 Madrid, Spain. E-mail: santiago@inia.es

climates (Lynch and Lande 1993; Kremer *et al.* 2012). At the species range scale, such adaptation lags can generate mosaics of selected alleles and increase population differentiation at selected loci, depending on the species' life-history traits and the geographic extent at which selective pressures are operating (Savolainen *et al.* 2007; Hancock *et al.* 2011). For example, in *Arabidopsis thaliana*, a model annual selfing plant, new advantageous mutations associated with fitness and local climate occur in such mosaics, together with more common and widely distributed favorable variants (Hancock *et al.* 2011). On the other hand, evidence suggests that in widespread outcrossing species with long generation times, such as forest trees, local adaptations mostly develop from alleles already present in the gene pools (*i.e.*, from standing genetic variation), which often results in the establishment of gene clines along environmental gradients (Savolainen *et al.* 2007; Eckert *et al.* 2010a; Chen *et al.* 2012; Prunier *et al.* 2012; Alberto *et al.* 2013b). Compared to well-studied selfing annuals, forest trees have to face highly variable selection pressures that result from environmental changes over long periods of time. It follows then that the fitness consequences of molecular maladaptation to climate have to be explored in such long-lived outcrossing taxa.

Many characteristics of forest trees, particularly conifers, make them ideal systems for studying climate (mal)adaptation. They are distributed in recurrently shifting geographic ranges, whose changes can be traced back in time through paleobotanical or phylogeographic studies (Petit and Hampe 2006), they exhibit large and significant differences in adaptive phenotypic traits in common garden experiments (Rehfeldt *et al.* 1999), and they bear low levels of proximal linkage disequilibrium (that decays within gene limits for most species; Brown *et al.* 2004; Heuertz *et al.* 2006). This last feature facilitates the identification of polymorphisms associated with adaptive traits (Neale and Savolainen 2004; González-Martínez *et al.* 2007, 2008), although it also implies that a large number of markers is needed to saturate the genome and adequately capture the genomic signals of adaptation. From this point of view, candidate-gene approaches are particularly attractive as they provide direct links with gene functions and allow targeting potential adaptive traits, gene networks, and/or selection drivers without the need for scanning the whole genome (Neale and Savolainen 2004; Neale and Ingvarsson 2008).

Here, we aimed to identify genetic polymorphisms related to climate adaptation in maritime pine (*Pinus pinaster* Ait.), a long-lived outcrossing forest tree, and used a common garden evaluated for survival under extreme (hot and dry) climatic conditions to validate SNP–climate associations and demonstrate the utility of these genetic markers to predict maladaptation to future climates. Maritime pine is an economically important conifer that forms large populations in southwestern Europe, inhabiting both wet-coastal and seasonally dry continental forests, and that exhibits a strong population structure associated with past climate and demographic changes (Bucci *et al.* 2007; Santos-Del-Blanco *et al.* 2012). Once this phylogeographic structure was accounted for to avoid spurious allele-frequency clines generated by historical factors, we

observed that mean and extreme summer and winter temperatures and winter rainfall were important ecological drivers for adaptation in this species. The spatial distribution of the selected alleles suggested that adaptive forces operate on the standing genetic variation at different scales and that they depend on the climatic heterogeneity experienced by each gene pool. Reduced average population fitness was observed in the common garden as the frequency of locally advantageous alleles diminished. This suggested that contrasting levels of future forest decline may occur in distinct maritime pine gene pools confronted with new (*i.e.*, more arid) climates.

Materials and Methods

Sampling, SNP genotyping, and scoring

Needles were collected from 772 individuals distributed in 36 natural populations across the maritime pine range. Total DNA was extracted with the Invisorb DNA Plant HTS 96 kit following the manufacturer's instructions. Samples were genotyped with two oligo pool assays (OPA) that included 2646 and 384 SNPs by using the Illumina Infinium and Illumina VeraCode platforms, respectively. The first assay contained randomly chosen SNPs detected *in silico* from transcriptome sequence data and are expected to represent neutral variants (hereafter, control SNPs; see Chancerel *et al.* 2013 for more details). Briefly, SNPs were identified from >600,000 sequences obtained from a set of 17 cDNA libraries of different tissues without any prior experimental treatment and were integrated into a unigene set previously used for linkage mapping (Chancerel *et al.* 2013).

The second OPA comprised 384 SNPs distributed in 221 candidate genes (Supporting Information, File S2; see also Budde *et al.* 2014) that included drought-stress candidates identified in Mediterranean (*P. pinaster* and *P. halepensis*) and American (*P. taeda*) pines (Eveno *et al.* 2008; Eckert *et al.* 2010a,b; Grivet *et al.* 2011) and genes overexpressed under abiotic stress (Lorenz *et al.* 2011; Perdiguero *et al.* 2013) and/or that have shown some evidence of adaptive value in different conifers (*e.g.*, González-Martínez *et al.* 2007, 2008; Eveno *et al.* 2008; Eckert *et al.* 2010a,b; Grivet *et al.* 2011; Lepoittevin *et al.* 2012; Mosca *et al.* 2012). SNPs were selected after resequencing these genes in a discovery panel that covered the entire natural range of *P. pinaster* (see Chancerel *et al.* 2011; Budde *et al.* 2014).

For both OPAs, loci in Hardy–Weinberg equilibrium, bearing minor allele frequencies (MAF) >5% and call rates >0.8 (average of 0.97) were retained (1745 and 266 control and candidate SNPs, respectively). Scoring was performed with BeadStudio v. 3.2 (Illumina Inc.) after manual adjustment of genotype clusters. To ensure that genotypes were correctly called, three high-quality DNAs included as positive controls in each genotyped plate were doubled checked for all SNPs.

Microsatellite genotyping and phylogeographic analyses

To corroborate that historical and demographic processes were adequately accounted for, 12 nuclear microsatellites (SSRs) distributed in 8 of the 12 linkage groups of the species (Chancerel *et al.* 2011; de-Miguel *et al.* 2012) were genotyped following Santos-Del-Blanco *et al.* (2012) and references therein. However, three of them (*epi6*, *ctg275*, *NZPR544*) were excluded from further analyses after detecting high frequencies of null alleles, which led to significant deviations from Hardy–Weinberg expectations in more than half of the populations (see Table S1). Genetic structure for these markers and for the control SNPs was estimated independently with a principal components analysis (PCA), performed in R v. 3.0.0 (R Development Core Team 2013) and with the Bayesian approach available in Structure v. 2.3.3 (Pritchard *et al.* 2000). Clustering with Structure was determined with an admixture model on correlated allele frequencies and a burn-in of 100,000 steps followed by 1,000,000 iterations. The number of clusters (K) was set from 1 to 15, and 10 runs were performed for each K . Similarity across runs with the same K was calculated with Clumpp (Jakobsson and Rosenberg 2007), and the most plausible number of clusters was determined following Evanno *et al.* (2005). Both the Structure membership coefficients (Q -matrix) and the first six PC scores of each individual for each independent data set were kept as explanatory variables for the environmental association analyses below. The six PC scores were retained in both cases (SSRs and control SNPs) because they each accounted for >5% of the total genetic variation.

Climate data

Summary climate data for the years 1950–2000 were retrieved for 32 variables from Worldclim (Hijmans *et al.* 2005) and a regional climatic model (Gonzalo 2007) for the 12 non-Spanish and the 24 Spanish populations, respectively. Climate variables included monthly mean, highest, and lowest temperatures and mean precipitation (Table S2). Gonzalo's (2007) model was favored for climate data in Spain because it considers a much denser network of meteorological stations than Worldclim, which is known to underperform in this region. Two independent PCAs, again performed in R v. 3.0.0 (R Development Core Team 2013), were used to summarize the climate variation of each population for the summer (June to September) and winter (December to March) seasons. Population scores for the first three PCs of each season (Table S2), which explained >95% of the total variance, were kept for further analyses (see below).

SNP–climate associations

Significant SNP–climate associations were identified by combining three different approaches. First, candidate SNP-allele frequencies were correlated with the climate PCs with multivariate logistic regressions (*mlr*) using independently the control SNP- and the SSR-PCs as covariates to account for historical and demographic processes that could have generated allele-frequency clines in the absence of selection (Grivet

et al. 2011; De-Mita *et al.* 2013). Identical associations were obtained when using the Structure membership coefficients (Q -matrix) as covariates (data not shown). These analyses were performed in R v. 3.0.0 (R Development Core Team 2013) for each separate candidate SNP by using the *glm* function, the Akaike Information Criterion for model selection, and the Benjamini–Hochberg procedure to control for false discovery rate (FDR; Benjamini 2010).

Second, another set of SNP–climate correlations was performed with the Gibbs sampler and the latent factor mixed models available in the software LFMM (Frichot *et al.* 2013). Following the population clustering results above (Structure and PCA), the number of latent factors (k) was set to six, as they should capture most of the underlying population structure (see Results). Then, 10 runs of one million sweeps were performed after discarding 100,000 iterations as burn-in. Significance was assessed by using the same FDR procedure as in the *mlr* analysis. Preliminary runs made with higher values of k yielded identical results, and tests performed with fewer latent factors resulted in correlations with lower statistical support and less convergence across independent runs (data not shown). Only those significant SNP–climate correlations that overlapped between the *mlr* and LFMM approaches (including all LFMM runs) were retained.

Third, a covariance matrix was built based on the control SNP data set (1745 SNPs) and used as a null model to further test for candidate SNP–climate correlations with the geographical Bayesian association analysis in Bayenv 2.0 (Coop *et al.* 2009; Günther and Coop 2013), a method showing great promise in comparative performance evaluations (Lotterhos and Whitlock 2014). To counterbalance the possibility that any gene-derived marker can be potentially involved in adaptation and can therefore be suboptimal to capture the population structure in the covariance matrix, a series of analyses was performed to validate its suitability. First, the covariance matrices produced by three independent Bayenv 2.0 runs were compared to each other to verify convergence to similar results. Second, these matrices were transformed to correlation matrices with the *cov2cor* function in R v. 3.0.0 and correlated to pairwise- F_{ST} matrices generated in Arlequin v. 3.5 (Excoffier and Lischer 2010) from this and the SSR data sets using Pearson's coefficient in R v. 3.0.0. Third, the Q -matrices produced in the Structure clustering analyses for the two putatively neutral data sets (control SNPs and SSRs) were similarly correlated in R v. 3.0.0.

After these validation steps (see File S1), a Bayes factor (BF) describing the deviation between the null demographic model (*i.e.*, the covariance matrix) and the alternative one (the candidate SNP–climate correlations) was estimated for each SNP. Model convergence was assured by performing three independent runs of 100,000 iterations and a burn-in of 10,000 chains with random seeds. Statistical support was assessed by Spearman's rank correlation (ρ) tests for those associations exhibiting unusually high BF (Eckert *et al.* 2010b; De La Torre *et al.* 2014). This was complemented by comparing these unusually high BFs to those observed for

SNPs that did not show any associations with the environment in any of the two previous methods (*mlr* and LFMM; see Figure S1). Only those candidate SNP–climate associations with a BF >10 (corresponding to strong support according to Jeffreys' scale) and a p -value of 0.25 or higher were conserved for validation and fitness prediction in the common garden (see below).

All three sets of analyses were repeated at the regional scale for the Iberian Mediterranean and Atlantic regions, where sampling was more intensive in terms of number of populations and individuals. The Mediterranean region of the Iberian Peninsula is a climatically heterogeneous area characterized by a high seasonality with very dry summers, in which maritime pine forms scattered populations near the coast, along altitudinal gradients in different mountain systems, and in the central plateau (Alía *et al.* 1996). The Atlantic region of the Iberian Peninsula is climatically more homogeneous, with wet-temperate climate and low seasonality. In this region, *P. pinaster* exhibits large continuous populations, partly due to plantations of local origin (Alía *et al.* 1996).

Genetic diversity and spatial structure

Expected and observed heterozygosity (H_E and H_O) and standardized genetic differentiation (G'_{ST}) were calculated with Smogd (Crawford 2010) for each population and gene pool (*i.e.*, Iberian Mediterranean and Atlantic regions) by using separately the SSRs, the candidate SNPs associated with climate, and the control SNPs. The existence of spatial autocorrelation was surveyed for the last two sets of markers by using the spatial structure analysis available in SAM v. 4.0 (Rangel *et al.* 2010). Briefly, a relative Moran I index (I/I_{max}) was determined for each marker by using a Gabriel connectivity matrix with symmetric distance classes. Significance was assessed with 999 permutations and the Bonferroni criterion for multiple testing.

Fitness predictions

Survival data for 19 representative populations were obtained from a common garden established in February 2004 in northeastern Spain (Cálcena; see Figure 1). The climate in this region is arid (average temperature 11.6°, annual rainfall 502 mm, summer rainfall 101 mm, for the period 1975–2008) and characteristic of the drier extreme of maritime pine's climatic breadth, which provides an ideal setting in which to test for differences in fitness due to climate maladaptation as expected under climate change (aridification and warming) in the core natural distribution of the species. The trial was designed as a nested structure of families within populations. Briefly, 1-year-old seedlings from 520 families were planted at a spacing of 2 m × 3 m in an α -lattice incomplete block design with 3 replications of 65 blocks, 8 families per incomplete block, and 4 plants per family plot (total number of 3 × 65 × 8 × 4 = 6240 plants). Survival was estimated as the ratio of seedlings alive per experimental unit when the plants were 5 years old. The climate in the test site during these 5 years was representative of the regional average (average temperature 11.8°, annual rainfall 533 mm, summer rainfall 99 mm; compare with

data for the period 1975–2008 above), with no particular extreme events in any year that might have affected tree survival.

Empirical best linear unbiased predictors (EBLUPs) for survival were computed for each factor (populations and families within population) after arcsine transformation (adequate for percentage data) using a mixed model with Gaussian error distribution that considered both factors as random. Climate data for the test site were used to determine the expected locally selected alleles for each of the 18 SNPs associated with climate (see Results) using the logistic regressions fitted above. The average frequency of locally advantageous alleles at each source population was then calculated and correlated with the EBLUPs for survival using Pearson's (r) and (nonparametric) Kendall's (τ) coefficients and associated significance tests. Then, given that SNPs associated with climate may have heterogeneous allelic effects on fitness, which would render the use of average frequencies from different loci inadequate, single-locus regressions with survival were also obtained for each of the 18 candidate markers to climate adaptation using standard linear regression (*lm* function) in R v. 3.0.0.

To demonstrate that significant correlations were not due to population structure, we first performed a resampling procedure as follows. A series of 1000 random samples of 18 SNPs with frequencies matching those of the retained candidates were obtained from the control SNP data set. Then, correlation coefficients with survival were computed for each 18-SNP sample as described above, and the probability of having equal or higher correlation coefficients than the ones obtained with the retained SNPs was estimated by comparing this value with the distribution of random correlation coefficients. Second, we computed a new PCA on the complete set of candidate SNPs tested for adaptation to climate (266 SNPs) and a new set of correlations with survival were obtained for the 18 SNPs that best explained population subdivision in this PCA.

Results

A total of 1745 control (*i.e.*, putatively neutral) SNPs and 9 nuclear SSRs were retained and reliably genotyped for 772 individuals from 36 maritime pine populations and used to infer background phylogeographic structure (Figure 1). The first six principal components (PCs) accounted for >67 and 61% of the total genetic variation, for the SNP and SSR data sets, respectively, with each individual component explaining at least 5% of the total variance. The most plausible number of genetic groups (following Evanno *et al.* 2005) produced by the Bayesian clustering analyses (Pritchard *et al.* 2000) was equal to six and coincided with the genetic structure identified by the PCs (Figure S2). These groups, hereafter called “gene pools,” coincided with those previously reported from independent chloroplast and nuclear SSR data sets for maritime pine (Bucci *et al.* 2007; Santos-Del-Blanco *et al.* 2012). Maritime pine gene pools were centered in Morocco,

Figure 1 (A) Geographic distribution of the six gene pools obtained from (B) nine nuclear SSRs in 36 natural populations of maritime pine. (C) Genetic partition is also shown for 1745 control (*i.e.*, putatively neutral) SNPs. The shading in A denotes the species natural range, and the star indicates the location of the common garden used to evaluate fitness (Cálçena, Spain). Populations included in the common garden are in boldface italics.

Corsica, the Atlantic coast of France, and the Atlantic and Mediterranean (southeastern and central) regions of the Iberian Peninsula (Figure 1 and Figure S2); they are probably the result of the expansion of as many glacial refugia (Bucci *et al.* 2007; Santos-Del-Blanco *et al.* 2012).

Climate records (32 variables) for each population were reduced to three PCs for the summer (June to September) and winter months (December to March), explaining 96.3 and 95.9% of the total climatic variation for each season, respectively (Table S2 and Figure S3). The three winter PCs

were mainly loaded by the lowest and mean temperatures (PC-Winter1), the mean precipitation (PC-Winter2), and the highest temperature (PC-Winter3), while the summer axes mostly corresponded to the mean precipitation (PC-Summer1), and the lowest (PC-Summer2) and mean (PC-Summer3) temperatures, respectively. Population PC scores were tested for correlation (after correcting for phylogeographic structure) with the genotypes of the 266 candidate SNPs that were retained from the second OPA. At the range-wide scale, the *mlr* revealed 29 significant allele frequency–climate correlations for 24 SNPs, while the LFMM produced 49 significant correlations for 41 SNPs. In total, 21 correlations for 18 SNPs overlapped for both methods, including those for five nonsynonymous polymorphisms; all of these correlations were also significant in the Bayesian environmental association analysis (Coop *et al.* 2009) (Table 1 and Table S3).

The 18 retained SNPs were located in 17 genes involved in growth, synthesis of secondary metabolites, membrane transport, and abiotic stress response, among others (Table 1). Sixteen of them were correlated with axes loaded by temperature (PC-Winter1, PC-Winter3, or PC-Summer2), while only five showed significant correlations with components associated with precipitation (e.g., PC-Winter2 or PC-Summer1; Table 1, Table S3, and Figure 2). Interestingly, four of the amplicons containing these SNPs (*2_1014_02*, *0_4032_02*, *2_3919_01*, and *CL813*; see functional annotations in Table 1) were overexpressed in maritime pine under experimental drought treatments (Perdiguerio *et al.* 2013), including the two examples given in Figure 2.

Average population differentiation (e.g., G'_{ST}) was higher for the candidate SNPs associated with climate than for the control (i.e., putatively neutral) ones or the nuclear SSRs (Tukey's post-hoc test, $P < 0.01$; see Figure 3A). Likewise, H_O and H_E were higher for these potentially adaptive loci when compared to the control markers (Tukey's post-hoc test, $P < 0.005$; Figure 3B), while relative Moran's I indices were also consistently significant for SNPs associated with climate (mean $I/I_{max} = 0.51$). These last values indicated a range-wide spatial autocorrelation for the alleles of these loci, which were more clumped, albeit widely distributed, than those of the control SNPs (mean $I/I_{max} = 0.19$).

Analyses were repeated at the regional scale for the Iberian Atlantic and Mediterranean regions, which represent independent gene pools currently experiencing contrasting climates (Bucci *et al.* 2007; Gonzalo 2007; Santos-Del-Blanco *et al.* 2012; see also Figure 1). In the more arid and climatically heterogeneous Iberian Mediterranean region, four SNPs (*m8*, *m80*, *m657*, and *m1309*) exhibited significant correlations with climate (Figure S4), while no associations were detected across the more humid and climatically homogeneous Iberian Atlantic region. Moreover, the 18 SNPs previously retained in the range-wide analyses displayed higher population differentiation than that of the other SNPs or SSRs within the first gene pool (Tukey's post-hoc test, $P < 0.01$), while in the second one all types of markers showed virtually the same population differentiation (Figure 3,

E and F). Likewise, genetic diversity was higher in the Mediterranean region for the SNPs associated with climate (Tukey's post-hoc test, $P < 0.02$; Figure 3D), while similar genetic diversity levels were found across regions for the control SNPs (H_E of 0.198 and 0.224, respectively). Alleles of the SNPs associated with climate (when variable) were also more clumped in the Mediterranean region than in the Atlantic one (mean $I/I_{max} = 0.46$ vs. 0.16).

Seedling survival after 5 years in a common garden test under the extreme climatic conditions of northeastern Spain was low and varied greatly across the 19 source populations surveyed (Figure 1 and Figure 4). Such a low (and variable) survival was somehow expected, as the common garden was located at the drier extreme of maritime pine's climatic breadth. Climate data for the common garden were used to determine the expected locally selected alleles for the 18 retained and potentially adaptive SNPs using the fitted logistic regressions above. The average frequency of these locally advantageous alleles in the source populations was positively and significantly correlated to survival in the common garden when using both parametric (Pearson's $r = 0.58$, $P = 0.0093$) and nonparametric (Kendall's $\tau = 0.44$, $P = 0.0082$) methods (Figure 4). Single-locus regressions provided additional support for these predictions, with 14 candidate SNPs (out of 18) each explaining $>5\%$ of the genetic variance for survival (Figure 4 and Table S4). Furthermore, only 0.2% ($P = 0.002$) of the 1000 random sets of 18 SNPs resampled from the control SNP data set (matched by allele frequency) had equal or higher correlation coefficients with survival than the one obtained with the SNPs that showed environmental associations. Finally, the 18 SNPs that best explained population subdivision for each of the four PCs explaining $>5\%$ of the variance in a PCA (accounting for $\sim 60\%$ of the total variance) did not produce any significant correlation (Table S5). Altogether, these tests suggest that the correlation between selected markers and climate maladaptation was not caused by chance or population structure.

Discussion

The utility of candidate-gene approaches in long-lived species with large genomes

In this study, we showed that the frequency of locally advantageous alleles at SNPs from carefully selected candidate genes can be used as predictors of climate maladaptation. This approach should be particularly appealing for outcrossing long-lived species like forest trees, for which establishing common gardens is expensive and time consuming. In addition, for taxa with extremely large genomes, such as conifers (Birol *et al.* 2013; Nystedt *et al.* 2013; Neale *et al.* 2014), the use of a candidate gene strategy seems an interesting and feasible cost-efficient alternative to genome-wide selection (e.g., Westbrook *et al.* 2013), for which millions of markers evenly distributed across the genome might be necessary to have a good predictive power (e.g., Grattapaglia and Resende 2011; Desta and Ortiz 2014). In this context, the adequate preselection of candidate

Table 1 Identity, annotation, and summary estimates of genetic diversity, population differentiation, and spatial structure for 18 SNPs exhibiting significant association with climate principal components (climate PCs)

EST contig or amplicon	<i>A. thaliana</i> locus	E-value	Annotation	LG ^a	SNP name (type of change) ^b	Accession number ^c	Climate PC	H ₀	H _E	G' _{ST}	I / I _{max} ^d
0_11649	At2g29550	2.4 E-85	β-Tubulin	12F / 11	m783 (A/G) ^{syn}	836187519	PC-Winter1	0.34	0.33	0.23	0.47**
0_4032_02	At1g29340	2.8 E-24	U-box containing protein (protein kinase)	—/—	m8 (A/T) ^{syn}	325995579	PC-Winter1	0.34	0.33	0.33	0.47**
0_9486_01	NA	NA	Unknown protein	—/—	m1513 (A/C) ^{nonsyn}	836187520	PC-Summer2	0.23	0.25	0.37	0.47**
2_1014_02	At4g36990	2.5 E-07	Heat-stress transcription factor	—/7	m1250 (A/G) ^{NA}	836187522	PC-Winter2	0.41	0.42	0.20	0.40**
2_3919_01	At3g53720	8.1 E-11	Cation/H(+) antiporter	—/—	m80 (A/G) ^{nonsyn}	836187523	PC-Summer2	0.34	0.34	0.30	0.40**
2_4183_01	At1g08960	4.7 E-52	Cation exchanger	—/10	m1309 (A/G) ^{nonsyn}	836187524	PC-Winter1	0.18	0.22	0.52	0.62***
2_9542_01	At5g51690	8.3 E-54	1-Aminocyclopropane-1-carboxylate synthase	—/11	m100 (A/G) ^{syn}	836187525	PC-Winter3	0.30	0.31	0.27	0.51***
4cPt_c	NA	NA	4-Coumarate-CoA ligase	—/7	m973 (A/G) ^{syn}	836187526	PC-Summer2	0.36	0.38	0.25	0.41**
CL813	At5g20830	5.6 E-18	Sucrose-UDP glucosyltransferase	10M/10	m137 (A/G) ^{syn}	836187527	PC-Summer1	0.35	0.36	0.22	0.37**
Cobra2	At3g02210	1.2 E-19	Cobra-like protein	—/—	m607 (A/C) ^{noncod}	325995573	PC-Winter3	0.41	0.41	0.16	0.50***
CT1487	NA	NA	Unknown protein	—/1	m1156 (A/G) ^{noncod}	184956809	PC-Winter1	0.26	0.25	0.18	0.34**
CT2134	At5g48480	2.8 E-18	Glyoxalase I-like protein	7M/—	m1211 (C/G) ^{NA}	184975277	PC-Winter1	0.31	0.34	0.30	0.52***
CT2331	At5g14040	0.0	Mitochondrial phosphate transporter	—/—	m646 (C/G) ^{nonsyn}	184975314	PC-Winter1	0.30	0.32	0.22	0.38**
CT2334	At1g14320	4.9 E-159	60S ribosomal protein	5F	m657 (A/C) ^{nonsyn}	184975344	PC-Winter1	0.42	0.40	0.17	0.46**
CT3538	NA	0.0	Chlorophyll a/b binding protein	3F/3	m658 (A/G) ^{syn}	184975347	PC-Summer2	0.34	0.34	0.31	0.56***
CT384	At2g29960	9.2 E-89	Peptidyl-prolyl cis-trans isomerase	—/—	m685 (A/G) ^{syn}	184975431	PC-Winter2	0.34	0.32	0.22	0.64***
CT4109	At2g40590	2.6 E-30	40S ribosomal protein (predicted)	—/—	m1196 (A/C) ^{syn}	184975463	PC-Summer3	0.43	0.39	0.33	0.67***
					m705 (A/G) ^{noncod}	184975489	PC-Summer2	0.29	0.33	0.27	0.52***

Only those loci displaying significant correlations at $P < 0.001$ and with Bayes factors > 10 are shown (see also Table S3).

^a Linkage groups retrieved from two independent maritime pine mapping populations: Chancerel *et al.* (2011) and de-Miguel *et al.* (2012).

^b Abbreviations: syn, synonymous; nonsyn, nonsynonymous; noncod, noncoding; NA, unknown/data unavailable. The locally selected alleles at the common garden used to evaluate fitness, judging from logistic models, appear in boldface type.

^c Accession numbers as provided by NCBI's dbSNP.

^d Levels of significance: ** $p < 0.01$; *** $p < 0.005$.

m80 (Heat stress transcription factor)

Figure 2 Examples of allele-climate associations for maritime pine populations at the range-wide scale. (A and D) Scatter plots, (B and E) box plots, and (C and F) distribution of minor allele frequencies (MAF) are shown for 2 of the 18 potentially adaptive SNPs. Lines within the scatter plots indicate clines of allele frequencies under a logistic regression model. Boxes denote the interquartile range and horizontal lines within boxes represent genotypic means.

genes becomes a fundamental step toward capturing a large part of the adaptive/phenotypic variance that needs to be used to perform such predictions.

Herein, we performed the preselection of genes by pinpointing candidates from previous population and functional studies. These genes included drought-stress-response candidates detected by population genetics (e.g., Eveno *et al.* 2008; Grivet *et al.* 2011; Mosca *et al.* 2012; see also Budde *et al.* 2014) or expression studies (Lorenz *et al.* 2011; Perdiguero *et al.* 2013), as well as variants previously associated to wood properties, cold hardiness, and growth in maritime pine (Pot *et al.* 2005; Eveno *et al.* 2008; Lepoittevin *et al.* 2012) and other conifers (González-Martínez *et al.* 2007, 2008; Eckert *et al.* 2010a,b). Many of these genes belong to gene families that have been associated with adaptive responses in other plants. For instance, amplicons *CL813* and *CT2134* contain genes directly involved in osmotic adjustments that are related to the metabolism of carbohydrates in both maritime pine and *Arabidopsis* (Seki *et al.* 2002; Pot *et al.* 2005), while contig *O_11649* encodes for a β -tubulin, a family of genes whose expression changes under low temperatures in different plants (Chu *et al.* 1993; Seki *et al.* 2002; Perdiguero *et al.* 2013).

This preselection of genes allowed us to survey many of the potential targets of selection and climate adaptation in

maritime pine, but it may also have led to ascertainment bias and to limiting our power to identify selection drivers in this species (see also Morin *et al.* 2004; Namroud *et al.* 2008). For instance, by selecting candidates from studies not only in maritime pine but also in other conifers, we assumed that adaptive processes mostly occur by convergent evolution in the same set of genes, while there is evidence suggesting that species adaptation to identical environments may involve separate genes and a certain number of possible paths (e.g., Tenaillon *et al.* 2012). For example, only 2 of the 18 genes surveyed in latitudinal clines of two Eurasian boreal spruces showed identical associations with bud phenology (Chen *et al.* 2012, 2014), while only 3 outlier genes related to climate were shared between two sympatric *Picea* species in eastern Canada (Prunier *et al.* 2011). Similar results have been reported for adaptation to freshwater in sticklebacks (Deagle *et al.* 2012) and to high elevation in humans (Bigham *et al.* 2010). Further bias may originate from excluding noncoding regions, micro-RNAs, or copy-number and presence-absence variants (including transposon insertions), which have been directly associated with adaptive responses in many taxa, including conifers (e.g., González *et al.* 2008; Yakovlev *et al.* 2010; Fischer *et al.* 2011; Hanikenne *et al.* 2013). While genome-wide approaches are appropriate to capture most of this variability,

Figure 3 (A) Range-wide and (E and F) regional population differentiation (G'_{ST}) for nuclear SSRs, control SNPs (*i.e.*, putatively neutral), and SNPs associated with climate across natural populations of maritime pine. (B) Overall genetic diversity (H_E) for SNPs associated with climate and control SNPs; (D) genetic diversity (H_E) by SNP type in the different regions. (C) Euclidean distances (d) on climate variability at the regional scale are also shown. Boxes denote the interquartile range, whiskers the 95% confidence intervals, and horizontal lines within boxes represent genotypic (or climatic) means.

having a satisfactory representation of the whole segregating genomic diversity is still a challenge for nonmodel species with large genomes, such as conifers (*e.g.*, Birol *et al.* 2013; Nystedt *et al.* 2013; see also Tiffin and Ross-Ibarra 2014).

Identifying SNP–climate associations and potential selection drivers

Complementary statistical analyses were used herein for detecting genotype–climate associations, which should be adequate to minimize the detection of false positives (De-Mita *et al.* 2013). Additional support for the adaptive role of the retained associations (involving 18 SNPs) was obtained from

a common garden experiment under extreme climate. This strategy (*i.e.*, environmental associations plus common gardens to evaluate fitness) should enrich traditional association studies that provide links between genotype and phenotype (mostly in the form of candidate genes lists), by identifying some of the genotype–environment relationships involved in adaptation (*i.e.*, the selection drivers) and by estimating the fitness effects of ecologically relevant variants (ideally under multiple environments). It can also provide a much needed validation step (Ioannidis *et al.* 2009; König 2011).

Most of the SNPs associated with climate in maritime pine were related to PC axes loaded by temperature, during

Figure 4 Scatter plot of average frequency of locally advantageous alleles in the common garden (Cálcena, Spain) at each source population and EBLUPs (empirical best linear unbiased predictors) for survival, a proxy for fitness. The inset shows the distribution of allelic effects for the 18 SNPs associated with climate.

both the winter and the summer months, and, to a lesser extent, by winter precipitation. Adaptation to these elements are critical for assuring survival in plants (Condit *et al.* 1995), which supports the view that maladaptation to increasing temperature and drought can become an important source of vulnerability for forests under climate change (Carnicer *et al.* 2011; Choat *et al.* 2012). The overrepresentation of correlations with winter temperatures in our analyses (53% of the total; Table 1) might seem counterintuitive for a Mediterranean conifer that should be more affected by drought stress during the summer, according to modeled distributions (Benito-Garzón *et al.* 2011). However, the genetic control of growth cessation and cold hardiness is capital for the survival of many plants, including forest trees (Tanino *et al.* 2010; Cooke *et al.* 2012). Moreover, given that both freezing and drought trigger analogous physiological responses in conifers (Bigras and Colombo 2001; Blödner *et al.* 2005), the same genes are expected to control (at least partially) responses to both types of stress (Urano *et al.* 2010). Indeed, four of the amplicons whose SNPs were related to winter temperatures were also overexpressed under drought conditions in maritime pine (Perdiguerro *et al.* 2013), while four of these SNPs were also associated with summer PCs, including a nonsynonymous variant located on a gene coding for a putative heat-stress transcription factor (*m80*; Table 1 and Figure 2). This is consistent with results from *Arabidopsis thaliana*, where 30 genes were upregulated by both cold temperatures and drought, including four amplicons coding for heat-shock proteins (Seki *et al.* 2002). Other genes associated with environment that are worth highlighting include the above-mentioned *CL813*, *CT2134*, and *O_11649*, which were correlated with winter PCs (Table 1; Seki *et al.* 2002; Pot *et al.* 2005), and *4cl-Pt_c*. This last gene belongs to the *4cl* family, which is involved in the phenylpropanoid metabolism and the biosynthesis of lignin (Yun *et al.* 2005; González-Martínez *et al.* 2007; Wagner *et al.* 2009) and has previously exhibited environmental associations at the haplotype level in maritime pine and other Mediterranean pine species (Grivet *et al.* 2011).

It must be noted, however, that although promising, SNP–environment correlations alone only hint at the real drivers of selection and local adaptation. For instance, in maritime pine, any other factor covarying with extreme temperatures and low

precipitation should exhibit similar correlations with the retained candidate SNPs. Indeed, it can be argued that these correlations might be biased if the used climate variables are spatially structured or partially match the genetic structure of the species. For example, two of the clusters detected for maritime pine visually coincide with the regions of highest average precipitation (northern Spain and Atlantic France), while the southeastern Spain gene pool is mostly located in a region with overly high winter temperatures. Indeed, when nonlinear models were fitted to explain each climatic PC with either geography (latitude and longitude) or genetic structure (the PCs derived from the SSR and control SNP data sets), significant correlations were observed for PC-Winter3 ($r^2 = 0.85$ for geography and $r^2 = 0.6$ for the genetic PCs) and PC-Summer2 ($r^2 = 0.39$ for geography and $r^2 = 0.7$ for the genetic PCs; Table S6), two components loaded by extreme temperatures (Table S2). Although these results suggest that caution should be taken when interpreting genetic associations with these two particular PCs, most of the correlations detected herein (67%) involved climate components that showed no evidence of being spatially structured (Table 1 and Table S6), which lends support to the view that the selected SNPs are effectively associated with adaptation to climate.

Geographic extent of SNP–climate associations and modes of adaptation

Alleles from potentially adaptive SNPs were both widely distributed and locally clumped, while the control (*i.e.*, putatively neutral) ones showed a strong population structure at the range-wide scale and a lower aggregation at the local level (Table 1, Figure 1, and Figure 2). Such patterns, together with the particular biological traits of forest trees, suggest that selection could be acting on the standing genetic variation of maritime pine and at relatively small geographical scales (see Eckert *et al.* 2012 for a similar case in *P. contorta*). Forest trees are characterized by large population sizes, extensive gene flow, and long generation times, which favor the long-term maintenance of ancestral polymorphisms (Bouillé and Bousquet 2005; Petit and Hampe 2006); indeed, selection on standing genetic variation seems to be the rule in these taxa (Savolainen *et al.* 2011; Chen *et al.* 2012; Eckert *et al.* 2012; Prunier *et al.* 2012; Alberto *et al.* 2013a,b). Such attributes

would also allow tree populations to better respond to heterogeneous local selection (Kremer *et al.* 2012), leading to contrasted regional adaptive patterns, such as observed herein for the Atlantic and Mediterranean regions of the Iberian Peninsula. The genetic differentiation and diversity of the potentially adaptive SNPs reflect the environmental variability of these areas (Figure 3C), which supports the view that there should be a correlation between the within-population genetic variation at adaptive traits and the environmental heterogeneity at the regional scale (Yeaman and Jarvis 2006; Kremer *et al.* 2012). However, it must be noted that most of the statistical approaches currently used to identify loci related to local adaptation (including those employed herein) assume that their alleles exhibit antagonistic pleiotropy (*e.g.*, De Mita *et al.* 2013; see review in Tiffin and Ross-Ibarra 2014), that is, that their variants are beneficial in particular regions of the species range and deleterious elsewhere. Nevertheless, empirical studies suggest that many alleles contributing to local adaptation in particular environments are effectively neutral in other regions (*e.g.*, Anderson *et al.* 2011; Fournier-Level *et al.* 2011). Such conditional neutrality could be an alternative explanation for the differences observed herein between the Iberian Atlantic and Mediterranean regions.

Other than the action of selection on standing genetic variation, it can be argued that selection on newly arisen variants could have also played a role in modeling the patterns described above. When a new advantageous mutation appears and hard selective forces drive it to high frequency within a few generations, its geographic distribution is expected to be narrower than that of the alleles already present in the species gene pool (Hancock *et al.* 2011). Nonetheless, given enough time and relatively stable populations, extensive gene flow, such as the one of conifers, can rapidly spread this variant until matching the patterns of more ancient alleles (Kremer *et al.* 2012). Thus, to address the contribution of *de novo* adaptive variants, the time scale and intensity of the selective forces that are currently operating in maritime pine should be estimated. Such estimations are out of the scope of this study, but taking into account the selection intensities previously inferred for putatively adaptive SNPs in other conifers (*i.e.*, $s = 0.01 - 0.04$; Prunier *et al.* 2011; Eckert *et al.* 2012), an adaptive allele that arose just after the last glacial maximum (some 10,000 YBP) should currently have a frequency <0.03 , assuming an average generation time of 50 years. These rough calculations imply that, because of the use of markers with MAF >0.05 , our strategy should be adequate to account only for far more ancient mutations or for those that have been favored by much stronger selection coefficients. Such a case was recently reported for Norway spruce, where a new variant in the *PaFTL2* gene appeared and spread across most of Fennoscandia in <6500 years, until reaching frequencies >0.6 in some modern populations in northern Sweden (Chen *et al.* 2012). However, even in this extreme case, the new advantageous allele did still have a local distribution (*i.e.*, only Fennoscandia), which is something that was

not observed herein for the SNPs associated with climate in maritime pine.

Spatially heterogeneous patterns of adaptation can also be accounted for by demographic history (*i.e.*, the fixation of lineage-specific polymorphisms by drift), as proposed for boreal black spruce (Prunier *et al.* 2012). However, in contrast to this conifer, overall levels of neutral genetic diversity in Mediterranean maritime pine were not different across regions (Figure 3D), thus suggesting similar local demographic histories, likely modeled by past bottlenecks and local expansions (Grivet *et al.* 2011). Furthermore, despite boreal and Mediterranean conifers bearing equally high levels of pollen-driven gene flow (O'Connell *et al.* 2006; de-Lucas *et al.* 2008), they have differential capacities (higher in boreal conifers) for quick range-shifts following environmental changes (McLeod and MacDonald 1997; Rubiales *et al.* 2010; Santos-Del-Blanco *et al.* 2012). Thus, it would not be surprising to observe selection to play contrasting roles on adaptive alleles in these species (Kuparinen *et al.* 2010). Several theoretical and empirical works (reviewed by Kremer *et al.* 2012) have shown that fixation of adaptive polymorphisms can be facilitated under rapid dispersal scenarios, such as for boreal conifers (McLeod and MacDonald 1997; Savolainen *et al.* 2011; Chen *et al.* 2012; Prunier *et al.* 2012), while more local expansions, such as those of Mediterranean taxa (Bucci *et al.* 2007; Rubiales *et al.* 2010; Grivet *et al.* 2011), should result in allele-frequency divergence for adaptive variants, particularly under contrasting environments and by assuming either an antagonistic-pleiotropic or a conditionally neutral model (see above).

Species evolving under this local-expansion scenario should also be more sensitive to maladaptation when facing rapid environmental changes, as suggested herein. Indeed, when grown in a common garden at the drier extreme of their climatic breadth, maritime pine populations bearing low frequency of locally advantageous alleles at potentially adaptive SNPs showed increased mortality (*i.e.*, lower fitness), suggesting that different gene pools should exhibit contrasting responses to climate change (see also Benito-Garzón *et al.* 2011). Strong selective regimes, such as the one observed in our common garden, can also foster adaptation and ultimately result in population persistence when acting concomitantly with the high reproductive capacity and population growth of forest trees (Lynch and Lande 1993; Kuparinen *et al.* 2010; Savolainen *et al.* 2011). However, it is uncertain whether a species such as maritime pine, which has a fragmented distribution and is already at its ecological limit in large parts of its southern range, can cope with environments that will quickly become more arid in the near future (Loarie *et al.* 2009; Bellard *et al.* 2012; Alberto *et al.* 2013a).

Integrating molecular predictors into range-shift models

Incorporating fitness proxies to range-shift models based on occurrence data can substantially change their predictions (*e.g.*, Kearney *et al.* 2009; Benito-Garzón *et al.* 2011; see also Hoffmann and Sgrò 2011). The results of the present study suggest that population genomic information (*e.g.*, the frequency

of beneficial alleles for a particular site identified through environmental association) could be a good proxy for fitness and to be integrated into range-shift models. Moreover, breeding predictors could be estimated from adaptive marker genotypes for each particular population/gene pool under different climate change scenarios, and used to guide future reforestation programs. A marker-based approach might also simplify the prediction of evolutionary trajectories by considering the underlying changes in allele frequency of locally selected alleles and their covariance. Within this framework, new markers from genome-wide studies could be readily incorporated, thus allowing improved predictive models. The sequencing of the first conifer reference genomes is a promising step in this direction (Birol *et al.* 2013; Nystedt *et al.* 2013; Neale *et al.* 2014).

The use of haplotypes can further improve the accuracy of predictive models that rely solely on the marginal effects of individual loci (Calus *et al.* 2008; Eveno *et al.* 2008; Grivet *et al.* 2011), as they allow incorporating epistatic interactions of alleles (*i.e.*, genetic context), which also affects fitness. This is of particular interest given that beneficial allele combinations and common large-effect variants are the first to be captured after the establishment of new divergent selective forces, as may be the case for impending climate change, while rarer variants and small-effect alleles at individual loci are targeted only after >50 generations (Kremer and Le Corre 2012). In the absence of adequate LD estimates, genetic context might be approximated by using allele frequency covariation, but in species with strong population genetic structure, such as maritime pine, the utility of this approach is limited. In such cases, looking for an enrichment of coupling-phase LD among adaptive alleles might be an alternative, although newly developed approaches may provide the adequate framework to overcome these issues more easily (*e.g.*, Berg and Coop 2014). Finally, the incorporation of epigenetic variation into these predictive models must also be considered. Epigenetic factors have been shown to drive a substantial proportion of phenotypic variation and climate adaptation in Norway spruce (Yakovlev *et al.* 2010), which seems particularly rich in gene families involved in DNA and chromatin methylation (Nystedt *et al.* 2013). Thus, their role in modulating the expression of key adaptive genes in this and other taxa, including the candidates retained herein, still must be surveyed.

In conclusion, we have shown that, while new technology that allows in-depth studies of the huge conifer genomes is developed, carefully selected candidate genes can still be useful in identifying genetic variation underlying adaptation to climate. Adaptive patterns are expected to vary across geographically separate gene pools, as observed for the Iberian Mediterranean and Atlantic ranges of maritime pine. Thus, the success of programs to preserve biological diversity under impending climate change will largely depend on our capacity to identify and understand how adaptive variation in keystone species is distributed and evolves. Fitness experiments under extreme environmental conditions, as the one developed herein, do not only provide much-needed validation to association and outlier-locus studies but are also a first step

to integrate this knowledge into ecological models to foretell the fate of modern populations and species.

Acknowledgments

We thank T. Kawecki for discussing analytical approaches to predict maladaptation based on SNPs using common gardens under extreme environmental conditions. Thanks are extended to J. Majada, M. Zabal-Aguirre, Y. Kurt, C. García-Barriga, A. I. de-Lucas, C. Martínez-Alonso, C. Feito, and E. Hidalgo for field and laboratory assistance, J. Wegrzyn, T. Lang, and J.-M. Frigerio for bioinformatics support, and S. I. Wright and two anonymous reviewers for thoughtful comments on a previous version of this manuscript. CEGEN-ISCIII (<http://www.cegen.org>) provided SNP genotyping services. We acknowledge grants from the European Commission (FP6 NoE EvolTree and FP7 NovelTree Breeding), the Spanish National Research Plan (ClonaPin, RTA2010-00120-C02-01; VaMPiro, CGL2008-05289-C02-01/02; AdapCon, CGL2011-30182-C02-01; and AFFLORA, CGL2012-40129-C02-02), and the European Research Area-Net BiodivERsA (LinkTree project, EUI2008-03713), which included the Spanish Ministry of Economy and Competitiveness as national funder (part of the 2008 BiodivERsA call for research proposals). G.G.V. was funded by a grant from the Italian Ministry (Ministero dell'Istruzione, dell'Università e della Ricerca project Biodiversitalia, RBAP10A2T4), while J.P.J.-C., D.G., and M.H. were supported by postdoctoral fellowships (Juan de la Cierva and Ramón y Cajal) from the Spanish Ministry of Science and Innovation. Finally, S.C.G.-M. and M.H. acknowledge the support of Marie Curie Intra European Fellowships within the 7th European Community Framework Programme (FP7-PEOPLE-2012-IEF, project nos. 328146 and 329088, respectively).

Literature Cited

- Aitken, S. N., S. Yeaman, J. A. Holliday, T. Wang, and S. Curtis-McLane, 2008 Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* 1: 95–111.
- Alberto, F. J., S. N. Aitken, R. Alía, S. C. González-Martínez, H. Hänninen *et al.*, 2013a Potential for evolutionary responses to climate change: evidence from tree populations. *Glob. Change Biol.* 19: 1645–1661.
- Alberto, F. J., J. Derory, C. Boury, J. M. Frigerio, N. E. Zimmermann *et al.*, 2013b Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* 195: 495–512.
- Alía, R., S. Martín, J. de-Miguel, R. Galera, D. Agúndez *et al.*, 1996 *Las regiones de procedencia de Pinus pinaster Aiton*. Organismo Autónomo de Parques Nacionales, Madrid.
- Anderson, J. T., J. H. Willis, and T. Mitchell-Olds, 2011 Evolutionary genetics of plant adaptation. *Trends Genet.* 27: 258–266.
- Bellard, C., C. Bertelsmeier, P. Leadley, W. Thuiller, and F. Courchamp, 2012 Impacts of climate change on the future of biodiversity. *Ecol. Lett.* 15: 365–377.
- Benito Garzón, M., R. Alía, T. M. Robson, and M. A. Zavala, 2011 Intra-specific variability and plasticity influence potential tree species distributions under climate change. *Glob. Ecol. Biogeogr.* 20: 766–778.

- Benjamini, Y., 2010 Discovering the false discovery rate. *J. R. Stat. Soc. B* 72: 405–416.
- Berg, J. J., and G. Coop, 2014 A population genetic signal of polygenic adaptation. *PLoS Genet.* 10: e1004412.
- Bigham, A., M. Bauchet, D. Pinto, X. Mao, J. M. Akey *et al.*, 2010 Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* 6: e1001116.
- Bigras, F., and S. J. Colombo, 2001 *Conifer Cold Hardiness*. Kluwer, Dordrecht, The Netherlands.
- Biol, I., A. Raymond, S. D. Jackman, S. Pleasance, R. Coope *et al.*, 2013 Assembling the 20Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 29: 1492–1497.
- Blödner, C., T. Skroppa, Ø. Johnsen, and A. Polle, 2005 Freezing tolerance in two Norway spruce (*Picea abies* [L.] Karst.) progenies is physiologically correlated with drought. *J. Plant Physiol.* 162: 549–558.
- Bouillé, M., and J. Bousquet, 2005 Trans-species shared polymorphisms at orthologous gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. *Am. J. Bot.* 92: 63–73.
- Brown, G. R., G. P. Gill, R. J. Kuntz, C. H. Langley, and D. B. Neale, 2004 Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc. Natl. Acad. Sci. USA* 101: 15255–15260.
- Bucci, G., S. C. González-Martínez, G. Le Provost, C. Plomion, M. M. Ribeiro *et al.*, 2007 Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Mol. Ecol.* 16: 2137–2153.
- Budde, K. B., M. Heuertz, A. Hernández-Serrano, J. G. Pausas, G. G. Vendramin *et al.*, 2014 *In situ* genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). *New Phytol.* 201: 230–241.
- Calus, M. P. L., T. H. E. Meuwissen, A. P. W. de Roos, and R. F. Veerkamo, 2008 Accuracy of genomic selection using different methods to define haplotypes. *Genetics* 178: 553–561.
- Carnicer, J., M. Coll, M. Ninyerola, X. Pons, G. Sánchez *et al.*, 2011 Widespread crown condition decline, food web disruption, and amplified tree mortality with increased climate change-type drought. *Proc. Natl. Acad. Sci. USA* 108: 1474–1478.
- Chancerel, E., C. Lepoittevin, G. Le Provost, Y. C. Lin, J. P. Jaramillo-Correa *et al.*, 2011 Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine. *BMC Genomics* 12: 368.
- Chancerel, E., J. B. Lamy, I. Lesur, C. Noirot, C. Klopp *et al.*, 2013 High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biol.* 11: 50.
- Chen, J., T. Källman, X. Ma, N. Gyllenstrand, G. Zaina *et al.*, 2012 Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics* 191: 865–881.
- Chen, J., Y. Tsuda, M. Stocks, T. Källman, N. Xu *et al.*, 2014 Clinal variation at phenology-related genes in spruce: Parallel evolution in *FTL2* and *Gigantea*? *Genetics* 197: 1025–1038.
- Choat, B., S. Jansen, T. J. Brodribb, H. Cochard, S. Delzon *et al.*, 2012 Global convergence in the vulnerability of forests to drought. *Nature* 491: 752–755.
- Chu, B., D. P. Snustad, and J. V. Carter, 1993 Alteration of β -tubulin gene expression during low-temperature exposure in leaves of *Arabidopsis thaliana*. *Plant Physiol.* 103: 371–377.
- Condit, R., S. P. Hubbell, and R. B. Foster, 1995 Mortality rates of 205 neotropical tree and shrub species and the impact of a severe drought. *Ecol. Monogr.* 65: 419–439.
- Cooke, J. E. K., M. E. Eriksson, and O. Junttila, 2012 The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. *Plant Cell Environ.* 35: 1707–1728.
- Coop, G., J. K. Pickrell, J. Novembre, S. Kudaravalli, J. Li *et al.*, 2009 The role of geography in human adaptation. *PLoS Genet.* 5: e1000500.
- Crawford, N. G., 2010 SMOGD: software for the measurement of genetic diversity. *Mol. Ecol. Res.* 10: 556–557.
- De La Torre, A. R., D. R. Roberts, and S. N. Aitken, 2014 Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Mol. Ecol.* 23: 2046–2059.
- de-Lucas, A. I., J. J. Robledo-Arnuncio, E. Hidalgo, and S. C. González-Martínez, 2008 Mating system and pollen gene flow in Mediterranean maritime pine. *Heredity* 100: 390–399.
- de-Miguel, M., N. de-Maria, M. A. Guevara, L. Díaz, E. Sáez-Laguna *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.
- De-Mita, S., A. C. Thuillet, L. Gay, N. Ahmadi, S. Manel *et al.*, 2013 Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.* 22: 1383–1399.
- Deagle, B. E., F. C. Jones, Y. F. Chan, D. M. Absher, D. M. Kingsley *et al.*, 2012 Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc. R. Soc. B Biol. Sci.* 279: 1277–1286.
- Desta, Z. A., and R. Ortiz, 2014 Genome-selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19: 592–601.
- Eckert, A. J., H. Shahi, S. L. Datwyler, and D. B. Neale, 2012 Spatially variable natural selection and the divergence between parapatric subspecies of lodgepole pine (*Pinus contorta*, Pinaceae). *Am. J. Bot.* 99: 1323–1334.
- Eckert, A. J., J. van Heerwaarden, J. L. Wegrzyn, C. D. Nelson, J. Ross-Ibarra *et al.*, 2010a Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185: 969–982.
- Eckert, A. J., A. D. Bower, S. C. González-Martínez, J. L. Wegrzyn, G. Coop *et al.*, 2010b Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Mol. Ecol.* 19: 3789–3805.
- Evanno, G., S. Regnaut, and J. Goudet, 2005 Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14: 2611–2620.
- Eveno, E., C. Collada, M. A. Guevara, V. Léger, A. Soto *et al.*, 2008 Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Mol. Biol. Evol.* 25: 417–437.
- Excoffier, L., and H. E. L. Lischer, 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10: 564–567.
- Fischer, I., L. Camus-Kulandaivelu, F. Allal, and W. Stephan, 2011 Adaptation to drought in two wild tomato species: the evolution of the *Asr* gene family. *New Phytol.* 190: 1032–1044.
- Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt *et al.*, 2011 A map of local adaptation in *Arabidopsis thaliana*. *Science* 334: 86–89.
- Franks, S. J., and A. A. Hoffmann, 2012 Genetics of climate change adaptation. *Annu. Rev. Genet.* 46: 185–208.
- Frichot, E., S. D. Schoville, G. Bouchard, and O. François, 2013 Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* 30: 1687–1699.
- Gamache, I., and S. Payette, 2004 Height growth response of tree line black spruce to recent climate warming across the forest-tundra of eastern Canada. *J. Ecol.* 92: 835–845.
- González, J., K. Lenkov, M. Lipatov, J. M. MacPherson, and D. A. Petrov, 2008 High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol.* 6: e251.
- González-Martínez, S. C., N. C. Wheeler, E. Ersoz, C. D. Nelson, and D. B. Neale, 2007 Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175: 399–409.

- González-Martínez, S. C., D. Huber, E. Ersoz, J. M. Davis, and D. B. Neale, 2008 Association genetics in *Pinus taeda* L. II. Water use efficiency. *Heredity* 101: 19–26.
- Gonzalo, J., 2007 Phytoclimatic analysis of the Spanish Peninsula: update and geostatistical analysis. PhD Thesis, University of Valladolid, Palencia, Spain.
- Grattapaglia, D., and M. D. V. Resende, 2011 Genome selection in forest tree breeding. *Tree Genet. Genomes* 7: 241–255.
- Grivet, D., F. Sebastiani, R. Alía, and T. Bataillon, S. Torre *et al.*, 2011 Molecular footprints of local adaptation in two Mediterranean conifers. *Mol. Biol. Evol.* 28: 101–116.
- Günther, T., and G. Coop, 2013 Robust identification of local adaptation from allele frequencies. *Genetics* 195: 205–220.
- Hancock, A. M., B. Brachi, N. Faure, M. W. Horton, L. B. Jarymowycz *et al.*, 2011 Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334: 83–86.
- Hanikenne, M., J. Kroymann, A. Trampczynska, M. Bernal, P. Motte *et al.*, 2013 Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS Genet.* 9: e1003707.
- Heuertz, M., E. De Paoli, T. Källman, H. Larsson, I. Jurman *et al.*, 2006 Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce. [*Picea abies* (L.) Karst] *Genetics* 174: 2095–2105.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005 Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25: 1965–1978.
- Hoffmann, A. A., and C. M. Sgrò, 2011 Climate change and evolutionary adaptation. *Nature* 470: 479–485.
- Hughes, L., 2000 Biological consequences of global warming: Is the signal already apparent? *Trends Ecol. Evol.* 15: 56–61.
- Ioannidis, J. P. A., G. Thomas, and M. J. Daly, 2009 Validating, augmenting and refining genome-wide association signals. *Nat. Rev. Genet.* 10: 318–329.
- Jakobsson, M., and N. A. Rosenberg, 2007 CLUMPP: a software matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806.
- Kearney, M., W. P. Porter, C. Williams, S. Ritchie, and A. A. Hoffmann, 2009 Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Funct. Ecol.* 23: 528–538.
- König, I. K., 2011 Validation in genetic association studies. *Brief. Bioinform.* 12: 253–258.
- Kremer, A., and V. Le Corre, 2012 Decoupling of differentiation between traits and their underlying genes in response divergent selection. *Heredity* 108: 375–385.
- Kremer, A., O. Ronce, J. J. Robledo-Arnuncio, F. Guillaume, G. Bohrer *et al.*, 2012 Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol. Lett.* 15: 378–392.
- Kuparinen, A., O. Savolainen, and F. M. Schurr, 2010 Increased mortality can promote evolutionary adaptation of forest trees to climate change. *For. Ecol. Manage.* 259: 1003–1008.
- Lepointevin, C., L. Harvengt, C. Plomion, and P. H. Garnier-Géré, 2012 Association mapping for growth, straightness and wood chemistry traits in the *Pinus pinaster* Aquitaine breeding population. *Tree Genet. Genomes* 8: 113–126.
- Loarie, S. R., P. B. Duffy, H. Hamilton, G. P. Asner, C. B. Field *et al.*, 2009 The velocity of climate change. *Nature* 462: 1052–1055.
- Lorenz, W. W., R. Alba, Y. S. Yu, J. M. Bordeaux, M. Simões *et al.*, 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.
- Lotterhos, K. E., and M. C. Whitlock, 2014 Evaluation of demographic history and neutral parameterization on the performance of F_{st} outlier tests. *Mol. Ecol.* 23: 2178–2192.
- Lynch, M., and R. Lande, 1993 Evolution and extinction in response to environmental change, pp. 234–250 in *Biotic Interactions and Global Change*, edited by P. M. Kareiva, J. G. Kingsolver, and R. B. Huey. Sinauer Associates, Sunderland, MA.
- Malcom, J. R., A. Markham, R. P. Neilson, and M. Garaci, 2011 Estimated migration rates under scenarios of global climate change. *J. Biogeogr.* 29: 835–849.
- McLachlan, J. S., J. S. Clark, and P. S. Manos, 2005 Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86: 2088–2098.
- McLeod, T. K., and G. MacDonald, 1997 Postglacial range expansion and population growth of *Picea mariana*, *P. glauca* and *Pinus banksiana* in the western interior of Canada. *J. Biogeogr.* 24: 865–881.
- Morin, P. A., G. Luikart, and R. K. Wayne, and the SNP workshop Group, 2004 SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19: 208–216.
- Mosca, E., A. J. Eckert, E. A. Di Pierro, D. Rocchini, N. La Porta *et al.*, 2012 The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Mol. Ecol.* 21: 5530–5545.
- Namroud, M. C., J. Beaulieu, N. Juge, J. Laroche, and J. Bousquet, 2008 Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* 17: 3599–3613.
- Neale, D. B., and P. K. Ingvarsson, 2008 Population, quantitative and comparative genomics of adaptation in forest trees. *Curr. Opin. Plant Biol.* 11: 149–155.
- Neale, D. B., and O. Savolainen, 2004 Association genetics of complex traits in conifers. *Trends Plant Sci.* 9: 325–330.
- Neale, D. B., J. L. Wegrzyn, K. A. Stevens, A. V. Zimin, D. Puiu *et al.*, 2014 Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* 15: R59.
- Nystedt, B., N. Street, A. Wetterbom, A. Zuccolo, Y. C. Lin *et al.*, 2013 The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584.
- O'Connell, L. M., A. Mosseler, and O. P. Rajora, 2006 Impact of forest fragmentation on the mating system and genetic diversity of white spruce (*Picea glauca*) at the landscape level. *Heredity* 97: 418–426.
- Perdiguer, P., M. C. Barbero, M. T. Cervera, C. Collada, and A. Soto, 2013 Molecular responses to water stress in two contrasting Mediterranean pines (*Pinus pinaster* and *Pinus pinea*). *Plant Physiol. Biochem.* 67: 199–208.
- Petit, R. J., and A. Hampe, 2006 Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.* 37: 187–214.
- Pot, D., L. McMillan, C. Echt, G. Le Provost, P. H. Garnier-Géré *et al.*, 2005 Nucleotide variation in genes involved in wood formation in two pine species. *New Phytol.* 167: 101–112.
- Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Prunier, J., J. Laroche, J. Beaulieu, and J. Bousquet, 2011 Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Mol. Ecol.* 20: 1702–1716.
- Prunier, J., S. Gerardi, J. Beaulieu, and J. Bousquet, 2012 Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Mol. Ecol.* 21: 4270–4286.
- R Development Core Team, 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org>.
- Rangel, T. F., J. A. F. Diniz-Filho, and L. M. Bini, 2010 SAM: a comprehensive application of spatial analysis in macroecology. *Ecography* 33: 46–50.
- Rehfeldt, G. E., C. C. Ying, D. L. Spittlehouse, and D. A. Hamilton, 1999 Genetic responses to climate in *Pinus contorta*: niche

- breadth, climate change, and reforestation. *Ecol. Monogr.* 69: 375–407.
- Rubiales, J. M., I. García-Amorena, L. Hernández, M. Génova, F. Martínez *et al.*, 2010 Late quaternary dynamics of pinewoods in the Iberian Mountains. *Rev. Palaeobot. Palynol.* 162: 476–491.
- Santos-del-Blanco, L., J. Climent, S. C. González-Martínez, and J. R. Pannell, 2012 Genetic differentiation for size at first reproduction through male vs. female functions in the widespread Mediterranean tree *Pinus pinaster*. *Ann. Bot.* 110: 1449–1460.
- Savolainen, O., T. Pyhäjärvi, and T. Knürr, 2007 Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* 38: 595–619.
- Savolainen, O., S. T. Kujala, C. Sokol, T. Pyhäjärvi, K. Avia *et al.*, 2011 Adaptive potential of northernmost tree populations to climate change, with emphasis on Scots pine (*Pinus sylvestris* L.). *J. Hered.* 102: 526–536.
- Seki, M., M. Narusaka, J. Ishida, T. Nanjo, M. Fujita *et al.*, 2002 Monitoring the expression profile of 7000 Arabidopsis genes under drought, cold and high salinity stresses using a full-length cDNA microarray. *Plant J.* 31: 279–292.
- Tanino, K. K., L. Kalcsits, S. Silim, E. Kendall, and G. R. Gray, 2010 Temperature-driven plasticity in growth cessation and dormancy development in deciduous woody plants: a working hypothesis suggesting how molecular and cellular function is affected by temperature during dormancy induction. *Plant Mol. Biol.* 73: 49–65.
- Tenaillon, O., A. Rodríguez-Vergugo, R. L. Gaut, P. McDonald, A. F. Bennett *et al.*, 2012 The molecular diversity of adaptive convergence. *Science* 335: 457–461.
- Tiffin, P., and J. Ross-Ibarra, 2014 Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* 29: 673–680.
- Urano, K., Y. Kurihara, M. Seki, and K. Shinozaki, 2010 “Omics” analyses of regulatory networks in plant abiotic stress responses. *Curr. Opin. Plant Biol.* 13: 132–138.
- Wagner, A., L. Donaldson, H. Kim, L. Phillips, H. Flint *et al.*, 2009 Suppression of 4-Coumarate-CoA ligase in the coniferous gymnosperm *Pinus radiata*. *Plant Physiol.* 149: 370–383.
- Westbrook, J. W., M. F. Resende, Jr., P. Munoz, A. R. Walker, J. L. Wegrzyn *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 Phytol.* 199: 89–100.
- Yakovlev, I. A., C. G. Fossdal, and Ø. Johnsen, 2010 MicroRNAs, the epigenetic memory and climate adaptation in Norway spruce. *New Phytol.* 187: 1154–1169.
- Yeaman, S., and A. Jarvis, 2006 Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. *Proc. R. Soc. B Biol. Sci.* 273: 1587–1593.
- Yun, M. S., W. J. Chen, F. Deng, and Y. Yogo, 2005 Selective growth suppression of five annual plant species by chalcone and naringenin correlates with the total amount of 4-coumarate: coenzyme A ligase. *Weed Biol. Manage.* 9: 27–37.
- Zhu, K., C. W. Woodall, and J. S. Clark, 2011 Failure to migrate: lack of tree range expansion in response to climate change. *Glob. Change Biol.* 18: 1042–1052.

Communicating editor: S. I. Wright

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.173252/-/DC1>

Molecular Proxies for Climate Maladaptation in a Long-Lived Tree (*Pinus pinaster* Aiton, Pinaceae)

Juan-Pablo Jaramillo-Correa, Isabel Rodríguez-Quilón, Delphine Grivet, Camille Lepoittevin,
Federico Sebastiani, Myriam Heuertz, Pauline H. Garnier-Géré, Ricardo Alía,
Christophe Plomion, Giovanni G. Vendramin, and Santiago C. González-Martínez

Supporting Information

Molecular Proxies for Climate Maladaptation in a Long-Lived Tree

(*Pinus pinaster* Aiton, Pinaceae)

Juan P. Jaramillo-Correa, Isabel Rodríguez-Quilón, Delphine Grivet, Camille Lepoittevin, Federico Sebastiani, Myriam Heuertz, Pauline Garnier-Géré, Ricardo Alía, Christophe Plomion, Giovanni G. Vendramin, Santiago C. González-Martínez*

*To whom correspondence should be addressed. E-mail: santiago@inia.es

Figure S1 Bayes Factors (BFs) describing the deviation of the genotype-environment associations of 266 putative candidate SNPs from a null demographic model (i.e. a co-variance matrix built with 1,745 control SNPs) obtained from BAYENV 2.0. Red dots are those SNPs above the BF=10 threshold (horizontal discontinuous line), and which were assumed to be significantly associated with climate.

Figure S2 Scatterplot of the population scores for the first three principal components derived from the variation of nine nuclear SSRs on 36 natural stands of maritime pine.

Figure S3 Scatterplot of the population scores for the first three principal components derived from the (A) winter and (B) summer climate variation of 36 natural stands of maritime pine.

Figure S4 (continued in next page)

Figure S4 Allele-climate associations obtained for maritime pine populations in Mediterranean Spain. (A, C, E, G, I) Scatterplots and (B, D, F, H, J) box-plots are shown for all SNPs exhibiting significant associations (see main text). Lines within scatterplots indicate clines of allele frequencies under a logistic regression model. Boxes denote the interquartile range and horizontal lines within boxes represent genotypic means.

File S1

Performance of the covariance matrix in null models of Bayesian environmental association using BAYENV 2.0

When performing Bayesian association analyses, it is important to verify that the covariance matrix captures adequately the underlying population structure of the taxon of interest (Coop *et al.* 2009; Günther and Coop 2013). Herein, this matrix was built from variation at 1,745 control SNPs that were assumed to be good representatives of the whole-genome neutral variation of maritime pine (see Materials and Methods in the main text). Control and candidate SNP markers had similar allele frequency distributions, as shown by a Kolmogorov-Smirnov test. In particular, control SNPs did not seem to be enriched by rare or common alleles, which could have indicated different levels of purifying selection (see Nielsen 2005). However, given the absence of factual information concerning their neutrality, some unnoticed non-stochastic bias might be present in the covariance matrix derived from these markers, as virtually any gene-derived SNP is a potential target of selective processes. For instance, if some of these SNPs were involved in the same adaptive processes aimed to be detected, they may have modified the covariance matrix in a way that does not fully represent the underlying neutral population structure, and thus would have reduced the statistical power for environmental associations.

To validate this covariance matrix, three different comparisons were performed. First, the matrices produced by each BAYENV 2.0 run were compared to each other with correlation tests, which were all highly significant (mean $r^2 = 0.943$; $P < 0.0001$), indicating that independent runs were converging to similar results. These covariance matrices were afterwards converted into correlation matrices using the *cov2cor* function in R and compared to a pairwise- F_{ST} matrix derived from this same SNP dataset and to a second matrix derived from SSR markers. Again, the correlation between matrices was extremely high (mean $r^2 = 0.974$; $P < 0.001$ for the SNP- F_{ST} matrix; mean $r^2 = 0.812$; $P < 0.005$ for the SSR one), implying that the covariance matrices were reflecting similar population differentiation values than the F_{ST} statistics. Finally, two independent STRUCTURE analyses were performed for the control-SNP and SSR datasets and their Q -matrices were correlated. For both types of markers, the best partition was that of six genetic groups (Figure 1), which matched perfectly those previously reported in other studies of this species (Bucci *et al.* 2007; Santos-del-Blanco *et al.* 2012). The correlation coefficients (Pearson's r) between the Q -matrices were relatively high and significant (correlation coefficients of 0.662; $P < 0.01$), although the groups obtained from the SNP dataset appeared better resolved (i.e. with less admixed individuals) than those determined with SSRs (see Figure 1). This is probably due to lower homoplasy and higher number of markers in the SNP dataset, which then should be more adequate to capture the underlying population structure of maritime pine than SSRs. This finding strongly supports our choice of the SNP dataset (and not

the SSRs) to compute the covariance matrix in null models for Bayesian association analyses. It also explains why correlation coefficients at the individual level are just moderately high between SNP and SSR Q-matrices.

LITERATURE CITED

- Bucci, G., S. C. González-Martínez, G. Le Provost, C. Plomion, M. M. Ribeiro *et al.*, 2007 Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Mol. Ecol.* **16**: 2137-2153.
- Coop, G., J. K. Pickrell, J. Novembre, S. Kudaravalli, J. Li *et al.*, 2009 The role of geography in human adaptation. *PLoS Genetics* **5**: e1000500.
- Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard, 2010 Using environmental correlations to identify loci under local adaptation. *Genetics* **185**: 1411-1423.
- Günther, T., and G. Coop, 2013 Robust identification of local adaptation from allele frequencies. *Genetics* **195**: 205-220.
- Nielsen, R., 2005 Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**:197-218.
- Santos-del-Blanco, L., J. Climent, S. C. González-Martínez, and J. R. Pannell, 2012 Genetic differentiation for size at first reproduction through male versus female functions in the widespread Mediterranean tree *Pinus pinaster*. *Ann. Bot.* **110**: 1449-1460.

File S2

**Design file for the 384-SNP (Single Nucleotide Polymorphism) candidate gene OPA used to genotype 36 populations of
maritime pine (*Pinus pinaster*)**

File S2 is available for download as an Excel file at

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.173252/-/DC1>

Table S1 Mean and standard errors for the number of individuals genotyped (*N*), the observed number of alleles (*Na*), the effective number of alleles (*Ne*), the observed (*Ho*) and expected (*He*) heterozygosity, and the consanguinity index (*F*) estimated for 36 populations of maritime pine (*Pinus pinaster*) with 12 microsatellite loci. The number of populations deviating from Hardy-Weinberg Equilibrium (HWE) at *P* < 0.01 is shown in the last column. Shaded loci were removed from further analyses.

Locus	<i>N</i>		<i>Na</i>		<i>Ne</i>		<i>Ho</i>		<i>He</i>		<i>F</i>		HWE ^a
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	
<i>A6F03</i>	19.436	2.092	4.667	0.189	2.854	0.115	0.568	0.021	0.622	0.019	0.075	0.028	4
<i>3pet</i>	10.846	1.503	5.923	0.707	3.884	0.459	0.568	0.060	0.572	0.059	0.006	0.021	5
<i>NZPR1078</i>	19.231	2.001	3.154	0.086	2.366	0.073	0.609	0.023	0.560	0.015	-0.091	0.033	1
<i>NZPR544</i>	19.410	2.059	2.359	0.107	1.881	0.046	0.316	0.025	0.457	0.013	0.311	0.049	12 (0.312)
<i>ctg4363</i>	20.487	2.125	4.154	0.174	2.701	0.082	0.650	0.024	0.617	0.011	-0.053	0.035	3
<i>rptest1</i>	19.923	2.142	3.692	0.133	2.770	0.115	0.646	0.024	0.612	0.018	-0.057	0.027	1
<i>ctg275</i>	20.385	2.142	7.949	0.380	4.090	0.229	0.649	0.027	0.721	0.018	0.097	0.033	12 (0.191)
<i>2669</i>	19.795	2.052	4.026	0.231	1.985	0.099	0.448	0.034	0.442	0.030	-0.013	0.031	5
<i>epi5</i>	19.077	2.090	3.128	0.192	1.605	0.071	0.274	0.026	0.333	0.027	0.133	0.047	2
<i>epi6</i>	15.974	1.325	3.974	0.239	2.499	0.157	0.298	0.031	0.553	0.031	0.458	0.041	19 (0.283)
<i>gPp14</i>	19.513	2.048	2.923	0.129	1.605	0.075	0.315	0.028	0.329	0.028	0.016	0.035	3
<i>epi3</i>	18.538	2.072	5.154	0.274	2.953	0.153	0.621	0.029	0.634	0.015	0.027	0.034	4

^aAverage null-allele frequency for discarded loci in populations deviating from HWE is given between parentheses, as estimated with MICRO-CHECKER software (vanOosterhout *et al.* 2004).

Literature cited: vanOosterhout, C., W. F. Hutchinson, D. P. M. Willis, and P. Shipley, 2004 Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**: 535-538.

Table S2 Percentage of the variance explained (% Variance), factor loads, and population scores for the top-three principal components summarizing winter (December to March) and summer (June to September) climatic variation in 36 populations of maritime pine (*Pinus pinaster*).

	Winter			Summer		
	PC1	PC2	PC3	PC1	PC2	PC3
% Variance	67.861	20.981	7.462	61.752	27.838	6.299
<i>Factor</i>						
Mean Temperature - December	0.966	-0.132	-0.002	---	---	---
Mean Temperature - January	0.980	-0.148	-0.087	---	---	---
Mean Temperature - February	0.982	-0.175	0.015	---	---	---
Mean Temperature - March	0.949	-0.235	0.020	---	---	---
Highest Temperature - December	0.927	-0.063	0.291	---	---	---
Highest Temperature - January	0.869	-0.090	0.468	---	---	---
Highest Temperature - February	0.717	-0.198	0.580	---	---	---
Highest Temperature - March	0.941	-0.089	0.209	---	---	---
Lowest Temperature - December	0.911	-0.197	-0.352	---	---	---
Lowest Temperature - January	0.927	-0.206	-0.306	---	---	---
Lowest Temperature - February	0.914	-0.212	-0.314	---	---	---
Lowest Temperature - March	0.921	-0.191	-0.325	---	---	---
Mean Precipitation - December	0.527	0.832	-0.018	---	---	---
Mean Precipitation - January	0.489	0.862	-0.007	---	---	---
Mean Precipitation - February	0.428	0.843	0.004	---	---	---
Mean Precipitation - March	0.514	0.820	-0.067	---	---	---
Mean Temperature - June	---	---	---	0.882	0.356	0.200
Mean Temperature - July	---	---	---	0.973	-0.050	0.138
Mean Temperature - August	---	---	---	0.979	0.055	0.126
Mean Temperature - September	---	---	---	0.816	0.511	0.104
Highest Temperature - June	---	---	---	0.879	-0.316	0.305
Highest Temperature - July	---	---	---	0.803	-0.546	0.208
Highest Temperature - August	---	---	---	0.842	-0.489	0.213
Highest Temperature - September	---	---	---	0.893	-0.238	0.292
Lowest Temperature - June	---	---	---	0.414	0.887	-0.022
Lowest Temperature - July	---	---	---	0.590	0.780	-0.069
Lowest Temperature - August	---	---	---	0.576	0.795	-0.078
Lowest Temperature - September	---	---	---	0.346	0.918	-0.112
Mean Precipitation - June	---	---	---	-0.855	0.083	0.439
Mean Precipitation - July	---	---	---	-0.877	0.281	0.302
Mean Precipitation - August	---	---	---	-0.789	0.442	0.374

Mean Precipitation - September		---	---	---	-0.718	0.464	0.455
<i>Population</i>	Country						
Hourtin	France	-0.020	-0.390	-0.538	-1.760	1.482	0.943
Leverdon	France	-0.376	-0.670	-0.896	-1.447	1.593	0.488
Mimizan	France	1.440	0.315	0.108	-2.198	2.250	2.205
Olonne sur Mer	France	-0.863	-0.540	-1.626	-2.395	1.209	-0.676
Petrocq	France	1.414	0.380	0.113	-2.228	2.269	2.284
Pleucadec	France	-1.990	-0.300	-1.927	-3.886	0.368	-1.370
St-Jean des Monts	France	-0.866	-0.471	-1.626	-2.629	1.027	-0.800
Alto de la Llama	Spain	0.250	1.066	-0.491	-3.580	-0.714	-0.810
Armayán	Spain	0.235	0.818	0.169	-3.134	-0.640	-0.383
Cadavedo	Spain	3.377	0.502	-0.196	-3.718	1.747	0.129
Sierra de Barcia	Spain	3.123	0.753	-0.209	-3.821	1.446	-0.100
Castropol	Spain	2.476	0.102	-0.184	-4.465	0.576	-0.543
Lamuño	Spain	3.643	0.342	-0.277	-3.111	2.093	-0.049
Puerto de Vega	Spain	3.384	0.199	0.069	-3.149	1.732	0.303
Rodoiros	Spain	2.126	2.486	-0.984	-5.861	0.260	-0.260
San Cipriano de Ribaterme	Spain	1.915	3.717	0.025	-2.063	-1.495	0.296
Leiria	Portugal	5.809	-1.742	0.065	1.149	1.582	-2.328
Pineta	France (Corsica)	3.006	-2.410	-1.465	3.803	3.400	-1.500
Pinia	France (Corsica)	3.188	-2.479	-0.904	3.947	3.266	-1.108
Tabuyo del Monte	Spain	-6.793	1.849	-1.309	-3.328	-1.461	1.535
Arenas de San Pedro	Spain	1.213	2.568	1.372	4.554	-1.084	2.135
Valdemaqueda	Spain	-2.344	-0.548	-0.126	1.681	-1.214	-0.339
Cenicientos	Spain	-1.699	-0.071	-1.018	1.941	-0.123	-0.996
Coca	Spain	-2.842	-1.725	1.224	2.042	-2.668	0.641
Cuellar	Spain	-2.915	-1.405	1.082	1.975	-2.861	0.590
Bayubas de Abajo	Spain	-5.089	-0.448	-0.127	-0.406	-3.358	0.008
San Leonardo	Spain	-6.337	1.121	-0.683	-2.916	-4.342	-0.540
Boniches	Spain	-3.909	-0.146	-0.766	0.014	-1.770	-0.092
Olba	Spain	-2.068	-2.192	0.565	0.736	-0.146	0.257
Sinarcas	Spain	-2.963	-1.836	0.041	1.554	0.215	-0.242
Sierra Calderona	Spain	-1.294	-2.118	-0.064	1.584	1.339	-0.410
Quatretonda	Spain	3.256	-1.819	1.889	4.210	1.295	0.944
Cazorla	Spain	-2.020	4.007	0.061	0.922	-2.711	0.002
Oria	Spain	-1.341	-2.193	1.204	3.574	-1.543	-0.429
Tamrabta	Morocco	-4.910	1.378	1.428	0.604	-3.725	-0.828
Tabarka	Tunisia	7.272	-1.045	1.043	8.234	3.671	0.930

Table S3 Strength of genotype-environment correlations (Akaike Information Criterion in multivariate logistic models –*mlr*, *P*-values in latent factor mixed models, and Bayes Factors in Bayesian environmental associations with BAYENV 2.0) for 18 maritime pine SNPs associated with climate.

SNP name (number; see Fig. S1)	Climate PC	<i>mlr</i>	LFMM	BAYENV ^a	
		AIC	<i>P</i>	BF1	BF2
<i>m783</i> (227)	PC-Winter1	21.6	1.28×10^{-19}	36.81	39.64
<i>m8</i> (89)	PC-Winter1	9.3	5.95×10^{-19}	26.43	25.62
<i>m1513</i> (53)	PC-Summer2	15.4	1.29×10^{-15}	29.41	33.84
<i>m1250</i> (142)	PC-Winter2	29.7	9.09×10^{-20}	34.46	38.03
<i>m80</i> (218)	PC-Summer2	7.3	2.18×10^{-23}	22.31	21.04
<i>m80</i> (218)	PC-Winter1	12.4	8.95×10^{-25}	40.84	38.89
<i>m1309</i> (176)	PC-Winter1	10.1	5.69×10^{-21}	25.37	24.12
<i>m100</i> (93)	PC-Winter3	13.6	8.21×10^{-11}	29.13	27.48
<i>m100</i> (93)	PC-Summer2	9.6	9.96×10^{-13}	18.46	15.84
<i>m973</i> (118)	PC-Winter2	10.2	7.65×10^{-11}	12.33	12.98
<i>m137</i> (46)	PC-Summer1	7.3	8.73×10^{-15}	20.31	18.44
<i>m607</i> (194)	PC-Winter3	18.5	4.12×10^{-15}	16.78	18.44
<i>m1156</i> (131)	PC-Winter1	16.4	1.24×10^{-11}	14.16	13.96
<i>m1211</i> (159)	PC-Winter1	22.8	2.34×10^{-24}	15.26	14.01
<i>m646</i> (19)	PC-Winter1	11.6	3.93×10^{-21}	17.29	15.52
<i>m657</i> (238)	PC-Winter1	15.9	2.21×10^{-23}	20.11	18.34
<i>m658</i> (47)	PC-Summer2	32.1	5.45×10^{-19}	21.91	22.43
<i>m685</i> (41)	PC-Winter2	16.0	7.52×10^{-11}	14.32	12.97
<i>m685</i> (41)	PC-Summer3	19.8	3.78×10^{-16}	12.58	11.86
<i>m1196</i> (238)	PC-Winter2	24.3	2.33×10^{-14}	14.87	16.12
<i>m705</i> (127)	PC-Summer2	13.6	3.18×10^{-15}	14.31	15.96

^aResults are shown for two independent BAYENV 2.0 runs made with a covariance matrix built using 1,745 control SNPs (see Materials and Methods for more details).

Table S4 Single-locus regressions of advantageous-allele frequency with survival estimated under extreme climate conditions for each of 18 SNPs associated with climate in maritime pine (*Pinus pinaster*).

SNP name	Advantageous allele	DF	<i>F</i>	<i>P</i>	<i>Adj-r</i> ²
<i>m783</i>	A	17	4.037	0.061	0.144
<i>m8</i>	A	17	5.773	0.028	0.210
<i>m1513</i>	C	17	5.182	0.036	0.188
<i>m1250</i>	A	17	0.095	0.762	-0.053
<i>m80</i>	A	17	7.178	0.016	0.256
<i>m1309</i>	G	17	7.865	0.012	0.276
<i>m100</i>	A	17	3.929	0.064	0.140
<i>m973</i>	G	17	1.976	0.178	0.051
<i>m137</i>	G	17	0.133	0.720	-0.051
<i>m607</i>	C	17	2.083	0.167	0.057
<i>m1156</i>	G	17	0.085	0.774	-0.054
<i>m1211</i>	C	17	8.380	0.010	0.291
<i>m646</i>	G	17	0.802	0.383	-0.011
<i>m657</i>	A	17	5.891	0.027	0.214
<i>m658</i>	G	17	0.825	0.376	0.046
<i>m685</i>	A	17	2.484	0.133	0.076
<i>m1196</i>	C	17	2.284	0.149	0.067
<i>m705</i>	A	17	5.485	0.032	0.200

Table S5 Correlation between the 18 SNPs that best explained each Principal Component (PC) in a PCA including all 266 candidate-gene SNPs and survival estimates obtained in a common garden under extreme (dry and hot) climate; ns: not significant.

Principal Components	% variance explained	Pearson's correlation coefficient (<i>r</i>)
PC1	24.30	0.28 (ns)
PC2	15.55	0.27 (ns)
PC3	10.23	-0.41 (ns)
PC4	9.16	0.26 (ns)

Table S6 Correlation coefficients for the non-linear models that best fitted the climate principal components (see Table S2) of maritime pine (*Pinus pinaster*) populations with their geographic location (latitude and longitude) and population structure loads (control SNP and SSR datasets). ** $P < 0.01$

Climate PC	Geography (latitude - longitude)	Population Structure
PC-Winter1	0.176	0.179
PC-Winter2	0.140	0.153
PC-Winter3	0.849**	0.593**
PC-Summer1	0.196	0.181
PC-Summer2	0.392**	0.706**
PC-Summer3	0.110	0.114

ACKNOWLEDGEMENTS

Llegó el momento de poner punto y final a esta interesante fase vital. No sin antes dar las gracias a todos aquellos que me han acompañado en el camino y han hecho posible esta tesis.

En primer lugar, gracias a mis directores. Gracias Ricardo por elegir a una agrónoma para lo que pensé que iba a ser una tesis de "mejora de pinos". Desde el primer día, cercano y maestro al mismo tiempo. Gracias a ti, ahora formo parte del apasionante mundo de la genética cuantitativa y de la conservación forestal. Además estoy en deuda contigo por haber hecho de mis años de doctorado un período lleno de actividades formativas, cursos, congresos y estancias a lo largo y ancho del mundo. Siempre dispuesto a ayudar, y, como todos los genios, muy despistado. Pero, por encima de todo, una de las mejores personas que he conocido nunca. Santi vino después. Realmente, en este caso fui yo quien le eligió a él, y no al contrario. Fui afortunada de ser adoptada como doctoranda compartida. Poco a poco mi gran admiración distante de los comienzos fue transformándose en respeto hacia el trabajo duro de una persona tremendamente inteligente. Me ayudaste a entender el *ricardiano* (idioma en el que mi otro director de tesis, que también fue el suyo, se comunica a veces por escrito, faltando sujetos o predicados) y pusiste el contrapunto práctico a la tesis, ayudando enormemente a transformar grandes ideas en artículos publicables. Gracias por tu tiempo, por las horas de conversación científica y por la etapa suiza compartida con la familia. Además de un codirector genial, gané un amigo.

Gracias a todo el equipo del CIFOR por su ayuda durante estos años. Ha sido un placer trabajar con vosotros. Gracias a Delphine y Juanjo por las discusiones científicas, y a Jose y Regina por su ayuda. Gracias al equipo de campo y laboratorio: Diana, Fer y Edu. Sin vosotros no habría datos, y sin datos, no habría tesis. Gracias Jesús por tu ayuda con los mapas. Gracias también a Toñi por hacer más fácil el papeleo en secretaría.

Mi tesis me ha llevado además a muchos sitios fuera de nuestras fronteras. A Praga, París, Uppsala y Edimburgo para asistir a cursos y congresos, a Lisboa, Burdeos, Lausana y Estados Unidos de estancia. I want to thank Andrew J. Eckert for his time and hard work in our paper during my research stay in Richmond. Je voudrais remercier Christophe Plomion pour leur aide lors de mon séjour à l'INRA de Bordeaux et de l'entretien et de la mesure du test clonal de pin maritime. Agradecimentos a José Rodrigues por sua ajuda com a análise da qualidade da madeira de pinho durante a minha estadia em Lisboa. I would also like to thank John Pannell for his help and hospitality during our stay in Lausanne. Mi tesis también me ha llevado a ciudades españolas como Vitoria, Pamplona o Sevilla, y a disfrutar de

muchas muchas horas de campo en Fundão (Portugal), Madrid, Cáceres, Valencia, Asturias y Burdeos. Ha sido duro, pero divertido.

No puede faltar un párrafo dedicado a la sala de becarios. Durante estos años hemos compartido mucho, destacando los desayunos, las alegrías vitales y los sufrimientos tésicos. Cuando llegué a la sala ¡ni siquiera cabíamos! Gracias a todos lo compartido, fue muy fácil desde el principio sentirse parte de la familia. Enhorabuena por todas las tesis terminadas y mucho ánimo a los poquitos que quedan en su camino hacia la luz. Gracias especiales a Javi y Katha por su simpatía y comprensión y por el tiempo compartido, y a Gregor, que empezó siendo mi compi de la derecha, para convertirse después en compañero de fatigas en el duro camino del doctorando. Gracias por las horas de conversación y los cafés, por haber seguido descubriendo Madrid contigo y por esos dulces austriacos que nos han ayudado a todos a hacer más llevaderos los baches de la tesis. Tienes que acabar, me lo debes. Y si hablamos de terminar tesis, Maje es la razón de que la mía esté ya acabada. Gracias por ese empujón, por los ánimos y por esa sonrisa permanente que hace pequeñito cualquier obstáculo.

Gracias a mi familia por su apoyo incondicional, por su ayuda y por ser como son. En cada paso me siento tremendamente afortunada por tenerlos cerca. Gracias a Elmo por acompañarme siempre, por su cariño infinito, su fuerza y su alegría.

Gracias a mi tesis por presentarme a Luis. Y gracias a Luis por su ayuda desde los orígenes. Empezaste enseñándome qué era un modelo mixto, y ahora me enseñas a ser mejor persona cada día estando a tu lado. Finalmente, gracias a Sofía por hacer nuestra vida más completa y apasionante.