

ORIGINAL ARTICLE

The population genomic signature of environmental selection in the widespread insect-pollinated tree species *Frangula alnus* at different geographical scales

H De Kort¹, K Vandepitte¹, J Mergeay², KV Mijnsbrugge^{2,3} and O Honnay¹

The evaluation of the molecular signatures of selection in species lacking an available closely related reference genome remains challenging, yet it may provide valuable fundamental insights into the capacity of populations to respond to environmental cues. We screened 25 native populations of the tree species *Frangula alnus* subsp. *alnus* (Rhamnaceae), covering three different geographical scales, for 183 annotated single-nucleotide polymorphisms (SNPs). Standard population genomic outlier screens were combined with individual-based and multivariate landscape genomic approaches to examine the strength of selection relative to neutral processes in shaping genomic variation, and to identify the main environmental agents driving selection. Our results demonstrate a more distinct signature of selection with increasing geographical distance, as indicated by the proportion of SNPs (i) showing exceptional patterns of genetic diversity and differentiation (outliers) and (ii) associated with climate. Both temperature and precipitation have an important role as selective agents in shaping adaptive genomic differentiation in *F. alnus* subsp. *alnus*, although their relative importance differed among spatial scales. At the 'intermediate' and 'regional' scales, where limited genetic clustering and high population diversity were observed, some indications of natural selection may suggest a major role for gene flow in safeguarding adaptability. High genetic diversity at loci under selection in particular, indicated considerable adaptive potential, which may nevertheless be compromised by the combined effects of climate change and habitat fragmentation.

Heredity (2015) 115, 415–425; doi:10.1038/hdy.2015.41; published online 6 May 2015

INTRODUCTION

Natural selection is an important driver of evolution, as advantageous alleles can be inherited across generations to maximize fitness in the local environment (Kawecki and Ebert, 2004; Aitken *et al.*, 2008). Although neutral processes such as drift and migration exert a genome-wide effect, divergent selection may result in clear allele frequency shifts at loci under selection and/or subtle changes in adaptive allelic combinations (Latta, 1998; Le Corre and Kremer, 2003; Kremer and Le Corre, 2012; Eckert *et al.*, 2013). These genomic signatures of selection are expected to be more pronounced with increasing geographical scale because of larger environmental differences and decreased gene flow across longer distances (Volis *et al.*, 2005; Fraser *et al.*, 2011). Unraveling the molecular signature of selection at different geographic scales and along different environmental gradients may reveal fundamental insights into the evolutionary processes underlying species' abilities to respond to environmental cues, and it may assist management decisions with regard to the preservation of adaptive differentiation within species (Eklblom and Galindo, 2011; Breed *et al.*, 2012; Funk *et al.*, 2012; Narum *et al.*, 2013).

Although the identification of loci that are under selection remains challenging in non-model species without reference genome sequences, several methods exist that aim to detect genetic variation putatively driven by selective forces (Seeb *et al.*, 2011; Schoville *et al.*,

2012; Jones *et al.*, 2013; Joost *et al.*, 2013; Weing *et al.*, 2014). First, the genome can be screened for outlier loci, by identifying genetic marker variation deviating from overall expected neutral distributions (Luikart *et al.*, 2003; Excoffier *et al.*, 2009). This may, however, render false negative results in the case of weak divergent selection or when small-effect loci are involved, resulting in limited differences in allele frequencies among populations, and therefore low power to detect weak or polygenic selection (Narum and Hess, 2011; De Villemereuil *et al.*, 2014). Moreover, complex demographic history and population structure tend to increase the number of false positives generated by outlier detection methods (Foll and Gaggiotti, 2008; Jones *et al.*, 2013). Landscape genomic approaches do not rely on F_{ST} -based neutrality assumptions but attempt to find associations between allele distributions and environmental variables presumed to be the drivers of selection (Manel *et al.*, 2003; Eckert *et al.*, 2010; Jones *et al.*, 2013; Manel and Holderegger, 2013; Sork *et al.*, 2013). Although landscape genomics approaches considerably increase the likelihood of finding weak and polygenic selection (De Mita *et al.*, 2013), correlations among environmental variables (that is, spatial autocorrelation), and lack of knowledge regarding the true drivers of selection, may increase the amount of false positive results (De Villemereuil *et al.*, 2014). In addition, it is assumed that there was sufficient time for selection to drive patterns of allele frequency distributions caused by the

¹Biology Department, University of Leuven, Plant Conservation and Population Biology, Kasteelpark Arenberg 31, Heverlee, Belgium; ²Department of Genetic Diversity, Research Institute for Nature and Forest, Gaverstraat 4, Geraardsbergen, Belgium and ³Department of Nature Conservation, Agency for Nature and Forest, Koning Albert II laan 20, Brussels, Belgium

Correspondence: Dr H De Kort, Biology Department, KU Leuven, Kasteelpark Arenberg 31, Heverlee 3000, Belgium.

E-mail: hanne.dekort@bio.kuleuven.be

Received 27 August 2014; revised 20 February 2015; accepted 20 March 2015; published online 6 May 2015

environmental agents of selection (Joost *et al.*, 2013). Yet when used together, population genomic and landscape genomic approaches may complement each other in terms of power and accuracy (Joost *et al.*, 2013; De Villemereuil *et al.*, 2014; Weinig *et al.*, 2014).

Despite the importance of tree species with regard to ecosystem service provisioning (for example, carbon fixation timber production, water regulation and erosion control), ecogenomic studies on tree species that lack closely related reference genomes are scarce. Although two studies have investigated the genetic signature of selection using amplified fragment length polymorphism markers in tree species hitherto lacking information on gene functions (*Alnus glutinosa*, Cox *et al.*, 2011, and *Fagus sylvatica*, Jump *et al.*, 2006), to our knowledge, only one study so far has screened the uncharacterized genome of a tree species for genome-wide patterns of natural selection using completely anonymous single-nucleotide polymorphisms (SNPs) (*A. glutinosa*, De Kort *et al.* (2014)). In contrast, tree genera for which genomic resources are available (for example, *Pinus* sp., *Picea* sp., *Eucalyptus* sp., and *Populus* sp.) have been frequently examined for patterns of environmental selection (for example, Grivet *et al.*, 2011; Prunier *et al.*, 2011; Keller *et al.*, 2011; Bradbury *et al.*, 2013; Steane *et al.*, 2014). Hence, although genomic markers provide valuable information on the capacity of tree species to respond to climatic cues, little is known about the relative effects of climate and geographical scale on the genomic differentiation in non-model tree species. A most effective strategy for gathering genomic resources in non-model tree species involves pooled restriction-sitesite-associated DNA (RAD) sequencing, which relies on the sequencing of a few hundreds of base pairs adjacent to restriction enzyme recognition sites (Baird *et al.*, 2008). Pooling a limited number of individuals sampled across the species' distribution range greatly reduces the costs of discovering thousands of SNPs (Vandepitte *et al.*, 2013). Moreover, pooled RADseq opens the possibility to genotype a selected set of candidate SNPs in a large population samples at relatively low costs using allele specific primers (Cuppen, 2007; Liu *et al.*, 2012; Vandepitte *et al.*, 2013), as the discovered SNPs can be functionality annotated by using the genome of a model species as a reference (Paterson *et al.*, 2010). SNPs with putatively adaptive functions constitute excellent candidates for population and landscape genomic research, whereas a set of random SNPs lacking functional annotations can be used to simulate neutral expectations.

Here, we performed a population and landscape genomic investigation of *Frangula alnus* subsp. *alnus* Mill. (Rhamnaceae), an insect-pollinated, deciduous small tree with a high seed dispersal capacity and a wide geographical range (Medan, 1994; Hampe, 2004; Hampe, 2005). Our major aims were (i) to evaluate the relative strength of selection in shaping the distribution of genomic variation at different spatial scales and (ii) to identify putative environmental drivers. To achieve these goals, population and landscape genomics approaches were used to screen a previously developed panel of annotated SNPs (De Kort *et al.*, 2013) for the molecular imprints of selection across a continental scale (Europe), a regional scale (the Flemish region, Belgium) and an intermediate scale (the regional scale together with a southern Belgian and a French region). More specifically, a standard Bayesian population genomic outlier screen (BAYESCAN, Foll & Gaggiotti, 2008) was complemented with individual-based latent factor mixed modeling (LFMM, Frichot *et al.*, 2013), and multivariate redundancy analyses (RDA, Borcard *et al.*, 2011), to identify (i) SNPs putatively under selection and (ii) the main environmental drivers of population differentiation while accounting for spatial autocorrelation in allele frequencies. Although the use of candidate SNPs has been the standard method to find loci putatively under selection (for example,

Namroud *et al.*, 2008; Alberto *et al.*, 2013; Guichoux *et al.*, 2013; Olson *et al.*, 2013), they may compromise the relevance of classical outlier detection methods due to overestimation of neutral distributions. This stresses the importance of integrating landscape genomic methods to reliably reveal loci putatively under selection and the selective agents underlying the observed patterns (De Villemereuil *et al.*, 2014; Weinig *et al.*, 2014).

MATERIAL AND METHODS

Study species

Frangula alnus (Rhamnaceae) is a small, self-incompatible, insect-pollinated tree that occurs throughout most of Europe and in northern Morocco (Medan, 1994; Hampe, 2005). The most prevalent subspecies (*F. alnus alnus*) occurs throughout most of the continental range, whereas the two other subspecies are limited to Central and East Anatolia (*F. alnus pontica*), and South Iberia and North Morocco (*F. alnus baetica*). The fruits of the species are dispersed by birds and water which may transfer alleles over large distances (Hampe, 2004; Kremer *et al.*, 2012). *F. alnus* grows on soils with diverse humidity levels, varying from relatively dry sandy soils over ditches and wet forest edges to wetlands (Godwin, 1943; Hampe *et al.*, 2003; Fiedler and Landis, 2012).

Samples and genotyping

We collected samples from 25 populations of *F. alnus* subsp. *alnus* across Europe, largely covering the extent of the subspecies' distribution (Figure 1, Table 1). The 25 populations were sampled across eight regions, with three populations from each of four regions in Flanders (northern Belgium: Sandy region, Western Brabant, Eastern Brabant and Campine), three populations from the Ardennes region (southern Belgium), three populations from both the Picardy region (northern France) and the Medelpad region (Sweden) and four populations from the Tuscany region (Italy). The populations from Medelpad and Tuscany represent the northern and southern edge of the *F. alnus* subsp. *alnus* geographical distribution range, respectively. We explored patterns of genetic clustering and natural selection across three nested spatial scales: the regional scale (four regions in Flanders), the intermediate scale (the four Flemish regions + Picardy + Ardennes; six regions) and the continental scale, also including Tuscany and Medelpad (eight regions).

A total of 619 *F. alnus* subsp. *alnus* trees were genotyped for 250 putative SNPs, which were selected from a large set of 7383 previously discovered SNPs (for the full discovery, genotyping and annotation procedure see De Kort *et al.* (2013)). To reduce SNP discovery induced ascertainment bias, which refers to the systematic underrepresentation of existing SNPs, and which mainly affects estimates of population size among other demographic parameters (Nielsen *et al.*, 2004; Rosenblum & Novembre, 2007), we assembled our ascertainment panel from individuals sampled across the distribution range of *F. alnus* subsp. *alnus* (Morin *et al.*, 2004; Nielsen *et al.*, 2004). Furthermore, the high sequencing coverage (on average 39.7x), the exclusion of rare alleles, and the small genome of *F. alnus* (see De Kort *et al.*, 2013) reduce the occurrence of sequencing and read assembly errors that may mimic SNPs (Da Silva *et al.*, 2014).

Briefly, paired-end sequencing of restriction-site-associated DNA (RAD-PE; Etter *et al.*, 2011; Vandepitte *et al.*, 2013) was used on a pooled set of eight individuals distributed across the species' natural range, to find polymorphisms randomly distributed across the *F. alnus* subsp. *alnus* genome. The contig sequences containing the 7383 detected biallelic putative SNPs were subsequently subjected to a blastx against all plant proteins available at the Swiss-Prot database

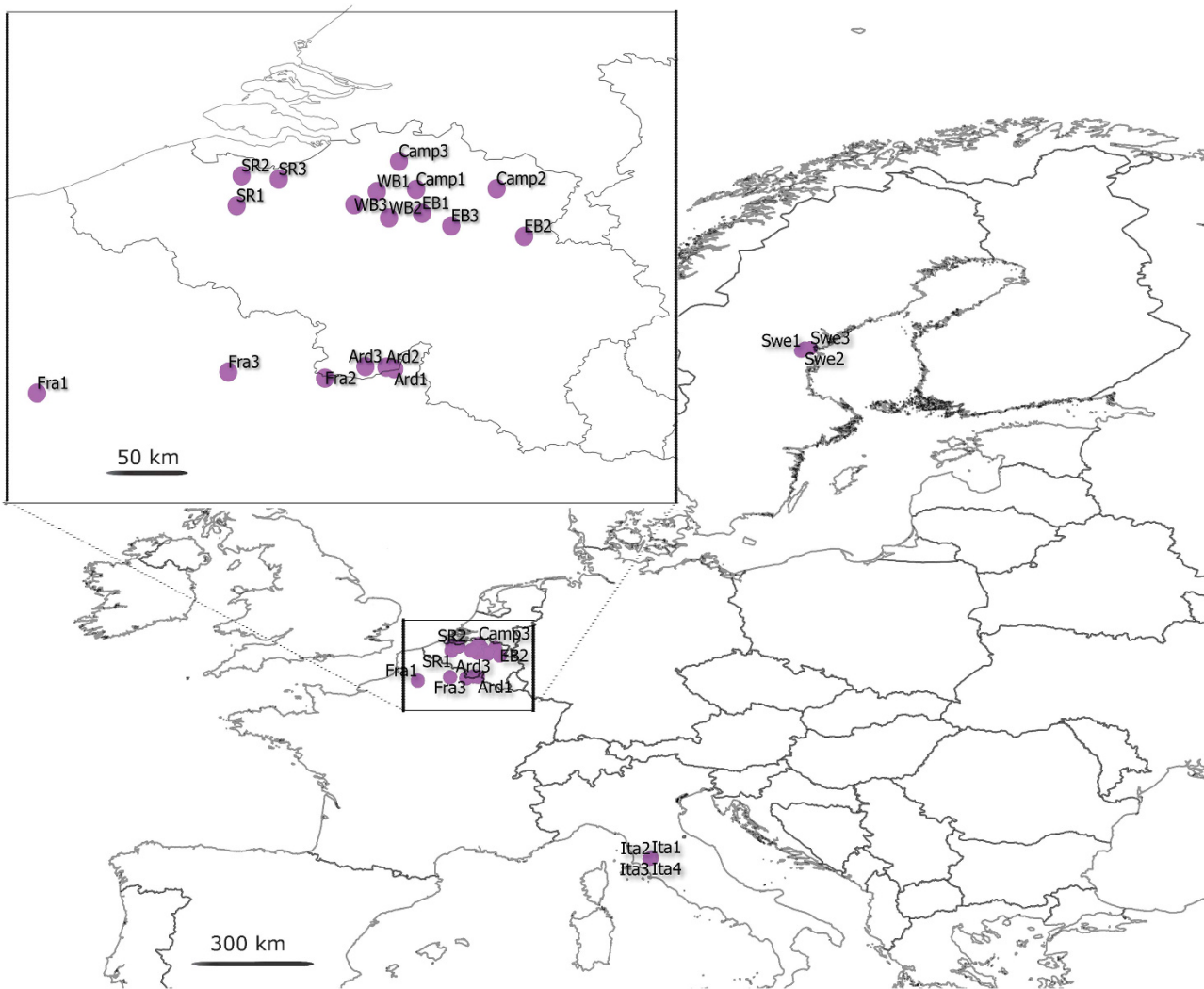


Figure 1 Sampling locations of *F. alnus* subsp. *alnus*. Twelve populations in Flanders (northern Belgium), three populations in the Ardennes (southern Belgium), three populations in Picardy (France), three populations in Medelpad (Sweden) and four populations in Tuscany (Italy). Population codes are explained in Table 1.

using Blast2Go, and hits were mapped to their corresponding gene annotations using the Gene Ontology (GO) database and several additional database files for functional categorization. This rendered annotation results for 306 SNPs, of which the vast majority ($n=250$) were selected for genotyping using allele-specific primer extension genotyping (KASPar, LGC Genomics, Teddington, UK). First, we selected all SNPs assigned to one or more of the following, broad higher level GO slim (<http://www.geneontology.org/>) categories: responses to stimulus (137 GO terms), reproduction (84 GO terms) and developmental processes (225 GO terms) (Supplementary Information 2, De Kort *et al.*, 2013). This yielded a set of 180 putative SNPs which was then complemented with 70 SNPs taken randomly from other categories. The selection of SNPs representing a wide array of gene functions should ensure that overall patterns approximate genome-wide genic variation (see Vandepitte *et al.*, 2014 for a similar approach). Moreover, the use of candidate SNPs may increase the likelihood of finding loci putatively associated with selective agents considerably. Of the 250 assayed SNPs, 183 had fewer than 10% missing data within individuals, and could be genotyped successfully by using a

allele-specific primer extension (De Kort *et al.*, 2013). This set of 183 SNPs was used for population genomic data analysis (see below).

Population structure and genomics

For each population, estimates of expected (H_E) and observed (H_O) heterozygosity were obtained by using GENALEX 6.4 (Peakall and Smouse, 2006). A mixed linear model was built to compare H_E estimates between neutral and outlier SNPs (see below for the outlier detection method). To account for non-independency among data within populations and regions, we included 'region' and 'population nested within region' as random factors. For each pair of populations, we computed pairwise F_{ST} values with ARLEQUIN 3.5 (Excoffier and Lischer, 2010), using 10 000 permutations. A principal coordinates analysis on genetic distances (GENALEX 6.4) was performed to detect major genetic clusters (K) at the three nested spatial scales. Only the neutral SNPs were used to make sure that the revealed clusters represent neutral population structure.

To identify putative loci under selection at the three scales, we used BAYESCAN 2.1., which compares the allele frequencies of each

Table 1 Population code and location, sample size, expected and observed heterozygosity of the sampled populations

Population	Code	Region	Country	No. individuals	Latitude (°N)	Longitude (°E)	H_E	H_O
De Pinte	SR1	Sandy region	Belgium	25	50.98565	3.649552	0.346	0.361
Evergem	SR2	Sandy region	Belgium	36	51.16598	3.681092	0.327	0.334
Moerbeke	SR3	Sandy region	Belgium	23	51.14555	3.918045	0.338	0.348
Sint-Katelijne-Waver	WB1	WB	Belgium	24	51.07125	4.543055	0.337	0.336
Herent	WB2	WB	Belgium	37	50.91215	4.619936	0.324	0.330
Zemst	WB3	WB	Belgium	27	50.99224	4.398996	0.328	0.334
Heist-op-den-Berg	Camp1	Campine	Belgium	30	51.08424	4.793124	0.343	0.352
Beringen	Camp2	Campine	Belgium	31	51.08847	5.305648	0.320	0.329
Zoersel	Camp3	Campine	Belgium	32	51.25348	4.683878	0.334	0.341
Holsbeek	EB1	EB	Belgium	18	50.93981	4.830925	0.341	0.360
Tongeren	EB2	EB	Belgium	21	50.80119	5.480932	0.307	0.309
Kortenaken	EB3	EB	Belgium	16	50.86372	5.01641	0.329	0.349
Thiérache	Ard1	Ardennes	Belgium	11	50.00368	4.653925	0.315	0.333
Regniëssart	Ard2	Ardennes	Belgium	20	50.01361	4.603547	0.337	0.346
Brûly-de-Pesche	Ard3	Ardennes	Belgium	33	50.01711	4.469786	0.341	0.350
Storåsmýran	Swe1	Medelpad	Sweden	16	62.33861	16.87907	0.297	0.304
Aborttjärn	Swe2	Medelpad	Sweden	25	62.42415	17.01771	0.295	0.308
Ottsjön	Swe3	Medelpad	Sweden	24	62.4421	17.23451	0.311	0.335
Casina Rossa	Ita1	Tuscany	Italy	12	43.16505	11.22042	0.257	0.257
Famelunga	Ita2	Tuscany	Italy	11	43.12181	11.17654	0.273	0.282
Luriano	Ita3	Tuscany	Italy	16	43.114894	11.157618	0.219	0.248
Fosso Bolza	Ita4	Tuscany	Italy	11	43.123665	11.221991	0.252	0.264
Boves	Fra1	Picardy	France	7	49.85694	2.378333	0.309	0.305
Saint-Michel	Fra2	Picardy	France	5	49.9475	4.212778	0.341	0.363
Hannapes	Fra3	Picardy	France	15	49.98417	3.596389	0.332	0.344

population with a common migrant gene pool (Foll and Gaggiotti, 2008). Hence, effective size and migration rate are allowed to differ among populations. Log10 values of the posterior odds (PO) > 0.5 and 1.5 were taken as ‘substantial’ and ‘very strong’ evidence for selection, respectively (Jeffreys, 1961; Foll and Gaggiotti, 2008). The false discovery rate was set at 0.05 and 0.01 to adjust the log10(PO) significance thresholds corresponding to the 0.5 and 1.5 values. Because we were mainly interested in patterns of adaptation driven by environmental gradients, we focused on outlier patterns indicating divergent selection (F_{ST} significantly higher than neutral expectations) rather than outliers for balancing selection (F_{ST} significantly lower than neutral expectations). BAYESCAN analyses were complemented with a hierarchical genome scan using ARLEQUIN 3.5 which, however, resulted in similar results and is not discussed further (see Table 2, Supplementary Information 1).

Selection of environmental variables

To test for associations between outlier loci and environment, we collected soil and climate data. For each population, a bulked soil sample was analyzed to quantify cation exchange capacity, moisture content and organic material content. To deal with the collinearity among the soil variables, a principal component analysis (SPSS v20, Armonk, NY, USA) was performed. The first component, which was positively correlated with all measured soil variables and explained 92% of the original variance (Supplementary Information 3), was used in all further analyses. Climate variables related to temperature and precipitation (maximum temperature of the warmest month, minimum temperature of the coldest month, mean temperature of the warmest quarter, mean temperature of the coldest quarter, precipitation of the wettest month, precipitation of the driest month) were downloaded from WorldClim (Hijmans *et al.*, 2005) with a 2.5 arc-minutes resolution. Both the extreme and average climatic variables

can be expected to drive natural selection. Values for each climate variable were extracted for each population using Quantum GIS (QGIS 1.9.0, <http://www.qgis.org>). To deal with collinearity among climate variables, a principal component analysis analysis on all climatic variables was performed. The climate variables were reduced to two principal components (explaining 83% of the variation), the first one positively related to temperature, and the second negatively related to precipitation (Supplementary Information 3). Taken together, one soil and two climate variables (soil, temperature and precipitation) were used for all subsequent analyses.

Landscape genomics: LFMM

We applied LFMM, to find significant associations with environmental factors while accounting for neutral population structure (Frichot *et al.*, 2013). For the analysis at the continental scale, three major genetic clusters ($K=3$) were chosen for the demographic background model, as revealed by the clustering of populations into three main groups (Supplementary Information 4). At the intermediate and regional scales, $K=1$ was chosen due to limited genetic structure (Supplementary Information 4). Although other K values may be more appropriate in cases of hidden population structure, results with different K values rendered equivalent results on all scales (Supplementary information 2). To estimate the LFMM parameters, an MCMC algorithm was used for each of the five environmental variables (temperature, precipitation, soil, latitude, longitude), using 1000 sweeps for burn-in and 10 000 additional sweeps to compute l -scores for all loci (Frichot *et al.*, 2013). Significance was determined using a standard Gaussian distribution and corrected for multiple testing using the R package q-value v1.43 (Dabney *et al.*, 2011). We considered that a SNP effect was significant when its l -score was > 4, corresponding to a q -value of $q < 10^{-5}$. For each SNP, we were particularly interested in the environmental variable with the highest

Table 2 Outlier SNPs and their significant ($|z|$ -score > 4) associations with environmental variables using LFMM

SNP	Gene description	Outlier F_{ST}		LFMM				
		Arlequin	Bayescan	Lat.	Long.	Soil	Temp	Prec.
<i>Regional scale</i>								
a2	Auxin-binding protein	0.240**	0.176**		6.604		7.662	6.212
a23	Pentatricopeptide repeat	0.218**	0.161*	6.196	9.244			
a34	CBF5 ribonucleoprotein	0.448**	0.202**		6.038	4.864	9.229	6.761
a107	Laccase17	0.333**	0.186**		10.356		4.939	
n55	Pentatricopeptide repeat	0.229**	0.171**		6.030		7.376	5.352
<i>Intermediate scale</i>								
a2	Auxin-binding protein	0.233**	0.157*		5.541			
a9	ABC transporter G	0.166*	0.152*	7.426			8.243	8.030
a17	Amino acid transporter 1	0.161*	0.169*	4.416		5.489	5.337	5.025
a34	CBF5 ribonucleoprotein	0.446**	0.204*		6.207			
a63	UDP-glycosyltransferase 88A1	0.235**	0.206**	4.111	9.094			
a107	Laccase17	0.275**	0.216**		11.728			
n7	Chlorophyll-binding protein	0.194**	0.155*	8.433			8.188	8.171
n55	Pentatricopeptide repeat	0.215**	0.150*		4.8678			
<i>Continental scale</i>								
a2	Auxin-binding protein	0.309*	0.270**		6.100			4.624
a9	ABC transporter G	0.235(*)	0.333*		8.114	5.994		10.655
a23	Pentatricopeptide repeat	0.232(*)	0.250*	7.986	4.524	5.532	6.772	
a34	CBF5 ribonucleoprotein	0.355*	0.247*	7.986	4.524		5.612	
a63	UDP-glycosyltransferase 88A1	0.255*	0.306**		6.553		4.245	4.006
a107	Laccase17	0.330**	0.351**	4.012				7.916
ap29	Tubby-like F-box protein	0.239(*)	0.264*		11.601	8.200	6.635	7.631
ap44	PAX-interacting protein	0.307**	0.256*		9.657	4.931		17.171*
ap51	Purple acid phosphatase	0.369**	0.284**	11.604			6.190	7.017
n19	Pectinesterase inhibitor	0.439**	0.289*		8.800			13.776
n55	Pentatricopeptide repeat	0.302*	0.267**		6.151	4.044		4.834
n69	L-asparaginase	0.247*	0.277**		7.191	7.862	5.806	6.472

Bold numbers reflect the strongest associations for each outlier SNP. Boxed values reflect associations with a $|z|$ -score at least 4 larger than all other associations with the same SNP (second highest $|z|$ -score +4 or more). Dark gray boxes reflect SNPs that are significant outliers at all three scales. Light grey boxes reflect SNPs that are significant outliers at two scales. For a full description of the gene functions, refer to Supplementary Information 2 and De Kort *et al.*, 2013. Asterisks refer to the significance level: * $P < 0.05$ and 'substantial' for Arlequin and Bayescan, respectively; ** $P < 0.01$ and 'very strong' for Arlequin and Bayescan, respectively.

$|z|$ -scores, as additional environmental associations are likely the result of spatial autocorrelation resulting in spurious associations (Supplementary information 3). Simulations have shown that tests based on LFMM significantly reduce the rates of false positive associations in highly structured populations (Frichot *et al.*, 2013; De Villedereuil *et al.*, 2014).

Landscape genomics: RDA

RDA is similar to canonical correlation analysis and has the advantage of eliminating redundant information in allele frequencies and their associations with environmental variables, which allows extracting the amount of genetic variance explainable by significant environmental variables (Borcard *et al.*, 2011). The RDA analyses were performed using the Vegan package in R, to reveal how much of the variation in allele frequencies can be explained by spatial vs abiotic variables (Oksanen *et al.*, 2008). Temperature, precipitation and soil variables (abiotic variables), and longitude, latitude and their square-transformed values (spatial variables) were chosen as potential explanatory variables for the observed multivariate genetic structure (Legendre and Fortin, 2010). As raw polynomial terms are highly correlated, we computed the orthogonal polynomials of latitude and longitude using the function 'poly' of the R package stats. We

performed the forward selection procedure as proposed by Borcard *et al.* (2011) to determine which geographical variables were relevant at each spatial scale, using the R package packfor. Allele frequencies were Hellinger transformed prior to RDA (Legendre and Gallagher, 2001). An adjusted R^2 was calculated to correct for the number of explanatory variables and significance was determined by using 1000 permutations (Borcard *et al.*, 2011).

RESULTS

Population structure and genomics

Expected heterozygosity was on average 0.314 (± 0.013), and ranged from 0.219 (Luriano, Tuscany) to 0.346 (De Pinte, Flanders), with highest average values for the populations in Belgium and France (0.307–0.345) and lowest average values for the populations in Sweden and Italy (0.219–0.311) (Table 1). Genetic diversity was higher for SNPs associated with temperature and precipitation (average H_E 0.323 ± 0.036) than for neutral SNPs (average H_E 0.289 ± 0.029) ($F_{1, 3686.6} = 33.31$, $P < 0.001$) (Figure 2). Pairwise F_{ST} values ranged from 0.013 to 0.155 (average 0.064 ± 0.032 ; $P < 0.01$) at the regional scale, from 0.017 to 0.157 (average 0.081 ± 0.029 ; $P < 0.01$) at the intermediate scale and from 0.160 to 0.369 (average 0.239 ± 0.047 ; $P < 0.01$) at the continental scale (Supplementary Information 2).

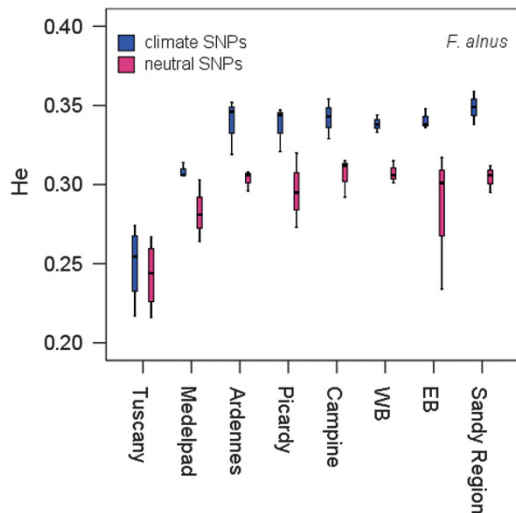


Figure 2 Expected heterozygosity in the different regions, on the set of neutral SNPs and on the set of putatively adaptive SNPs (SNPs associated to temperature and precipitation).

The principal coordinates analysis plot based on all 25 populations showed three main clusters, which correspond to (i) Flanders/Ardennes/Picardy, (ii) Medelpad and (iii) Tuscany (Figure 3; Supplementary Information 4).

At the continental scale, 12 loci (6.6%) were identified as loci putatively under positive selection (Figure 4), and 5 of these were assigned the gene ontology term ‘response to abiotic stimulus’ (Supplementary Information 2, De Kort *et al.*, 2013). Hence, 41.67% of the 12 outliers versus 29.51% (51 SNPs, Supplementary information 2) of all 183 SNPs were putatively associated with the GO term ‘response to abiotic stimulus’. At the intermediate scale, eight putative outlier loci (4.4%) were observed. Finally, five loci (2.7%) putatively under positive selection were detected at the regional scale (Figure 4, Table 2). Four outlier loci putatively under positive selection (2.2%; a2, a34, a107, n55) were found at all three spatial scales (Table 2), of which a2, a34 and a107 were associated with ‘response to abiotic stimulus’ Table 1 (Supplementary information 2). SNP n55 was not associated with an ecologically relevant function. All of the regional outliers and six out of the eight of the intermediate outliers were also found at the continental scale (Table 2).

Landscape genomics: LFMM

At the continental scale, a total of 143 SNPs (78.14%) could be associated to one or more environmental variables with $|z| > 4$ (Supplementary information 2). Temperature was the most significant variable in 17 SNP-environment associations (9.30%; Figure 5), with a mean $|z|$ -score (that is, strength of SNP associations) of 7.844. Precipitation could be associated to 47 SNPs (25.68%; Figure 5), with a mean $|z|$ -score of 6.645. Soil was the most significant variable in 13 SNP-environment associations (7.10%; Figure 5), with a mean $|z|$ -score of 5.678. Finally, latitude and longitude could be associated to 23 (12.57%) and 43 SNPs (23.50%), respectively (Figure 5). All 12 loci that were suggested as outlier loci under positive selection at the continental scale were associated with one or more abiotic variables (Table 2). More specifically, precipitation (10 associations), temperature (6 associations) and soil (6 associations) were correlated with outlier allele frequencies. The strongest associations were between precipitation and two SNPs, with locus n19 showing a striking pattern

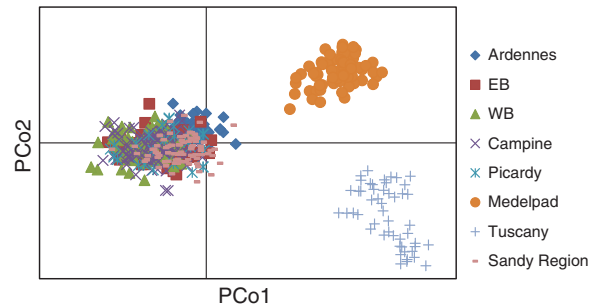


Figure 3 Principal coordinates analysis (GENALEX 6.4) of genetic distances among all populations. Samples were assigned the names of their corresponding region. PCo1 and PCo2 explain, respectively, 34.92% and 22.05% of the variation. See Supplementary Information 4 for genetic structure at the other scale levels using populations as labels. At the intermediate scale, PCo1 and PCo2 explain, respectively, 31.13% and 20.55% of the genetic variation, and at the regional scale, PCo1 and PCo2 explain respectively 34.10% and 15.19% of the variation. See Supplementary Information 5 for the genetic structure for all combinations of the first five PCo axes using regions as labels.

when considering the BAYESCAN and LFMM results combined (Figure 6, Table 2). We previously performed a Blastx of the n19 contig sequence against the *Arabidopsis thaliana* genome, and annotation of the corresponding *Arabidopsis* homolog revealed a very confident association with cell wall modification and pectinesterase activity (1.0E-42; AGI code AT2G26440, De Kort *et al.*, 2013).

At the intermediate scale, a total of 91 SNPs (49.73%) were associated with one or more environmental variables (Supplementary Information 2). Temperature was the most significant variable in 39 SNP-environment associations (6.01%; Figure 5), with a mean $|z|$ -score (that is, strength of association) of 5.639. Precipitation was associated with 10 SNPs (5.46%; Figure 5), with a mean $|z|$ -score of 5.520. Soil was the most significant variable in two SNP-environment associations (1.09%; Figure 5), with a mean $|z|$ -score of 4.269. Finally, latitude and longitude were the most important variables in 21 (11.48%) and 47 SNP-environment associations (25.68%) respectively (Figure 5). A total of three out of eight outlier loci were associated with ‘temperature’ and ‘precipitation’ according to the LFMM at the intermediate scale (Table 2).

At the regional scale, a total of 73 SNPs (39.89%) could be associated to one or more environmental variables (Supplementary information 2). Temperature was the most significant variable in 16 SNP-environment associations (8.74%; Figure 5), with a mean $|z|$ -score of 6.618. Precipitation was associated with 12 SNPs (6.56%; Figure 5), with a mean $|z|$ -score of 5.977. Soil was the most significant variable in two associations (1.09%; Figure 5), with a mean $|z|$ -score of 5.567. Finally, latitude and longitude could be associated to 13 (7.10%) and 38 (20.77%) SNPs, respectively (Figure 5). Temperature and/or precipitation were associated with 4 out of 5 loci that were suggested as outlier loci under positive selection (Table 2).

Landscape genomics: RDA

The redundancy analyses revealed strong non-linear isolation-by-distance, with latitude² explaining most of the genetic variation at the continental scale ($F=10.973$, $P<0.001$), whereas longitude and longitude² played an important role at the intermediate ($F=3.940$, $P<0.001$ and $F=2.519$, $P<0.001$ respectively) and regional scales ($F=5.579$, $P<0.001$ and $F=1.804$, $P<0.01$) (Figure 7, Supplementary information 2).

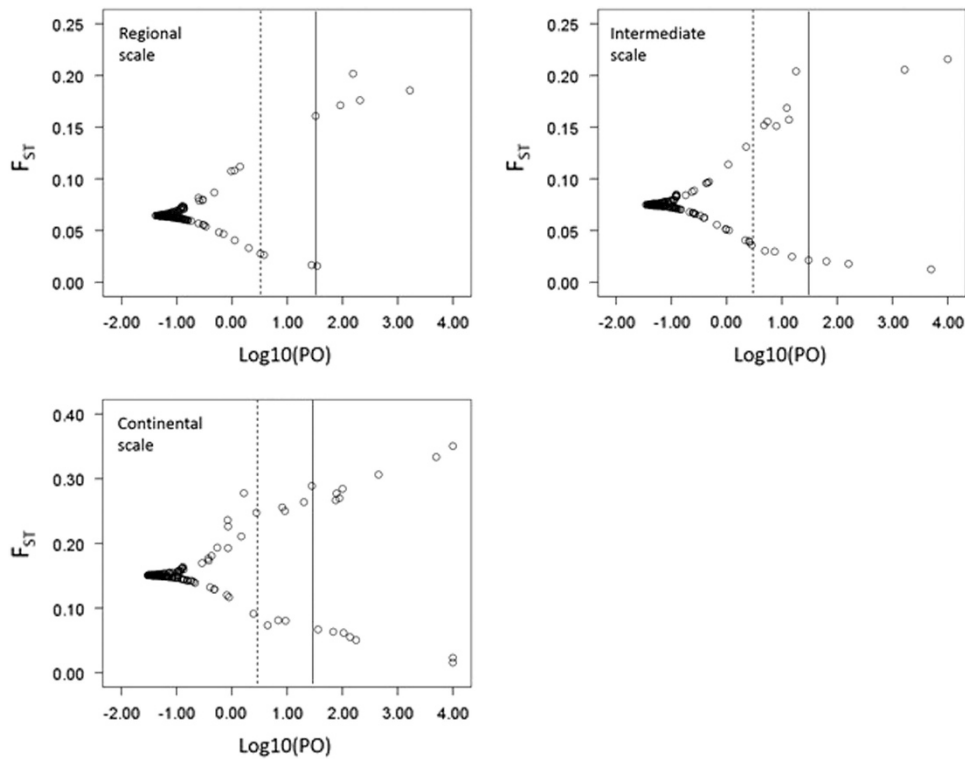


Figure 4 Outlier detection at the three spatial scales showing empirical F_{ST} values as a function of the $\log_{10}(PO)$, resulting from the Bayesian model implemented in BAYESCAN (Supplementary information 2). The dashed line represents the $\log_{10}(PO)$ of 0.5, whereas the solid line indicates the $\log_{10}(PO)$ of 1.5. \log_{10} values of the $PO > 0.5$ and 1.5 were taken as 'substantial' and 'very strong' evidence for selection, respectively. Positive outliers (under positive selection) have F_{ST} values that were higher than expected under neutrality, whereas negative outliers (under balancing selection) have F_{ST} values lower than neutral expectations.

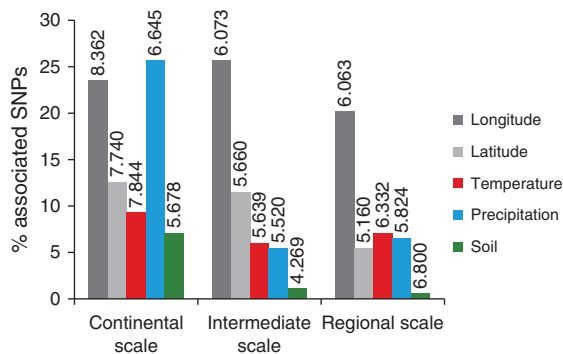


Figure 5 The percentage of SNPs that were associated with an environmental variable. To exclude spurious effects of correlated variables, only the most significant association for each SNP was counted. Numbers represent the average per-SNP $|z|$ -score (i.e., strength of selection).

Together, the geographical variables explained 31.60, 29.73 and 46.53% of the variation in allele frequencies at the continental, intermediate and regional scale respectively (Figure 7, Supplementary Information 2). Temperature was the most important explaining abiotic variable at the continental ($R^2_{adj} = 13.53\%$, $F = 3.805$, $P < 0.001$) and the regional scale ($R^2_{adj} = 4.68\%$, $F = 1.695$, $P < 0.05$), whereas precipitation explained most of the genetic variation at the intermediate scale ($R^2_{adj} = 11.11\%$, $F = 3.397$, $P < 0.001$). Soil had no significant effect on the variation in allele frequencies (Figure 7).

DISCUSSION

Both the population genomic and the landscape genomic approaches indicated a more distinct signature of selection with increasing geographical scale, where more outliers and associations with climate were observed. At the continental scale, large environmental differences and non-linear isolation-by-distance very likely decrease the effective gene flow among populations of *F. alnus* subsp. *alnus*, thereby increasing the frequency of beneficial alleles in local environments. At the regional scale, on the other hand, the limited abiotic differentiation may explain the relatively small number of outliers. Nevertheless, the occurrence of some outliers suggest that natural selection may effect genomic variation, even at these small scales where gene flow is expected to be extensive. In line with the outlier detection results, the number of SNP associations with environmental variables increased with increasing geographical scale, while the strength and type of selection pressure varied among scales. At the continental scale, precipitation (reflected by the number of SNP associations) and temperature (reflected by the strength of SNP associations) were identified as putatively important drivers of adaptive differentiation in *F. alnus* subsp. *alnus*. At the regional scale on the other hand, SNPs were mainly correlated with temperature (reflected by both the number and strength of SNP associations). Finally, high levels of expected heterozygosity, in particular for the putatively adaptive loci, suggest high potential to cope with climate change.

Evidence for selection at all three spatial scales

In line with previous studies on the tree species *A. glutinosa* (Cox *et al.*, 2011) and the fish species *Salmo trutta* (Meier *et al.*, 2011), the

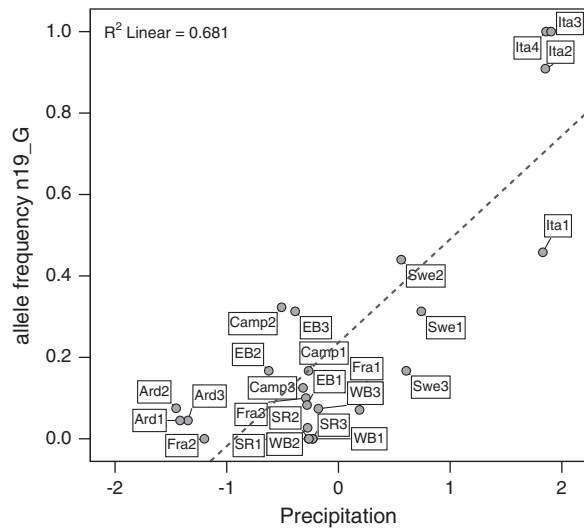


Figure 6 Comparison between the population-specific values of precipitation and allele (G) frequency of SNP n19. Negative values reflect high precipitation and vice versa. Population codes are explained in Table 2.

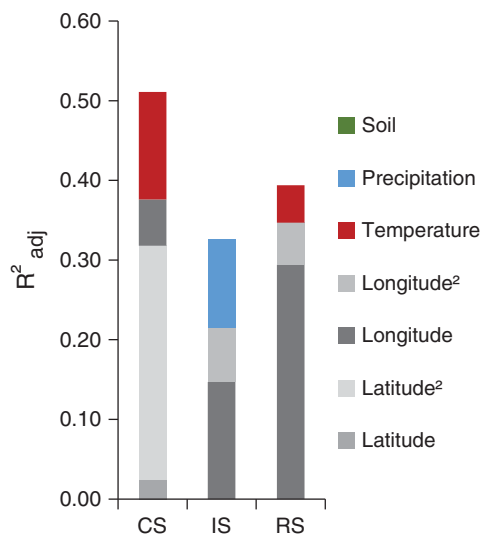


Figure 7 The percentage of genetic variation (R^2_{adj}) that could be explained by environmental variables at the continental scale (CS), the intermediate scale (IS) and the regional scale (RS).

number of outliers increased with increasing geographic scale (Figure 4, Table 2), from 2.7% outlier SNPs (5 outliers) at the regional scale, to 4.4% (8 outliers) at the intermediate scale and 6.6% (12 outliers) at the continental scale, suggesting changes in the nature of the selective pressures depending on the distance among the populations. Although restricting the study to SNPs derived only from coding regions may have overestimated the proportion of outlier loci to some extent, outlier detection studies in other tree species revealed similar proportions of outliers at a large scale (> 400 km), for example in *Picea glauca* (5.5% outliers, $F_{ST}=0.006$, genic SNPs) (Namroud *et al.*, 2008), *A. glutinosa* (3.7% outliers, $F_{ST}=0.057$, amplified fragment length polymorphisms; and 2.86%, $F_{ST}=0.01$, SNPs) (Cox *et al.*, 2011; De Kort *et al.*, 2014), and *Quercus robur* and *Q. petraea* (4% and 5% outliers resp., $F_{ST}=0.013$ and 0.015 resp., genic SNPs) (Guichoux *et al.*, 2013). Nevertheless, many of these

findings should be interpreted with considerable caution due to the use of candidate loci (except for those of Cox *et al.*, 2011, De Kort *et al.*, 2014), which tend to overestimate neutral distributions and therefore underestimate the true percentage of loci under selection, the latter being expected to be high when using loci related to adaptive functions or traits. The percentage of outlier loci we detected at a similar geographical scale (> 400 km, 6.6%) even slightly exceeds this range of outlier proportions, likely because of (i) stronger selection driven by environmental differences between populations, and (ii) limited gene flow between Italian, Swedish and Belgian/French populations as a consequence of geographical isolation by mountain and sea barriers combined with pollination by insects. Indeed, *F. alnus* subsp. *alnus* is insect pollinated, which may be featured by limited gene migration between populations, therefore potentially facilitating divergent selection, whereas the studies mentioned earlier all examined signatures of selection in wind pollinated species. These barriers are probably responsible for the unimodal genetic response to latitude revealed by the RDA analyses. Accordingly, pairwise F_{ST} values at the continental scale were rather high with an average of $0.239 (\pm 0.047)$.

The detection of five outlier loci at the regional scale, where little population structure and high genetic diversity were observed (Figures 2 and 3, Supplementary information 5, Table 1), suggests that low levels of adaptive differentiation may have emerged in the presence of strong among-population gene flow, strengthening the notion that local selection pressures can override the homogenizing effects of gene exchange (Lenormand, 2002; Alleaume-Benharira *et al.*, 2006; Neale and Ingvarsson, 2008; Yeaman and Otto, 2011; Richardson *et al.*, 2014).

Environmental agents of selection at the continental scale

Both temperature and precipitation seemed important selective agents driving adaptive differentiation in *F. alnus* subsp. *alnus* (9 outliers, Table 2), with the number and strength of SNP-climate associations differing among geographical scales (Table 2). Drought stress induced by different temperature and precipitation regimes may indeed facilitate selection of specific genetic variants, as the species naturally occurs across a variety of soil moisture levels, ranging from ephemeral to wet ecosystems (personal observations; Godwin, 1943; Fiedler and Landis, 2012; see also Zang *et al.*, 2014). Interestingly, whereas most SNP-climate associations at the continental scale involved precipitation (25.68% SNP-precipitation associations vs 9.30% SNP-temperature associations, Figure 5), the most significant associations involved temperature (mean $|z|$ -score of 7.844 and R^2_{adj} of 12.24%, Figures 5 and 6). This result may suggest a highly polygenic nature of traits responding to precipitation, whereas temperature may affect traits with a less complex genetic architecture.

Although we found strong neutral genetic differences between the Italian, Swedish and French/Belgian populations (Figure 2), it seems unlikely that neutral genetic divergence in the absence of gene flow has mimicked the signature of selection. First, the RDA analyses took both linear and quadratic geographical structure into account, whereas the LFMM approach accounted for the different genetic clusters. Second, the association of several outlier loci with climate was remarkable, especially at the continental scale, and did not unequivocally correspond to geographic distances.

Furthermore, correlated co-ancestry and demographic processes may increase the number of false positive outliers considerably (Excoffier *et al.*, 2009; Mosca *et al.*, 2012; Bierne *et al.*, 2013). However, the detection of the same outliers at different geographical scales, together with the association of most of our outliers with ecologically relevant functions and environmental variables, supports selective divergence of the corresponding loci (Limborg *et al.*, 2012;

Buehler *et al.*, 2012; Alberto *et al.*, 2013; Steane *et al.*, 2014). Overall, these findings also render it unlikely that our outlier SNPs represent sequence assembling errors, which can be introduced during the *de novo* sequence assembly preceding SNP discovery in non-model organisms (Da Silva *et al.*, 2014). A literature search on the gene homologs containing the outlier SNPs (Table 2) uncovered some patterns that corroborate the ecological relevance of their associations with specific environmental variables. First, locus n19 was annotated to an *A. thaliana* homolog that is known to affect pectinesterase activity in cell wall modifications, mediating plant responses to drought (Wang *et al.*, 2011; Verelst *et al.*, 2013). Correspondingly, this SNP correlated exceptionally well with patterns of precipitation, over a gradient of very little rainfall in Italy over moderate levels in Sweden and Flanders, to high levels of precipitation in the Ardennes (Figure 6). Second, SNP a107, which appeared to be an outlier at all geographical scales, was putatively associated with a laccase locus (*lac17*, Table 2). Laccases are mainly known for their role in the lignification process and therefore especially relevant for biotechnological purposes (Berthet *et al.*, 2011; Guerra *et al.*, 2013; Lu *et al.*, 2013). Moreover, because lignin quantity and quality affects cell wall fragility, laccase activity can affect tree growth and tolerance to various biotic and abiotic stresses (Peter and Neale, 2004). Although the role of laccase in responses to drought stress is hitherto poorly understood (but likely related to increased lignin production, Yoshimura *et al.*, 2008; Moura *et al.*, 2010), we observed a clear association between the allele frequency of locus a107 and precipitation at the continental scale. In concert with the exceptional genetic differentiation among populations of *F. alnus* subsp. *alnus* (Table 2), this finding denotes locus a107 as a candidate target for natural selection induced by drought.

Although the occurrence of the same outliers at different spatial scales may reflect excellent candidates for further research, selective loci may be different at different spatial scales because of variable selection pressures across scales, or because of correlations among loci underlying the same phenotypic trait (Nosil *et al.*, 2009; Alberto *et al.*, 2013). We therefore cannot exclude the potential adaptive value of outliers that do not appear consistently across all scales.

Environmental agents of selection at the intermediate and regional scale

At the intermediate scale, it is unclear which climatic variable is the most important selective agent (similar number of SNP-climate associations, but precipitation explains significantly more of the genetic variance according to the RDA analyses, Figures 5 and 6). This may be due to the strong correlation among the two variables at this scale (Spearman's $\rho=0.955$, Supplementary Information 3), resulting in spurious associations. At the regional scale, temperature imposes a stronger selective pressure than precipitation as putative selective agent, both through the number and the strength of the SNP-temperature associations (Figures 5 and 6). This finding may suggest that at a very small scale, only loci underlying adaptive traits with a less complex genetic architecture, putatively representing large-effect loci, may be able to override the homogenizing effects of gene flow. Nevertheless, a full genome scan is required to make final conclusions on the relation between environmental drivers of selection and the genetic architecture of the SNPs involved.

The dominating effect of temperature at the regional scale (Table 2) is in line with temperature differences following a longitudinal cline, with slightly higher minimum winter temperatures in the more Atlantic populations in the west (ca 0.5 °C) than the more continental populations in the east of Flanders (ca -0.5 °C). In parallel with its

association with temperature at the regional scale, outlier SNP a2 was annotated as an auxin-binding protein (Table 2). Auxin is an important plant hormone, modulating many aspects of plant growth, development and responses to biotic and abiotic stress (Wolters and Jürgens, 2009; Peleg and Blumwald, 2011). Variation at loci underlying auxin expression may thus have profound fitness consequences, as has been demonstrated in the context of temperature stress (for example, Swindell *et al.*, 2007; Lee *et al.*, 2014), plant-pathogen interactions (Kazan and Manners, 2009) and symbiotic interactions (for example, Reddy *et al.*, 2006). We correspondingly observed a significant association between SNP a2 and temperature at the regional scale (Table 2). Natural adaptive variation in auxin genes has also been demonstrated in several other tree species, including *Picea glauca* (Namroud *et al.*, 2008) and *Populus trichocarpa* (McKown *et al.*, 2014), stressing the role of micro-evolution in driving genetic differences in auxin levels among populations.

Genetic diversity and adaptive potential

The observed high heterozygosity in *F. alnus* subsp. *alnus* populations can be explained by the use of putatively adaptive candidate loci, which are expected to considerably increase the probability of finding signatures of selection. Correspondingly, we found significantly higher levels of expected heterozygosity at putatively adaptive SNPs than at the neutral SNPs. Allelic variation may be higher at adaptive loci rather than at neutral loci for several reasons, and we accordingly observed higher levels of genetic diversity at the loci that were associated with climate than the loci not associated with climate (Figure 2). First, climatic variation in time may occur within the life span of a tree, resulting in older trees bearing the alleles adapted to a former climate together with younger trees that were selected for alleles adapted to the new climate. Given the very strong selection during the juvenile stage of each tree generation (Petit and Hampe, 2006; Kremer *et al.*, 2012), long life spans may thus allow accumulation of selective alleles within populations to a greater extent than neutral alleles. Second, the complex genetic architecture underlying the adaptive traits may support high allelic diversity within populations for the loci underlying these traits (that is, adaptive loci). When many small-effect loci affect an adaptive trait, interactions among loci may prevent genetic drift of alleles that are neutral or slightly deleterious in the local environment but beneficial in a different environment (conditional neutrality and antagonistic pleiotropy resp.) (Latta, 1998; Colautti *et al.*, 2012; Alonso-Blanco and Méndez-Vigo, 2014). Hence, although divergent selection reduces population genetic variation at adaptive loci, at least over one generation, it may not negatively affect the overall population adaptive diversity, when considering the adaptive variation across generations. Our findings may therefore suggest substantial adaptive genetic variation that may provide resilience to climate change in *F. alnus* subsp. *alnus* (see also Jump *et al.*, 2009). Nevertheless, the current generation of trees may harbor the genetic variation that is associated with former demographic and climatic conditions, but not with future climate, which is predicted to be substantially different from historical conditions (Moritz and Agudo, 2013). Moreover, ongoing habitat loss and fragmentation may reduce population genetic diversity and increase inbreeding in many tree species, which may threaten the ability to cope with the synergistic effects of habitat reduction and climate change (Brook *et al.*, 2008; Hof *et al.*, 2011; Meier *et al.*, 2012; Vranckx *et al.*, 2012).

CONCLUSION

Whereas the majority of population and landscape genomic studies involve wind-pollinated tree species, our study demonstrates the

genetic effects of divergent selection across the genome of a non-model, insect-pollinated tree species. Based on the number of outliers, the number of associations between loci and climate, and the strength of environmental effects on allelic distributions, we were able to assess the molecular signature of selection at different spatial scales. Although patterns of divergent selection imposed by temperature and precipitation were strongest at the continental scale, their occurrence at the intermediate and regional scales, where high intra-population diversity and limited genetic clustering were observed, may suggest a major role of among-population gene exchange in safeguarding evolutionary potential by supporting the conservation of adaptive genetic variation. Taken together, we conclude that the integration of complementary population and landscape genomic approaches using annotated SNPs rendered evidence for considerable selection imposed by temperature and precipitation at various spatial scales in *F. alnus*.

DATA ARCHIVING

The SNPs used in this study are part of a larger set of previously discovered SNPs, which has been submitted to GenBank and can be accessed using the following accession number PRJNA203129. <http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA203129>. Sequences of the 250 SNPs as well as the GENALEX file containing the successfully genotyped SNPs (183) are accessible on Dryad using the following 10.5061/dryad.r12jk.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank K Van Acker, S Maccherini, G Decocq and BG Jonsson for assisting with sample collection. HDK holds a doctoral funding from the agency for Innovation by Science and Technology (IWT). The genetic analyses were funded by a grant of the Agentschap voor Natuur en Bos (ANB) to JM.

AUTHOR CONTRIBUTIONS

HDK performed the research and wrote the manuscript. KV, JM, KVM and OH assisted with the analyses and commented on the manuscript. JM also enabled the genotyping-by-sequencing by organizing financial support.

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

Alonso-Blanco C, Méndez-Vigo B (2014). Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. *Curr Opin Plant Biol* **18**: 37–43.

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

Alleaume-Benharira M, Pen IR, Ronce O (2006). Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *J Evol Biol* **19**: 203–215.

Baird NA, Etter PD, Atwood TS (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**: e3376.

Berthet S, Demont-Caulet N, Pollet B, Bidzinski P, Cézard L, Le Bris P *et al.* (2011). Disruption of Laccase4 and 17 results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* **23**: 1124–1137.

Bierne N, Roze D, Welch JJ (2013). Pervasive selection or is it...? why are F ST outliers sometimes so frequent? *Mol Ecol* **22**: 2061–2064.

Borcard D, Gillet F, Legendre P (2011). *Numerical Ecology with R*. Springer: New York, p 306.

Bradbury D, Smithson A, Krauss SL (2013). Signatures of diversifying selection at EST-SSR loci and association with climate in natural *Eucalyptus* populations. *Mol Ecol* **22**: 5112–5129.

Breed MF, Stead MG, Ottewell KM, Gardner MG, Lowe AJ (2012). Which provenance and where? Seed sourcing strategies for revegetation in a changing environment. *Conserv Genet* **14**: 1–10.

Brook BW, Sodhi NS, Bradshaw CJA (2008). Synergies among extinction drivers under global change. *Trends Ecol Evol* **23**: 453–460.

Buehler D, Poncet BN, Holderegger R (2012). An outlier locus relevant in habitat-mediated selection in an alpine plant across independent regional replicates. *Evol Ecol* **27**: 285–300.

Colautti RI, Lee C-R, Mitchell-Olds T (2012). Origin, fate, and architecture of ecologically relevant genetic variation. *Curr Opin Plant Biol* **15**: 199–204.

Cox K, Broeck Vanden, Van Calster H, Mergeay J (2011). Temperature-related natural selection in a wind-pollinated tree across regional and continental scales. *Mol Ecol* **20**: 2724–2738.

Cuppen E (2007). Genotyping by allele-specific amplification (KASPar). *CSH Protoc* **2007**: 172–173.

Dabney A, Storey JD, Warnes GR (2011). Qvalue: Q-value estimation for false discovery rate control. R package version 1.26.0.

Da Silva AG, Barendse W, Kijas JW, Barris WC, McWilliam S *et al.* (2014). SNP discovery in nonmodel organisms: strand bias and base-substitution errors reduce conversion rates. *Mol Ecol* **10.1111/1755-0998.12343**.

De Kort H, Vandepitte K, Bruun HH *et al.* (2014). Landscape genomics and a common garden reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Mol Ecol* **23**: 4709–4721.

De Kort H, Vandepitte K, Mergeay J, Honnay O (2013). Isolation, characterization and genotyping of single nucleotide polymorphisms in the non-model tree species *Frangula alnus* (Rhamnaceae). *Conserv Genet Resour* **6**: 267–269.

De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J *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.

De Villemereuil P, Fricot E, Bazin E, François O, Gaggiotti OE (2014). Genome scan methods against more complex models: when and how much should we trust them? *Mol Ecol* **23**: 2006–2019.

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

Eckert AJ, Wegrzyn JL, Liechty JD, Lee JM, Cumbie WP, Davis JM *et al.* (2013). The evolutionary genetics of the genes underlying phenotypic associations for loblolly pine (*Pinus taeda*, Pinaceae). *Genetics* **195**: 1353–1372.

Eklblom R, Galindo J (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1–15.

Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011). Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. *PLoS One* **6**: e18561.

Excoffier L, Hofer T, Foll M (2009). Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285–298.

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

Fiedler AK, Landis DA (2012). Biotic and abiotic conditions in Michigan prairie fen invaded by glossy buckthorn (*Frangula alnus*). *Nat Area J* **32**: 41–53.

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.

Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011). Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity* **106**: 404–420.

Fricot E, Schoville SD, Bouchard G, François O (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* **30**: 1687–1699.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012). Harnessing genomics for delineating conservation units. *Trends Ecol Evol* **27**: 489–496.

Godwin H (1943). *Frangula alnus* Miller (*Rhamnus frangula* L.). *J Ecol* **31**: 77–92.

Grivet D, Sebastiani F, Alía R, Bataillon T, Torre S, Zabal-Aguirre M *et al.* (2011). Molecular footprints of local adaptation two mediterranean conifers. *Mol Biol Evol* **28**: 101–116.

Guerra FP, Wegrzyn JL, Sykes R, Davis MF, Stanton BJ, Neale DB *et al.* (2013). Association genetics of chemical wood properties black poplar (*Populus nigra*). *New Phytol* **197**: 162–176.

Guichoux E, Garnier-Géré P, Lagache L, Lang T, Boury C, Petit RJ *et al.* (2013). Outlier loci highlight the direction of introgression in oaks. *Mol Ecol* **22**: 450–462.

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

Hampe A (2004). Extensive hydrochory uncouples spatiotemporal patterns of seedfall and seedling recruitment in a "bird-dispersed" riparian tree. *J Ecol* **92**: 797–807.

Hampe A (2005). Fecundity limits in *Frangula alnus* (Rhamnaceae) relict populations at the species' southern range margin. *Oecologia* **143**: 377–386.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* **25**: 1965–1978.

Hof C, Levinsky I, AraÚjo MB, Rahbek C (2011). Rethinking species' ability to cope with rapid climate change. *Global Change Biol* **17**: 2987–2990.

Jeffreys H (1961). *Theory of Probability*. 3rd edn. London: Oxford University Press, 432.

Jones MR, Forester BR, Teufel AI, Adams RV, Anstett DN, Goodrich BA *et al.* (2013). Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution* **67**: 3455–3468.

Joost S, Vuilleumier S, Jensen JD, Schoville S, Leempoel K, Stucki S *et al.* (2013). Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. *Mol Ecol* **22**: 3659–3665.

- Jump AS, Hunt JM, Martínez-Izquierdo JA, Penuelas J (2006). Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Mol Ecol* **15**: 3469–3480.
- Jump AS, Marchant R, Peñuelas J (2009). Environmental change and the option value of genetic diversity. *Trends in plant science* **14**: 51–58.
- Kawecki TJ, Ebert D (2004). Conceptual issues in local adaptation. *Ecol Lett* **7**: 1225–1241.
- Kazan K, Manners J (2009). Linking development to defence: auxin in plant-pathogen interactions. *Trends Plant Sci* **14**: 373–382.
- Keller SR, Soolanayakanahally RY, Guy RD, Silim SN, Olson MS, Tiffin P *et al.* (2011). Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, *Populus balsamifera* L. (Salicaceae). *Am J Botany* **98**: 99–108.
- Kremer A, Le Corre V (2012). Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* **108**: 375–385.
- Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R *et al.* (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol Lett* **15**: 378–392.
- Latta RG (1998). Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *Am Nat* **151**: 28.
- 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.
- Lee H, Jung L, Llorca LC, Kim SG, Lee S, Baldwin IT *et al.* (2014). FCA mediated thermal adaptation of stem growth by attenuating auxin action in *Arabidopsis*. *Nat Commun* **5**: 5473.
- Legendre P, Fortin M-J (2010). Comparison of the mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol Ecol Resour* **10**: 831–844.
- Legendre P, Gallagher E (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**: 271–280.
- Lenormand T (2002). Gene flow and the limits to natural selection. *Trends Ecol Evol* **17**: 183–189.
- Limborg MT, Helyar SJ, De Bruyn M, Taylor MI, Nielsen EE, Ogden R *et al.* (2012). Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Mol Ecol* **21**: 3686–3703.
- Liu J, Huang S, Sun M, Liu S, Liu Y, Wang W *et al.* (2012). An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods* **8**: 34.
- Lu S, Li Q, Wei H, Tunlaya-Anukit S, Kim H, Liu J *et al.* (2013). Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in *Populus trichocarpa*. *Proc Natl Acad Sci USA* **110**: 10848–10853.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003). The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* **4**: 981–994.
- Manel S, Holderegger R (2013). Ten years of landscape genetics. *Trends Ecol Evol* **28**: 614–621.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* **18**: 189–197.
- McKown AD, Klápště J, Guy RD (2014). Genome-wide associations implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytol* **203**: 535–553.
- Medan D (1994). Reproductive biology of *Frangula alnus* Miller in Southern Spain. *Plant Syst Evol* **193**: 173–186.
- Meier K, Hansen MM, Bekkevold D, Skaala Ø, Mensberg K-LD (2011). An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta* L.): footprints of selection at microsatellite DNA loci. *Heredity* **106**: 488–499.
- Meier ES, Lischke H, Schmatz DR, Zimmermann NE (2012). Climate, competition and connectivity affect future migration and ranges of European trees. *Global Ecol Biogeography* **21**: 164–178.
- Morin PA, Luikart GA, Wayne RKSnp workshop group (2004). SNPs in ecology, evolution and conservation. *Trends Ecol Evol* **19**: 208–216.
- Moritz C, Agudo R (2013). The future of species under climate change: resilience or decline? *Science* **341**: 504–508.
- Mosca E, Eckert AJ, Liechty JD, Wegrzyn JL, La Porta N, Vendramin GG *et al.* (2012). Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evol Appl* **5**: 762–775.
- Moura JCMS, Bonin CAV, Viana JDof, Dornelas MC, Mazzafera P (2010). Abiotic and biotic stresses and changes in the lignin content and composition in plants. *J Integ Plant Biol* **52**: 360–376.
- Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J (2008). Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol Ecol* **17**: 3599–3613.
- Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol* **22**: 2841–2847.
- Narum SR, Hess JE (2011). Comparison of F(ST) outlier tests for SNP loci under selection. *Mol Ecol Resour* **11**: 184–194.
- Neale DB, Ingvarsson PK (2008). Population, quantitative and comparative genomics of adaptation in forest trees. *Curr Opin Plant Biol* **11**: 149–155.
- Nielsen R, Hubisz MJ, Clark AG (2004). Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics* **168**: 2373–2382.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009). Divergent selection and heterogeneous genomic divergence. *Mol Ecol* **18**: 375–402.
- Oksanen J, Kindt R, Legendre P (2008). The Vegan Package version 1.15-1. Community ecology package.
- Olson MS, Levens N, Soolanayakanahally RY, Guy RD, Schroeder WR, Keller SR *et al.* (2013). The adaptive potential of *Populus balsamifera* L. to phenology requirements in a warmer global climate. *Mol Ecol* **22**: 1214–1230.
- Paterson AH, Freeling M, Tang H, Wang X (2010). Insights from the comparison of plant sequences. *Annu Rev Plant Biol* **61**: 349–372.
- Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.
- Peleg Z, Blumwald E (2011). Hormone balance and abiotic stress tolerance in crop plants. *Curr Opin Plant Biol* **14**: 290–295.
- Peter G, Neale D (2004). Molecular basis of the evolution of xylem lignification. *Curr Opin Plant Biol* **7**: 737–742.
- Petit RJ, Hampe A (2006). Some evolutionary consequences of being a tree. *Annu Rev Ecol Syst* **37**: 187–214.
- 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. *Mol Ecol* **20**: 1702–1716.
- Reddy SM, Hitchin S, Melayah D (2006). The auxin-inducible CH3 homologue Pp-GH3.16 is downregulated in *Pinus pinaster* root systems on ectomycorrhizal symbiosis establishment. *New Phytol* **172**: 391–400.
- Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol Evol* **29**: 165–176.
- Rosenblum EB, Novembre J (2007). Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. *J Heredity* **98**: 331–336.
- Schoville SD, Bonin A, François O *et al.* (2012). Adaptive genetic variation on the landscape: methods and cases. *Annu Rev Ecol Syst* **43**: 23–43.
- Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW *et al.* (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol Ecol Resour* **11**: 1–8.
- Sork VL, Aitken SN, Dyer RJ *et al.* (2013). Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genet Genomes* **9**: 901–911.
- Steane DA, Potts BM, Prober SM, Prober SM, Stock WD, Vaillancourt RE *et al.* (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Mol Ecol* **23**: 2500–2513.
- Swindell W, Huebner M, Weber A (2007). Plastic and adaptive gene expression patterns associated with temperature stress in *Arabidopsis thaliana*. *Heredity* **99**: 143–150.
- Vandepitte K, Honnay O, Mergeay J, Breyne P, Roldán-Ruiz I, De Meyer T *et al.* (2013). SNP discovery using paired-end RAD-tag sequencing on pooled genomic DNA of *Sisymbrium austriacum* (Brassicaceae). *Mol Ecol Resour* **13**: 269–275.
- Vandepitte K, De Meyer T, Helsen K, van Acker K, Roldán-Ruiz I, Mergeay J *et al.* (2014). Rapid genetic adaptation precedes the spread of an exotic plant species. *Mol Ecol* **23**: 2157–2164.
- Verelst W, Bertolini E, De Bort S, Vandepoel K, Demeulenaere M, Pè ME *et al.* (2013). Molecular and physiological analysis of growth-limiting drought stress in *Brachypodium distachyon* leaves. *Mol Plant* **6**: 311–322.
- Volis S, Yakubov B, Shulgina I, Ward D, Mendlinger S (2005). Distinguishing adaptive from nonadaptive genetic differentiation: comparison of Q(ST) and F(ST) at two spatial scales. *Heredity* **95**: 466–475.
- Vranckx G, Jacquemyn H, Muys B, Honnay O (2012). Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. *Conserv Biol* **26**: 228–237.
- Wang D, Pan Y, Zhao X, Zhu L, Fu B, Li Z *et al.* (2011). Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. *BMC Genomics* **12**: 149.
- Weinig C, Ewers BE, Welch SM (2014). Ecological genomics and process modeling of local adaptation to climate. *Curr Opin Plant Biol* **18**: 66–72.
- Wolters H, Jürgens G (2009). Survival of the flexible: hormonal growth control and adaptation in plant development. *Nat Rev Genet* **10**: 305–317.
- Yeaman S, Otto SP (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**: 2123–2129.
- Yoshimura K, Masuda A, Kuwano M, Yokota A, Akashi K (2008). Programmed proteome response for drought avoidance/tolerance in the root of a C3-xerophyte (wild watermelon) under water deficits. *Plant Cell Physiol* **49**: 226–241.
- Zang C, Hartl-Meier C, Dittmar C, Rothe A, Menzel A (2014). Patterns of drought tolerance in major European forest trees: climatic drivers and levels of variability. *Global Change Biol* **20**: 3767–3779.

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