MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3824–3835 doi: 10.1111/j.1365-294X.2010.04716.x

Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*

S. MANEL,*† B. N. PONCET,* P. LEGENDRE,‡ F. GUGERLI§ and R. HOLDEREGGER§

**Laboratoire d'Écologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, 2233 Rue de la Piscine, 38041 Grenoble Cedex 09, France, †Laboratoire Population Environnement Développement, UMR 151 UP/IRD, Université de Provence, 3 place Victor Hugo, 13331 Marseille Cedex 03, France, ‡Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7, §WSL Swiss Federal Research Institute, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland*

Abstract

A major challenges facing landscape geneticists studying adaptive variation is to include all the environmental variables that might be correlated with allele frequencies across the genome. One way of identifying loci that are possibly under selection is to see which ones are associated with environmental gradient or heterogeneity. Since it is difficult to measure all environmental variables, one may take advantage of the spatial nature of environmental filters to incorporate the effect of unaccounted environmental variables in the analysis. Assuming that the spatial signature of these variables is broad-scaled, broad-scale Moran's eigenvector maps (MEM) can be included as explanatory variables in the analysis as proxies for unmeasured environmental variables. We applied this approach to two data sets of the alpine plant *Arabis alpina*. The first consisted of 140 AFLP loci sampled at 130 sites across the European Alps (large scale). The second one consisted of 712 AFLP loci sampled at 93 sites (regional scale) in three mountain massifs (local scale) of the French Alps. For each scale, we regressed the frequencies of each AFLP allele on a set of eco-climatic and MEM variables as predictors. Twelve (large scale) and 11% (regional scale) of all loci were detected as significantly correlated to at least one of the predictors ($R^2_{\mathrm{adj}} > 0.5$), and, except for one massif, 17% at the local scale. After accounting for spatial effects, temperature and precipitation were the two major determinants of allele distributions. Our study shows how MEM models can account for unmeasured environmental variation in landscape genetics models.

Keywords: allele distribution model, genome scan, landscape genetics, local adaptation, principal coordinates analysis of neighbour matrices

Received 18 November 2009; revision received 22 March 2010; accepted 29 April 2010

Introduction

Studying adaptive genetic variation in natural environments is a major objective of landscape genetics (Holderegger *et al.* 2008; Manel *et al.* 2010). Genome scans are widely used to detect loci linked to genomic regions of adaptive relevance (Bonin *et al.* 2006; Oetjen & Reusch 2007; Bonin 2008; Herrera & Bazaga 2008). One way of identifying loci that might be under selection is to see which ones are correlated with environmental gradients using allele distribution models (Joost *et al.* 2007; Holderegger *et al.* 2008; Manel & Segelbacher 2009). The basic assumption of allele distribution models is that natural selection along environmental gradient or heterogeneity generates gradual changes (i.e. clinal variation) in allele frequencies at loci linked to selected genes (Endler 1986; Hamilton *et al.* 2002; Hirao & Kudo 2004; Skot *et al.* 2005; Schmidt *et al.* 2008). However, one of the major challenges facing landscape geneticists studying adaptive variation is to include in the analysis all the environmental variables that might be correlated with loci across the genome.

Correspondence: Stéphanie Manel, Fax: 0476514279;
E-mail: stephanie.manel@ujf-grenoble.fr

The investigation of allele distribution in response to environmental factors parallels a research question of great interest in ecology, namely the explanation of species distribution patterns (Legendre 1993; Gaston & Blackburn 2009), since beta diversity (i.e. the variation in species composition among sites) is highly relevant for environmental protection and management policies (Legendre *et al.* 2005). In addition to multi-facetted competition, the distribution of species is influenced by many environmental variables, many of which are spatially structured. In consequence, patterns of species distribution are space-dependent (Legendre 1993; Wagner & Fortin 2005). At large scale, geographical and environmental variations are mostly inter-dependent, leading to respective patterns at broad spatial scales (Wiens 1989). In contrast, biotic processes such as dispersal, mating or competition mostly invoke spatial patterns at intermediate or small spatial scales.

Recently, Jombart *et al.* (2009) reviewed several options to consider spatial variation in ecological data. The method of Moran's eigenvector maps (MEM) was identified as a particularly promising tool for separating and modelling spatial patterns comprised within environmental variation (Borcard & Legendre 2002; Dray *et al.* 2006; Garcia *et al.* 2009). MEM analysis allows the identification of relationships between a variable of interest (e.g. allele frequency) and environmental variables at multiple spatial scales in regression or canonical analyses. In this study, the MEM analysis will provide proxies for unaccounted environmental variables that will be used as predictors in regressions to explain allele frequencies distribution. MEM variables are the eigenvectors of a spatial weighting matrix calculated from the site geographic coordinates. MEM analysis produces uncorrelated spatial eigenfunctions used to dissect the spatial patterns of the studied variation (allele frequencies in the present context) across a range of spatial scales. The first few MEM variables, which have large Moran's I coefficients (this is a measure of spatial autocorrelation; Epperson 2003), can be used to model broad-scale processes (e.g. environmental variation in space), whereas subsequent MEM variables with smaller Moran's I coefficients can be used to model the spatial autocorrelation generated by biotic processes such as individual dispersal (Dray *et al.* 2006). The MEM method provides a flexible approach to incorporate geographic relationships in model structure dealing with patterns at multiples scales (Diniz-Filho *et al.* 2009). Examples of its application are rare because of its recent development (but see Borcard *et al.* 2004; Dray *et al.* 2006; Garcia *et al.* 2009; Gazol & Ibanez 2010 for examples).

Once the purely geographical component in an allele distribution model is considered, another question related to space arises: what is the best spatial scale to study adaptation? Adaptation likely operates at different spatial scales, as pointed out by Savolainen *et al.* (2007). Until recently, molecular biologists expected adaptation to occur in a repeated and predictive way at broader spatial scales, at least in mobile species. Recent studies, however, have challenged this belief, showing that molecular adaptation is often local (Conover *et al.* 2006), a view shared by most ecologists. In reality, adaptive processes probably form a complex multi-scale continuum, with natural selection being the result of this complex continuum of scales. In addition, for application and in the light of global changes, identifying the working scales of adaptive genetic variation is crucial for species conservation (Stockwell *et al.* 2003). Given the complexity of scale-dependence in ecological patterns, it is essential to evaluate adaptive responses along a range of spatial scales (Thompson & McGarigal 2002).

In the present study, we aim at identifying loci significantly correlated with environmental variables (i.e. loci linked to genes under selection and therefore of ecological relevance) in the alpine plant *Arabis alpina* at three spatial scales, i.e. large, regional, and local scales. We introduce a new approach to correlate allele frequencies derived from genome scans with a wide array of environmental variables and broad-scaled MEM variables, the latter representing unmeasured broad-scale environmental variation. Since it is difficult to measure all environmental variables, our approach takes advantage of the spatial nature of environmental gradients to include unaccounted environmental variation in the analysis. We addressed the following specific question: are explanatory (i.e. environmental) variables driving patterns of adaptive genetic variation the same at different spatial scales?

Materials and methods

We used two published genetic data sets (Gugerli *et al.* 2008; Alvarez *et al.* 2009; Herrmann *et al.* 2010; Poncet *et al.* 2010), which allowed for analyses at three different scales. The first data set was sampled over the entire European Alps (large scale) (Alvarez *et al.* 2009; Thiel-Egenter *et al.* 2009) (Fig. 1a). The second one, obtained in the French Alps (regional scale) (Herrmann *et al.* 2010; Poncet *et al.* 2010), included three separate mountain massifs of the French Alps (local scale; Fig. 1b).

Study species, study areas and genotyping

Arabis alpina L. (Brassicaceae) is a perennial arctic-alpine rosette herb, which is widely distributed in the

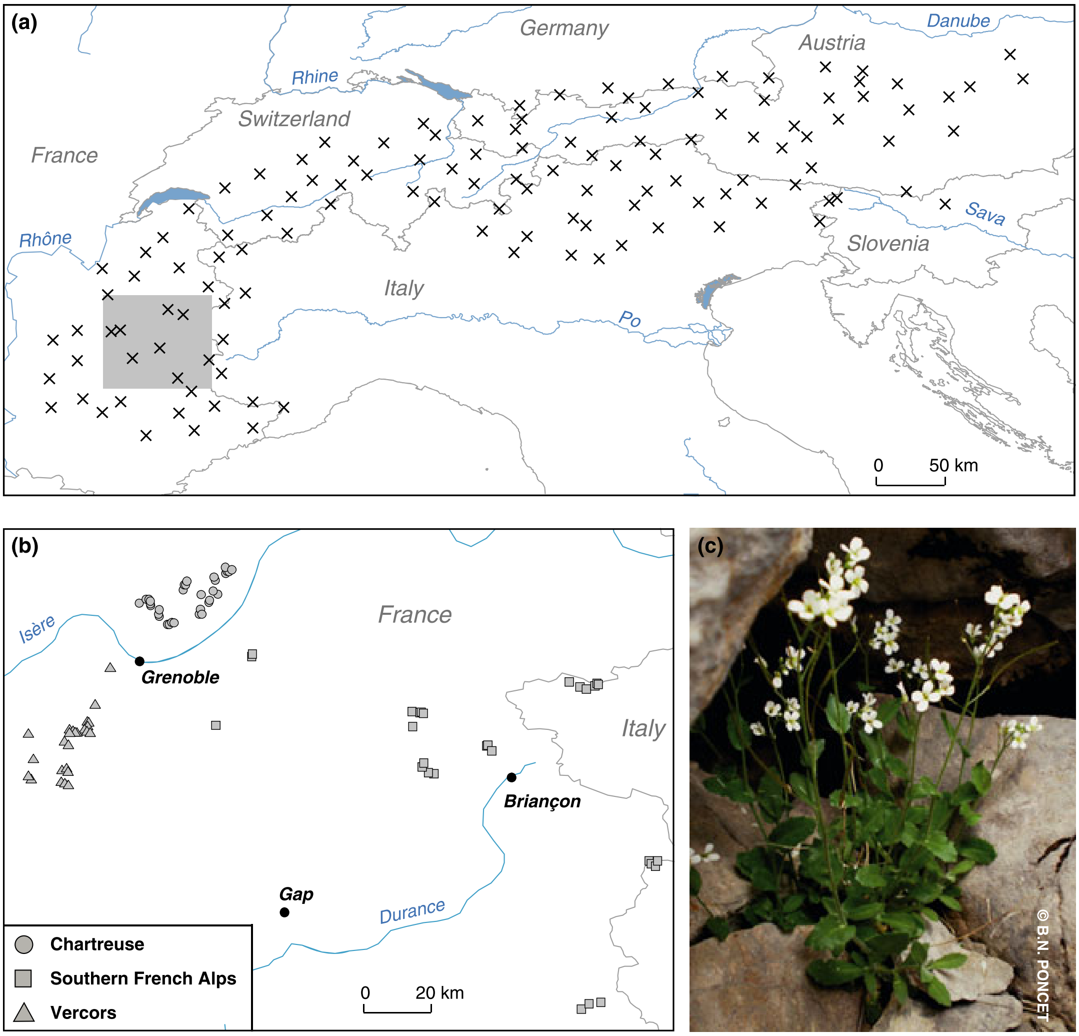

Fig. 1 Geographical arrangement of sampling locations (a) across the European Alps (crosses indicate the sampled populations) and (b) in the French Alps with three mountain massifs (Chartreuse, Vercors, southern French Alps) of (c) the Crucifer *Arabis alpina*. The grey square in (a) indicates the sampling area in the French Alps.

European Alps (Bovet *et al.* 2006; Ehrich *et al.* 2007). It has a large altitudinal distribution, ranging from the montane to the alpine and even nival vegetation belt. *A. alpina* prefers open, moist and rocky habitats but also grows at nutrient-rich, densely vegetated sites. It mainly reproduces sexually through seeds or asexually via stoloniferous growth (Ansell *et al.* 2008).

The samples of the first data set were collected and genotyped in the course of the European project INTRABIODIV (Gugerli *et al.* 2008). Leaf samples of *A. alpina* were collected across the European Alps (latitude: 44°48′ to 48°36′; longitude: 5°20′ to 15°40′) within a 12′ latitude × 20′ longitude rectangular grid system (ca. 23 km × 25 km) during summer of 2004. The elevation of the sampling locations ranged from 640 m to 2820 m above sea level. Three plants (at a minimum distance of 10 m) were sampled in 130 cells, resulting in a total of 385 samples distributed over 171′350 km^2 (Fig. 1a). Samples were dried in silica gel, and AFLP data were generated using a protocol inspired from Vos *et al.* (1995) as described in Gugerli *et al.* (2008). After electrophoresis on an ABI 3100 automated sequencer (Applied Biosystems), 140 polymorphic AFLP loci were scored for presence/absence using GENOGRAPHER (http://hordeum.oscs.montana.edu/genographer/). A mean error rate of 1.3% per locus was estimated based on 46 duplicated samples.

The second data set was sampled in the French Alps (Herrmann *et al.* 2010; Poncet *et al.* 2010). Plants were collected from 93 locations in three mountain massifs (Vercors, Chartreuse, southern French Alps; Fig. 1b). Elevation of the sampling locations ranged from 691 m to 3133 m above sea level. Fresh plant material from three to nine individuals per location was collected in the summer of 2006 and dried in silica gel. AFLP data were generated as outlined in Herrmann *et al.* (2010). After electrophoresis on an ABI 3100 automated sequencer, 1731 AFLP markers were scored for presence/absence from 321 individuals using GENEMAPER 3.7 (Applied Biosystems). An error rate of 1.2% per locus was estimated by duplicated analyses of 39 samples. Using the R script described in Herrmann *et al.* (2010), we automatically determined the presence or absence of AFLP fragments based on criteria related to peak height distribution, and we further removed loci of low reproducibility. This procedure resulted in a final data set containing 712 polymorphic AFLP loci. Note that the two genetic data sets described above did not comprise the same AFLP loci. We could thus compare the percentages of detected loci and of environmental predictors correlated to allele distributions but not directly of corresponding genomic regions between large vs. regional and local scales. For further analysis, we used the allele frequencies at AFLP loci per location.

Environmental variables and PCA

Fourteen monthly and annual environmental variables related to temperature, precipitation and topography were extracted per sampling location from published GIS eco-climatic layers from 1980 to 1989 (200-m resolution; Zimmermann & Kienast 1999) (Table 1). A principal component analysis (PCA) was applied to these environmental variables to examine possible correlations between eco-climatic variables and elevation and remove redundant variables (i.e. variables that were correlated at $|r| > 0.8$ and which were logically related). We first identified variables correlated to each retained axis, creating groups of variables. Within each group, we kept only one (or two) variables considered to be the most pertinent in terms of local adaptation in plants.

Statistical analysis

For the identification of AFLP loci of ecological relevance, we applied the suite of analyses described below for each of the three spatial scales separately (i.e. large, regional and local) with the objective of explaining AFLP allele distributions by environmental predictors.

First, we used only the environmental variables identified as being uncorrelated from the PCA analysis. These environmental variables, except aspect, were used both untransformed and transformed into cubic polynomials, the latter to account for nonlinear relationships between AFLP allele frequencies and these variables (Legendre & Legendre 1998). Aspect was transformed into sin(aspect) and cos(aspect), rendering this variable appropriate for use in linear models. Since environmental and geographical variation are considered to covary at broad spatial scales (see Introduction), we assumed that the spatial signature of unaccounted environmental variables was mostly broad-scaled. Therefore, we used broad-scaled MEM (i.e. the first half of the MEM eigenfunctions that model positive spatial correlation). They were used as proxies for unaccounted environmental variation. MEM are spatial eigenfunctions, computed from the geographic coordinates of the study sites, that describe the spatial relationships among the sites at all scales that can be perceived by the sampling design. The type of MEM variables computed in the present study were formerly called principal coordinates of neighbour matrices (PCNM) (Borcard & Legendre 2002; Borcard *et al.* 2004; Dray *et al.* 2006).

Table 1 Environmental (climatic and topographic) variables used to explain allele frequencies at amplified fragment length polymorphism (AFLP) loci in *Arabis alpina*. Variables finally included in the analyses are in bold

Yearly climatic layers (period 1980–1989)	ddeg: annual degree days above 0°C from daily climate maps [°C × days]
	prcpangy: mean annual precipitation sum [cm]
	srad: annual mean of daily global radiation (horizon- terrain-corrected) [kJ/m^2/day]
	tmaxavgiy: number of days with maximum temperature below freezing
	tmaxavgty: mean annual maximum temperature [°C]
	tminavgiy: number of days with minimum temperature below freezing
	tminavgty: mean annual minimum temperature [°C]
Seasonal climatic layers (period 1980–1989)	**prcp0305: spring seasonal precipitation: number of rain days from March to May**
	prcp0608: summer seasonal precipitation: number of rain days from June to August
Topography	dem: altitude
	slp: slope [%]
	asp: aspect
	topo: integrated topographic exposure map
	twi: potential soil humidity

Two joint tables of explanatory variables (i.e. predictors) were produced. The first one comprised the untransformed environmental variables (i.e. linear effects only, except for the cos/sin transformed aspect) and the broad-scale MEMs, while the second contained the transformed variables (i.e. polynomial effects and cos/sin aspect) and the broad-scale MEMs.

Second, for the large and regional scales, two multiple regressions were computed, between the allele frequencies per location and the two joint tables comprising respectively the linear and the polynomial effects. For the local scale, a multiple regression between allele frequencies per location was computed using only the first joint table containing the linear effects; the number of explanatory variables tested in the joint table with polynomial effects was too large given the number of study locations. Adjusted R^2 values were computed, which are unbiased estimators of the explanatory power provided by a particular set of variables (Ohtani 2000). Loci with an adjusted $R^2 \geq 0.5$ were considered to be significantly correlated to at least one predictor. In principle, several criteria could be used to identify loci of ecological relevance. (1) For example, one could test whether the regression of the loci on the environmental and broad-scaled MEM variables is significant, to identify the loci whose spatial variation can, at least to some extent, be explained by the environmental variables. This would be a liberal criterion: many loci with significant albeit small explained variation would be retained. One should, of course, use a correction for multiple testing (e.g. the Holm or Hochberg corrections; see Wright 1992) in order to have an experiment-wise error rate approximately equal to the significance level. (2) On the other hand, and as applied in this study, one could use the more conservative criterion of the proportion of variation of each locus explained by the environmental variables. One would use the adjusted R-square, R^2_{adj}, which is an unbiased estimator of the explained variation, and determine a threshold (here 0.50) combined with the criterion of a significant relationship; any other convenient threshold value could have been used.

Finally, we used linear regressions between allele frequencies at each of the identified significantly correlated loci and each predictor separately in order to estimate the explanatory power provided by each environmental variable, using R^2_{adj} values.

Results

PCA analysis of environmental variables

The first two axes of the PCA explained 60% of the variation in the environmental variables for the European Alps and 68% for the French Alps. At both scales, all temperature variables and elevation were highly correlated ($|r| > 0.8$) with the first PCA axis (Fig. 2). In the French Alps (Fig. 2b), precipitation variables were also correlated with the first PCA axis but less strongly than the temperature variables. For the European Alps (Fig. 2a), the precipitation variables (except summer seasonal precipitation, *prcp0608*) were correlated with the second PCA axis. We thus retained one variable from those correlated with axis 1 and two variables from those correlated to axis 2. Since we were interested in the biological interpretation of these variables, we preferred non-synthetic variables (by opposition to synthetic variable like PCA

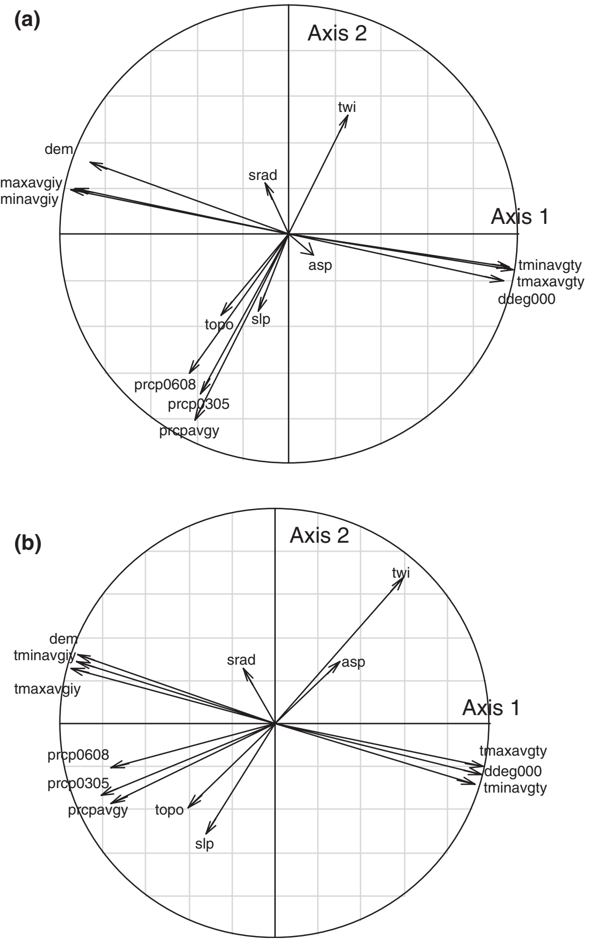

Fig. 2 Representation of the standardized variables on a principal component plot, with PCA axis 1 as the abscissa and axis 2 as the ordinate. The circle delimiting the graph corresponds to the circle of maximum fit of the standardized variables in the plot (i.e. correlation circle). The angles between variables and between variables and axes indicate their correlations. Percentage of variation explained by axes 1 and 2 respectively: (a) European Alps: 42%, 17%; (b) French Alps: 56%, 12%. For abbreviations of variables, see Table 1.

loadings, but see Manel *et al.* 2009 for a corresponding analysis and discussion). We retained the same eight environmental variables for the analysis at all scales: mean minimal temperature per year (*tminavgty*), annual mean of daily global radiation (*srad*), spring seasonal precipitation (*prcp0305*), summer seasonal precipitation (*prcp0608*), slope (*slp*), aspect (*asp*), integrated topographic exposure map (*topo*), and potential soil humidity (*twi*).

Detecting loci of ecological relevance

MEM analysis identified 11 broad-scaled MEM variables for the European Alps, 5 for the French Alps and 2 for each of the three local-scale analyses. The multiple linear regressions between allele frequencies and the joint table containing the transformed variables (i.e. polynomial effects and cos/sin aspect) and these broad-scaled MEMs, based on the R^2_{adj} criterion, detected 12% and 11% of all AFLP loci of ecological relevance at the large and regional scales, respectively (Fig. 3a, c; Table 2). Consideration of the linear environmental variables only reduced the percentages of loci with ecological relevance to 10% and 3%, respectively (Fig. 3b, d; Table 2). At the local scale and considering the linear explanatory variables only, we detected 3% of the loci as being of ecological relevance in Chartreuse (Fig. 3e), 16% in Vercors (Fig. 3f), and 17% in the southern French Alps (Fig. 3g; Table 2). Nine of the 21 loci of ecological relevance with a linear response identified at regional scale in the French Alps were also detected at the scale of single massifs (eight in the southern French Alps, one in Vercors). Two out of the 13 loci classified as being of ecological relevance in Chartreuse were also detected in Vercors, and Chartreuse and the southern French Alps also only shared two loci. Ten out of the 78 loci of ecological relevance identified in Vercors were also of ecological relevance in the Southern French Alps.

Environmental variables acting as potential selective pressures

To quantify the influence of the explanatory variables, R^2_{adj} values were estimated in separate regressions between each locus of ecological relevance and each of the eight explanatory variables, and then cumulated by explanatory variable using a bar plot (Fig. 4). At the scale of the European and French Alps, MEM variables had the highest explanatory power with high cumulated R^2_{adj} values over all loci of ecological relevance (Fig. 4a–d). The influence of MEM variables was less clear at the local scale (Fig. 4g). At that scale, however, only two explanatory MEM variables were used compared with 11 and 5 MEM variables at the large and regional scales.

After accounting for spatial effects through MEM variables, *tminavgty* was the environmental variable with the best explanatory power. It had the highest cumulated R^2_{adj} value at all spatial scales, except for the large scale (cubic polynomial) and for the local scale in massifs Vercors and Chartreuse where its scores came second after those of *prcp0608* or *prcp0305* (Fig. 4). The second major environmental driver of AFLP allele distributions was precipitation (*prcp0608* or *prcp0305*; Fig. 4).

Discussion

This paper proposes a new approach to identify loci linked to adaptive variation owing to their correlation with environmental variables, while also considering their spatial variation unaccounted for by the available environmental variables. This approach allowed us to identify loci of ecological relevance under natural conditions (Hamilton *et al.* 2002; Skot *et al.* 2002; Karrenberg & Widmer 2008; Schmidt *et al.* 2008). The originality of our approach lies in the use of broad-scaled MEM variables to model the spatial variation of the loci not accounted for by the environmental predictors included in the analysis (Borcard & Legendre 2002; Borcard *et al.* 2004; Dray *et al.* 2006). Moreover, large scale geographical effects on the structure described by genetic markers, which are also retrieved by MEM variables, relate to the influence of historical dynamics of *A. alpina*. This refinement of environmental association analysis provides a basis for identifying and characterizing genomic regions under selection and to subsequently evaluate their functionality using molecular and experimental approaches (Holderegger *et al.* 2008).

Methodological issues

Our approach combines a simple algorithm (i.e. linear regression) and a powerful measure of model validity (the adjusted R^2, Ohtani 2000). Other algorithms could be used instead of linear regression (e.g. logistic regression), but measuring the fit of the model would be more difficult because R^2_{adj} statistics are not available in generalized linear models (GLM) (Joost *et al.* 2007; Manel *et al.* 2009; Poncet *et al.* 2010). However, multi-model inferential approaches have been developed (Burnham & Anderson 2002) and can also be used in GLM to identify relevant selective pressures and select best models. Future studies on allele distribution models need to consider previous work on algorithms to model species distributions, such as comparative algorithm studies (Elith *et al.* 2006; Elith

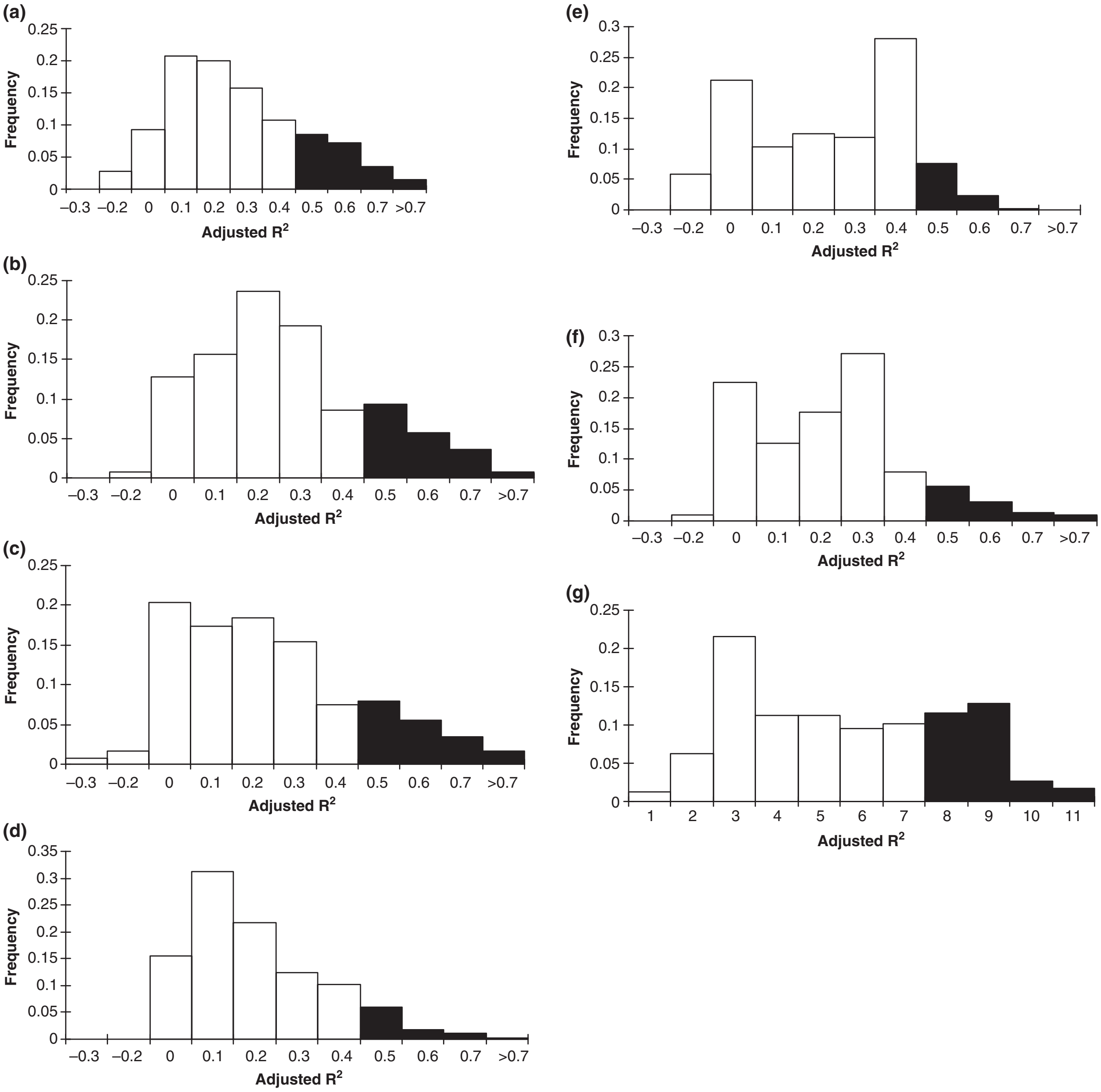

Fig. 3 Frequency of R^2_{adj} values from multiple linear regressions of *Arabis alpina* AFLP allele frequencies per locus and eight environmental variables (see Table 1) and broad-scaled MEMs. Environmental variables (except aspect) were either considered as linear (LE) or as cubic polynomials (PE). Aspect was sin- and cos-transformed. (a) European Alps—PE (17 out of 140 markers had $R^2_{\text{adj}} > 0.5 = 12\%$; (b) European Alps—LE (14 out of 140 = 10%); (c) French Alps—PE (76 out of 712 = 11%); (d) French Alps—LE (21 out of 712 = 3%); (e) Chartreuse—LE (13 out of 509 = 3%); (f) Vercors—LE (78 out of 498 = 16%); (g) Southern French Alps—LE (101 out of 588 = 17%).

& Graham 2009) or review and perspective analyses (Guisan *et al.* 2002; Thuiller *et al.* 2008).

If possible, it would also be useful to double-check the ecological relevance of the identified loci using population genomic approaches (Beaumont & Balding 2004; Foll & Gaggiotti 2008). Such approaches are used to detect outlier loci, i.e. those loci more strongly differentiated among populations than expected under neutrality (Beaumont & Nichols 1996; Beaumont & Balding 2004; Foll & Gaggiotti 2008). However, these genomic methods only deal with random samples from genetic populations, hence, they are not suitable for analyzing individuals continuously sampled over space. In our case, the adopted study designs of our two data sets precluded the application of methods relying on genetic population-based allele frequencies.

Table 2 Number of sites and loci analysed at each of three scales. Number of loci of ecological relevance detected either using cubic polynomials of the environmental variables and MEM variables or linear environmental variables and MEM variables as explanatory variables in multiple regressions

Scale	Study area	Number of sites	Number of loci	Number of loci (cubic)	Number of loci (linear)
Large	European Alps	44	140	17	14
Regional	French Alps	93	712	76	21
Local	Chartreuse	39	509	/	13
	Vercors	26	498	/	78
	Southern French Alps	30	588	/	101

Consistency of environmental predictors across different scales

The spatial variation measured by MEM variables seems to be more influential at large and regional scales than at a local scale (Fig. 4). A first explanation for this result is that 11 and 5 MEM variables explained AFLP allele frequencies at the larger scales, while only two MEM variables were identified at the local scale. A second explanation is that at smaller scales, environmental variables are less influenced by broad-scale spatial patterns.

After accounting for broad-scale spatial effects by MEMs, which represent purely geographical variation as well as the effects of unaccounted environmental variables, temperature and precipitation were identified as the two major drivers of allele distributions at all spatial scales, although this was less evident at a local scale (Fig. 4). A similar result was obtained for *A. alpina* sampled in the French and the Swiss Alps (Poncet *et al.* 2010). These authors used generalized estimation equations and found AFLP loci correlated with *tminavgty*, *prcp0305*, *slp* and *twi*. They identified 43 (5.2%) loci significantly related with mean annual minimum temperature in the French Alps, not considering massifs, and 10 (1.2%) in the Swiss Alps. The slightly higher percentage (12% and 11% at large and regional scales, respectively) of putative loci with adaptive relevance identified in the present study can partly be explained by the use of MEM variables that modelled the unaccounted environmental variation and by the higher number of environmental variables used.

Previous studies have shown the importance of temperature and precipitation in plant adaptation (Hamilton *et al.* 2002; Skot *et al.* 2002; St Clair *et al.* 2005; Richardson *et al.* 2009), and not only in alpine environments (Körner 2003). For example, Richardson *et al.* (2009) found an effect of both temperature and precipitation on AFLP loci in western white pine, *Pinus monticola*, from western North America, using population-based outlier loci detection methods (Beaumont & Balding 2004). We see the coincidental responses of temperature and precipitation as a result of these two clinal variables being highly correlated to elevation in the European Alps (Körner 2003). Therefore, both variables vary often in parallel at any scale. On the other hand, the topography-related variables included in our analysis did not co-vary with elevation and thus represented a rather local component of environmental variation. This scale-dependent responsiveness of groups of environmental predictors should be considered in future studies associating adaptive genetic variation to environmental gradients.

A. alpina occurs across a large altitudinal range, which requires specific adaptations to the respective clines in environmental conditions. At the same time, the species may be found in various types of habitats at a local scale, such as scree slopes, rocky outcrops, moist or eutrophic sites. Our analysis did not consider such small-scale micro site variation, which may be the reason why topography-related environmental variables, e.g. *topo* or *twi*, revealed low values of R^2_{adj} (Fig. 4). As such, the inferred pattern of adaptation at different scales nicely fitted the ecological niche of *A. alpina* in the study ranges.

Scale-specific patterns of adaptation

Our approach detected AFLP loci of ecological relevance at all spatial scales considered. Two and 8% of the loci showed a non-linear relationship of allele frequencies with the environmental variables at the large and the regional scales, respectively, and no significant relationship was detected when considering linear responses. This result illustrates that adaptation was not always a direct linear response to the environment. In addition, we found a higher proportion of loci of ecological relevance responsive to linear environmental predictors at a local scale, namely in the Vercors and the southern French Alps (16% and 17%, respectively); lower percentages of adaptive loci were related to linear predictors at the regional scale of the French Alps and

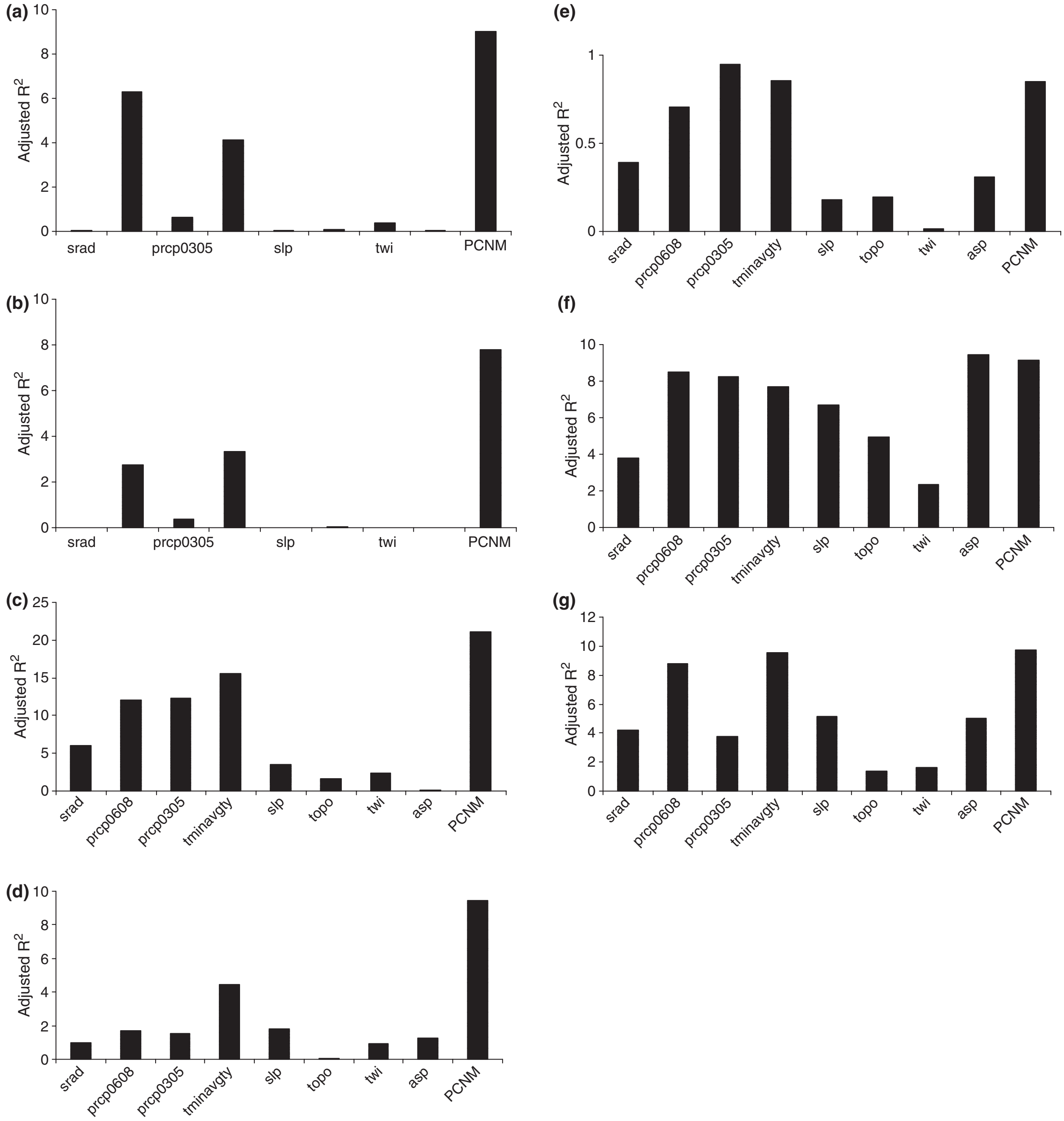

Fig. 4 Bar plots of R^2_{adj} values for the loci of ecological relevance presented in black in Fig. 3. R^2_{adj} values were calculated from simple linear regressions between allele frequencies and the eight environmental variables (either considered as linear (LE) or cubic polynomials (PE)), multiple regressions on sin(aspect) and cos(aspect), or multiple regressions on the broad-scale MEM variables. (a) European Alps—PE (11 MEMs); (b) European Alps—LE; (c) French Alps—PE (5 MEMs); (d) French Alps—LE; (e) Chartreuse—LE (2 MEMs); (f) Vercors—LE (2 MEMs); (g) Southern French Alps—LE (2 MEMs).

at the local scale of the Chartreuse massif. However, 43% of the loci of ecological relevance were in common between the regional and local scales.

The above results suggest that there may be two different types of adaptive responses acting on *A. alpina*. Many loci are probably involved in site-specific local adaptation, hence the large number of loci of ecological relevance at local scale, while other ecologically relevant loci are mainly involved in more general adaptive responses at larger geographical scales. The latter type likely reflects selective pressures consistent across scales in alpine plants (Körner 2003), such as adaptation to altitude or frost. In contrast, adaptive fine-tuning of gene regulation and expression acts at a local scale. The involvement of many genes in ecologically important traits as well as the action of local networks of gene regulation have recently been reviewed for plants and, in particular, for *Arabidopsis thaliana* (Alonso-Blanco *et al.* 2009). However, the interpretation of our comparisons across scales is preliminary for two reasons. First, there is large variation in the number of loci of ecological relevance among replicates at the local scale (3–17% of the loci). Second, we have no replicates available at the regional and large scales and, therefore, cannot assess the level of confidence in these estimates. Additional studies across various scales are needed before generalizations can be made.

Limitations and perspectives

Until recently, the AFLP technique was the most convenient method for obtaining large numbers of molecular markers for genomic studies in non-model organism, for which no a priori sequence knowledge was available (Meudt & Clarke 2007). We took advantage of this circumstance, making use of two large data sets to introduce our new approach. While the AFLP loci used in our study were not the same at large versus regional and local scales, we obtained a similar result for all scales, namely that temperature and precipitation are the major drivers of allele distributions after spatial signals were accounted for. This congruence gives strong support for the significance of the loci identified as ecologically relevant. Recent advances in next-generation sequencing technology and their increasing affordability (Hudson 2008) will allow us to replace AFLP by other, more informative markers such as expressed sequence tags (ESTs) or single-nucleotide polymorphisms (SNPs) (Manel *et al.* 2010) and to rely on more comprehensive genomic resources in environmental association studies. The very near future will see large SNP data sets with functional information for each SNP and large environmental databases (Manel *et al.* 2010) becoming available. However, whether SNPs (from candidate genes) or AFLPs are used, scarcely matters for this study: the only claim we make here is that our loci of ecological relevance are linked to genomic regions under selection. Even though it may be appealing to study SNPs within coding sequences, these may still refer to synonymous mutations with no change in protein function and, therefore, be non-adaptive like any anonymous marker. Experimental proof of the truly selective nature of a particular marker will ultimately be required in the case of AFLPs as well as SNPs.

Conclusions

Our results illustrate the potential of allele distribution models for identifying loci that are potentially influenced by natural selection along complex environmental gradients. Accounting for spatial and possibly unidentified environmental variation using MEM spatial eigenfunctions allowed us to detect a high number of loci of ecological relevance. However, allele distribution models only identify loci that are physically linked to genomic regions harbouring genes that display relevant adaptive variation. Simultaneously, they identify potential selective pressures. Hence, the next steps in the analysis of the adaptive value of such loci will be to characterize them using genomic information, pinpoint the underlying genes and their variation on a molecular basis, and test their ecological relevance in selection experiments (Gienapp *et al.* 2008; Hoffmann & Willi 2008). The rapid increase in sequence availability (Martin & Martin 2010), including sequences from non-model organisms, should pave the way for rapid progress in the study of adaptive response to environmental variation and change.

Acknowledgements

This work was conducted as part of the Landscape Genetics Working Group (SM, RH and PL) supported by the National Center for Ecological Analysis and Synthesis, a centre funded by NSF (Grant #DEB-0553768), the University of California, Santa Barbara, and the State of California. SM was also supported by the Institut Universitaire de France. Contributions of RH and FG were associated with the CCES-BIOCHANGE project of the ETH domain. We also thank the EC-supported project INTRABIODIV and its consortium for providing the data set across the European Alps. We are thankful to S. Dray for helpful comments on MEM analysis. We also thank Victoria Sork, Bryan Epperson, Deborah Zulliger and several anonymous reviewers for helpful comments on earlier versions of the manuscript.

References

Alonso-Blanco C, Aarts MGM, Bentsink L *et al.* (2009) What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*, **21**, 1877–1896.

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Ansell SW, Grundmann M, Russell SJ, Schneider H, Vogel JC (2008) Genetic discontinuity, breeding-system change and population history of *Arabis alpina* in the Italian Peninsula and adjacent Alps. *Molecular Ecology*, **17**, 2245–2257.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

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

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular Ecology*, **17**, 3583–3584.

Bonin A, Miaud C, Taberlet P, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826–1832.

Bovet L, Kammer PM, Meylan-Bettex M, Guadagnuolo R, Matera V (2006) Cadmium accumulation capacities of *Arabis alpina* under environmental conditions. *Environmental and Experimental Botany*, **57**, 80–88.

Burnham K, Anderson A (2002) *Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer, New York.

Conover DO, Clarke LM, Munch SB, Wagner GN (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology*, **69**, 21–47.

Diniz-Filho JAF, Nabout JC, Telles MPD, Soares TN, Rangel T (2009) A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genetics and Molecular Biology*, **32**, 203–211.

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.

Ehrich D, Gaudeul M, Assefa A *et al.* (2007) Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*, **16**, 2542–2559.

Elith J, Graham CH (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**, 66–77.

Elith J, Graham CH, Anderson RP *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Endler J (1986) *Natural Selection in the Wild*. Princeton University Press, New Jersey, USA.

Epperson B (2003) *Geographical Genetics*. Princeton University Press, New Jersey, USA.

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.

Garcia D, Rodriguez-Cabal MA, Amico GC (2009) Seed dispersal by a frugivorous marsupial shapes the spatial scale of a mistletoe population. *Journal of Ecology*, **97**, 217–229.

Gaston K, Blackburn T (2009) *Pattern and Process in Macroecology*. Blackwell Science, Oxford.

Gazol A, Ibanez R (2010) Variation of plant diversity in a temperate unmanaged forest in northern Spain: behind the environmental and spatial explanation. *Plant Ecology*, **207**, 1–11.

Gienapp P, Teplitsky C, Alho JS, Mills JA, Merila J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*, **17**, 167–178.

Gugerli F, Englisch T, Niklfeld H *et al.* (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation—a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 259–281.

Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89–100.

Hamilton NRS, Skot L, Chorlton KH, Thomas ID, Mizen S (2002) Molecular genecology of temperature response in *Lolium perenne*: 1. preliminary analysis to reduce false positives. *Molecular Ecology*, **11**, 1855–1863.

Herrera CM, Bazaga P (2008) Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. *Molecular Ecology*, **17**, 5378–5390.

Herrmann D, Poncet B, Manel S *et al.* (2010) Selection criteria for scoring amplified length polymorphisms (AFLPs) and their effect on the reliability of population genetic parameter estimates. *Genome*, **53**, 302–310.

Hirao AS, Kudo G (2004) Landscape genetics of alpine-snowbed plants: comparisons along geographic and snowmelt gradients. *Heredity*, **93**, 290–298.

Hoffmann A, Willi Y (2008) Detecting genetic response to environmental change. *Nature Reviews Genetics*, **9**, 421–432.

Holderegger R, Herrmann D, Poncet B *et al.* (2008) Land ahead: Using genome scans to identify molecular markers of adaptive relevance. *Plant Ecology and Diversity*, **1**, 273–283.

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources*, **8**, 3–17.

Jombart T, Dray S, Dufour AB (2009) Finding essential scales of spatial variation in ecological data: a multivariate approach. *Ecography*, **32**, 161–168.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.

Karrenberg S, Widmer A (2008) Ecologically relevant genetic variation from a non-*Arabidopsis* perspective. *Current Opinion in Plant Biology*, **11**, 156–162.

Körner C (2003) *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems*, 2nd edn. Springer, New York, USA.

Legendre P (1993) Spatial autocorrelation—trouble or new paradigm. *Ecology*, **74**, 1659–1673.

Legendre P, Legendre L (1998) *Numerical Ecology*, 2nd English edition. Elsevier, Amsterdam.

Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.

Manel S, Segelbacher G (2009) Perspectives and challenges in landscape genetics. *Molecular Ecology*, **18**, 1821–1822.

Manel S, Conord C, Després L (2009) Genome scan to assess the respective roles of host-plant and environmental constraints on the adaptation of a widespread insect. *BMC Evolutionary Biology*, **9**, 288.

Manel S, Joost S, Epperson BK *et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760–3772.

Martin NF, Martin F (2010) From Galactic archeology to soil metagenomics—surfing on massive data streams. *New Phytologist*, **185**, 343–347.

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science*, **12**, 106–117.

Oetjen K, Reusch TBH (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology*, **16**, 5156–5167.

Ohtani K (2000) Bootstrapping R^2 and adjusted R^2 in regression analysis. *Economic Modelling*, **17**, 473–483.

Poncet B, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance in two independent regional population samples of *Arabis alpina*. *Molecular Ecology*, doi:10.1111/j.1365-294X.2010.04696.x.

Richardson BA, Rehfeldt GE, Kim MS (2009) Congruent climate-related genecological responses from molecular markers and quantitative traits for western white pine (*Pinus monticola*). *International Journal of Plant Sciences*, **170**, 1120–1131.

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution and Systematics*, **38**, 595–619.

Schmidt PS, Serrao EA, Pearson GA *et al.* (2008) Ecological genetics in the north Atlantic: environmental gradients and adaptation at specific loci. *Ecology*, **89**, S91–S107.

Skot L, Hamilton NRS, Mizen S, Chorlton KH, Thomas ID (2002) Molecular genecology of temperature response in *Lolium perenne*: 2. association of AFLP markers with ecogeography. *Molecular Ecology*, **11**, 1865–1876.

Skot L, Humphreys MO, Armstead I *et al.* (2005) An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Molecular Breeding*, **15**, 233–245.

St Clair JB, Mandel NL, Vance-Boland KW (2005) Genecology of Douglas fir in western Oregon and Washington. *Annals of Botany*, **96**, 1199–1214.

Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology and Evolution*, **18**, 94–101.

Thiel-Egenter C, Gugerli F, Alvarez N *et al.* (2009) Effects of life history traits on the genetic diversity of high-mountain plants: a multi-species study across the Alps and the Carpathians. *Global Ecology and Biogeography*, **18**, 78–87.

Thompson CM, McGarigal K (2002) The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). *Landscape Ecology*, **17**, 569–586.

Thuiller W, Albert C, Araujo MB *et al.* (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.

Vos P, Hagers R, Bleeker M *et al.* (1995) AFLP: new technique for DNA fingerprinting. *Nucleic Acid Research*, **23**, 4407–4414.

Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: Concepts and statistics. *Ecology*, **86**, 1975–1987.

Wiens JA (1989) Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.

Wright SP (1992) Adjusted *P*-values for simultaneous inference. *Biometrics*, **48**, 1005–1013.

Zimmermann NE, Kienast F (1999) Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *Journal of Vegetation Science*, **10**, 469–482.