MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 2896–2907 doi: 10.1111/j.1365-294X.2010.04696.x

Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*

BÉNÉDICTE N. PONCET,*[1] DORIS HERRMANN,†‡[1] FELIX GUGERLI,† PIERRE TABERLET,* ROLF HOLDEREGGER,† LUDOVIC GIELLY,* DELPHINE RIOUX,* WILFRIED THUILLER,* SERGE AUBERT* and STÉPHANIE MANEL*§

*Laboratoire d'Écologie Alpine (LECA), CNRS UMR 5553, Grenoble Université, BP 53, 2233 Rue de la Piscine, 38041 Grenoble Cedex 9, France, †WSL Swiss Federal Research Institute, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland, ‡Indo-Swiss Collaboration in Biotechnology (ISCB), EPFL AI-VP CO-ISCB, 1015 Lausanne, Switzerland, §Laboratoire Population–Environnement–Développement, UMR 151 UP/IRD, Université de Provence, centre Saint-Charles, Case 10, 3, place Victor-Hugo, 13331 Marseille Cedex 3, France

Abstract

Understanding the genetic basis of adaptation in response to environmental variation is fundamental as adaptation plays a key role in the extension of ecological niches to marginal habitats and in ecological speciation. Based on the assumption that some genomic markers are correlated to environmental variables, we aimed to detect loci of ecological relevance in the alpine plant *Arabis alpina* L. sampled in two regions, the French (99 locations) and the Swiss (109 locations) Alps. We used an unusually large genome scan [825 amplified fragment length polymorphism loci (AFLPs)] and four environmental variables related to temperature, precipitation and topography. We detected linkage disequilibrium among only 3.5% of the considered AFLP loci. A population structure analysis identified no admixture in the study regions, and the French and Swiss Alps were differentiated and therefore could be considered as two independent regions. We applied generalized estimating equations (GEE) to detect ecologically relevant loci separately in the French and Swiss Alps. We identified 78 loci of ecological relevance (9%), which were mainly related to mean annual minimum temperature. Only four of these loci were common across the French and Swiss Alps. Finally, we discuss that the genomic characterization of these ecologically relevant loci, as identified in this study, opens up new perspectives for studying functional ecology in *A. alpina*, its relatives and other alpine plant species.

Keywords: adaptive genetic variation, amplified fragment length polymorphism, generalized estimating equations, landscape genomics, local adaptation

Received 15 January 2010; revision received 22 March 2010; accepted 26 April 2010

Introduction

Adaptation is the evolution of a plant population towards a phenotype that better fits the present environment (Orr 2005). The major driving force of adaptation is divergent selection (Kawecki 2008). Understanding the genetic changes resulting in divergent selection is a major objective in evolutionary genetics (Feder & Mitchell-Olds 2003), since adaptation plays a key role in the extension of the ecological niche to marginal habitats (Kawecki 2008) as well as in ecological speciation (Schluter & Conte 2009). Adaptation to new habitats is also a potential plant response to shifts in environmental conditions, which is all the more crucial in the context of climate change (Reusch & Wood 2007). Until recently, it has been difficult to detect genetic variation of adaptive relevance

Correspondence: S. Manel, Fax: +33 476514279;
E-mail: stephanie.manel@ujf-grenoble.fr
[1]The first two authors contributed equally to this paper.

© 2010 Blackwell Publishing Ltd

in nonmodel organisms (Bonin *et al.* 2006). Genome scans combined with landscape genomic approaches have the potential to improve our understanding of the genetic basis of adaptation in nonmodel species with unknown genomes (Storz 2005) through the investigation of a large number of loci distributed across a genome in a large number of individuals (Black *et al.* 2001; Luikart *et al.* 2003; Schlötterer 2003). In nonmodel plants, genome scans have been used to detect ecologically relevant loci (ERLs) under various conditions (Scotti-Saintagne *et al.* 2004; Savolainen *et al.* 2006; Parisod & Christin 2008). However, underlying environmental cues or even functional traits involved in adaptation have rarely been identified in these studies (Vasemägi & Primmer 2005). For example, Namroud *et al.* (2008) tracked candidate genes for local adaptation in six Canadian populations of white spruce (*Picea glauca*) using a genome scan of expressed sequence tags (EST). Amongst the genes identified as potentially being under natural selection, half were specific to populations in the warmest locations and half to populations experiencing different extreme conditions (most arid, coldest or most humid locations).

Allele distribution models (Holderegger *et al.* 2007), which relate allele frequencies to environmental variables, aim to identify genes of ecological relevance. However, patterns of genetic variation that seem to be related to natural selection may also result from historical and spatial effects (Endler 1986; Kawecki & Ebert 2004; Schmidt *et al.* 2008). Firstly, isolation by distance might restrict gene flow among populations. Thereby, neutral alleles will change in frequency due to genetic drift (Wright 1938). However, this trend would affect the entire genome and not only specific loci. Secondly, contact and admixture zones, where populations which have diverged in isolation (e.g. in different glacial refugia) come into secondary contact, may result in confusing genetic patterns due to past selection followed by a spatial gradient of population admixture (Endler 1977). Using replicated regions and searching for congruent patterns of ERLs across regions is one way of limiting the confounding influence of historical and spatial processes, as it is unlikely that these neutral processes and genetic drift would generate similar genetic patterns at a locus across independent environmental gradients (e.g. see Wark & Peichel 2010).

The objective of this study was to identify loci of ecological relevance in the perennial alpine herb *Arabis alpina*. We did so based on the premise that loci whose distributions are correlated to environmental variation are ecologically relevant and could be linked to genomic regions undergoing selection. Although loci of ecological relevance might well be region-specific, more weight will be given to loci identified in two independent regions. To investigate the above assumption, we developed a large data set of more than 800 amplified fragment length polymorphism (AFLP) loci, greatly exceeding the numbers usually tested in genome scan studies. Prior to the outlier analysis, we checked for linkage disequilibrium in the set of loci and elucidated historical and spatial effects by evaluating population structure, testing an admixture model and investigating isolation by distance. To correlate allele frequencies to environmental variables, we used generalized estimating equations (GEE) with binomial error. The GEE method makes it possible to consider the spatial autocorrelation of individuals within sampling locations (Carl & Kuhn 2007), which can be high in the case of sessile plants. Finally, the ERLs identified are discussed with regard to available genomic resources (i.e. the *Arabidopsis thaliana* whole genome sequence), opening up new perspectives for studies in functional ecology.

Materials and methods

Study species and area

Arabis alpina L. (Brassicaceae) is a perennial arctic-alpine rosette plant widely distributed over the Arctic and mountainous regions of the northern hemisphere. It is particularly common in the European Alps (Bovet *et al.* 2006), where it shows a large altitudinal distribution, ranging from mountain forests to alpine and even nival vegetation patches. *A alpina* prefers moist, open and rocky habitats, and reproduces sexually or asexually by stoloniferous growth (Ansell *et al.* 2008). Its small seeds are dispersed by wind.

Plants were sampled from 99 locations in the French Alps and from 109 locations in the Swiss Alps (Table S1, Supporting information). In the French Alps, *A. alpina* was mainly sampled in three massifs: Vercors, Chartreuse and Briançonnais. In the Swiss Alps, sampling locations were distributed across south-eastern Switzerland (Fig. 1). They were chosen to cover a wide range of habitats (e.g. scree, moist, nutrient rich) and elevations (440–3133 m above sea level). Geographical distances between locations ranged from 23 m to 143 km in the French Alps and from 164 m to 132 km in the Swiss Alps. Fresh plant material from nine individuals per location (in the vast majority of cases) was collected between July and September 2006 and immediately dried in silica gel.

Environmental variables

A set of 20 monthly and annual environmental variables related to temperature, precipitation and topography was extracted for each sampling location from published GIS eco-climatic layers at the spatial resolution of 200 m

© 2010 Blackwell Publishing Ltd

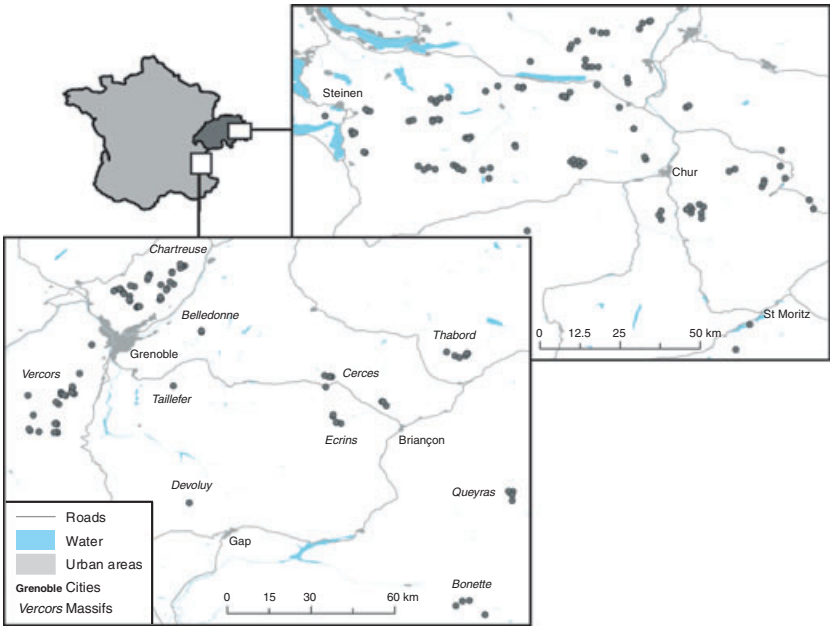

Fig. 1 Distribution of the 99 *Arabis alpina* sampling locations in the French Alps and of the 109 sampling locations in the Swiss Alps.

(Zimmermann & Kienast 1999). A principal component analysis was used to check for the correlation between these eco-climatic variables and elevation. Only the four most uncorrelated explanatory variables were selected from PCA: mean annual minimum temperature (tmin), sum of precipitation between March and May (prcp), slope index (slp) and topographic wetness index (twi; i.e. the ratio of the upslope contributing region on the tangent of the slope angle). The variable twi is a hydrological index developed by Beven & Kirkby (1979) as a measure of soil moisture at a spatial resolution of 400 m. These four selected variables showed variation among the sites sampled in the two study regions. They were used as explanatory environmental variables in the identification of loci considered to have undergone divergent selection.

DNA extraction and AFLP genotyping

Total DNA from three, and in few cases nine individuals per sampling location was extracted using the DNeasy 96 plant extraction kit (QIAGEN) according to the manufacturer's instructions. We used AFLPs (Vos *et al.* 1995), which have become the reference marker technique for genome scans (Bonin *et al.* 2007; Meudt & Clarke 2007), because AFLP loci are widely distributed across the genome. AFLP loci were generated for 768 samples. The AFLP procedure followed Vos *et al.* (1995) with minor modifications. Digestion of genomic DNA was performed for 2 h at 37 °C in a 20 μL mix using 2 U of *Mse*I and 5 U of *Eco*RI or *Pst*I (New England Biolabs). Double-stranded adaptors were then ligated to the digested DNA in a 40 μL volume for 2 h at 37 °C using 1 U of T4 DNA Ligase (Roche). Products were diluted 1:9, and pre-selective PCR (120 s at 94 °C, 30 cycles of 30 s at 94 °C, 30 s at 56 °C, 120 s at 72 °C with a final elongation of 10 min at 72 °C) was carried out in a 25 μL volume containing 1X PCR Buffer II pH 8.3, 1.5 mM of $MgCl_2$, 80 μM of each dNTP, 0.2 μM of each primer at 10 μM (*Eco*RI/*Mse*I or *Pst*I/*Mse*I), and 0.5 U of Ampli*Taq* DNA polymerase (Applied Biosystems). After 1:19 dilution of pre-selective PCR products, the selective amplification (10 min at 95 °C, 13 cycles of 30 s at 94 °C, 60 s at 65 °C to 56°C, and 60 s at 72 °C, 23 cycles of 30 s at 94 °C, 60 s at 56 °C, and 60 s at 72 °C with a final elongation of 10 min at 72 °C) was carried out in a 12.5 μL volume containing 2.5 μL diluted pre-selective PCR product, 1X PCR Buffer II, 2.5 mM of $MgCl_2$, 80 μM of dNTP mix, 0.2 μM of each primer (*Eco*RI+3/*Mse*I+3 or *Pst*I+3/*Mse*I+2), 8 μg/mL BSA and 0.5 U of Ampli*Taq* Gold DNA polymerase (Applied Biosystems). After preliminary tests, 19 primer combinations were chosen which resulted in clear bands of sufficient variability (Table 1). PCR products of *Eco*RI/*Mse*I combinations were purified using columns of half to half 5% Sephadex G50 and Sephacryl S200, diluted 1:9. PCR products from each *Eco*RI/*Mse*I primer pair were run separately, but multiplexed for *Pst*I/*Mse*I primer pairs for fragment length analysis. For *Eco*RI/*Mse*I combinations, 1.5 μL of the diluted, FAM-labelled products were mixed with 10 μL of HiDi formamide and 0.1 μL Genescan ROX 500 size standard (Applied Biosystems) and electrophoresed on an ABI

© 2010 Blackwell Publishing Ltd

Table 1 Number of DNA fragments generated by 19 AFLP primer–enzyme combinations used in *Arabis alpina*

Primer–enzyme combination*	Number of loci	Number of polymorphic loci†	Percentage of polymorphism	Mean error rate (%)‡
*Eco*RI-AAT/*Mse*I-CAC	100	37	37.0	1.2
*Eco*RI-AGC/*Mse*I-CAC	135	92	68.1	0.8
*Eco*RI-ATC/*Mse*I-CAC	126	61	48.4	0.7
*Eco*RI-AGG/*Mse*I-CAC	90	39	43.3	1.2
*Eco*RI-ACG/*Mse*I-CAG	102	61	59.8	1.4
*Eco*RI-AGC/*Mse*I-CTG	112	49	43.8	0.8
*Eco*RI-ACG/*Mse*I-CTC	106	48	45.3	1.2
*Pst*I-AAG/*Mse*I-CA	108	44	40.7	1.3
*Pst*I-ACT/*Mse*I-CA	84	42	50.0	1.0
*Pst*I-ATC/*Mse*I-CA	63	25	39.7	1.2
*Pst*I-AAC/*Mse*I-CT	96	48	50.0	1.5
*Pst*I-AGA/*Mse*I-CG	49	21	42.9	1.1
*Pst*I-ACA/*Mse*I-CA	49	24	49.0	1.2
*Pst*I-AAC/*Mse*I-CA	142	61	43.0	1.0
*Pst*I-AGA/*Mse*I-CA	101	35	34.7	1.5
*Pst*I-ACA/*Mse*I-CG	65	31	47.7	1.3
*Pst*I-AAC/*Mse*I-CG	89	50	56.2	1.2
*Pst*I-AGA/*Mse*I-CT	66	33	50.0	1.9
*Pst*I-ACA/*Mse*I-CT	48	24	50.0	2.3
Total	1731	825		
Mean	91.1	43.4	47.3	1.2

*Only selective bases are given for each enzyme-specific primer.
†A locus was considered polymorphic if at least three individuals showed a different allele than all other samples.
‡Mean error rates were calculated by summing differences between control profiles (i.e. samples that were genotyped twice) per primer–enzyme combination.

PRISM 3100 capillary sequencer (Applied Biosystems). For *Pst*I/*Mse*I combinations, 1 μL of each of the diluted PCR products was mixed with 10 μL of HiDi formamide and 0.15 μL Genescan ROX 500 size standard (Applied Biosystems) and electrophoresed on an ABI PRISM 3100-Avant capillary sequencer (Applied Biosystems). Raw data were sized with GeneMapper 3.7 (Applied Biosystems).

AFLP loci were obtained using a semi-automated procedure. Loci were defined manually in GeneMapper and scored for the presence or absence of AFLP fragments in each individual. The quality of all loci was then automatically checked with an R script as described in Herrmann *et al.* (2010; http://www-leca.ujf-grenoble.fr/logiciels.htm). This script assesses presence/absence within a particular locus based on peak height distribution and further removes loci with large peak height variation or low reproducibility. AFLP fragments shorter than 50 bp for *Eco*RI/*Mse*I or 75 bp for *Pst*I/*Mse*I primer–enzyme combinations were discarded. We included 40 control samples from DNA extraction throughout AFLP analysis to check for locus reliability (replicates between and within plates). In addition, we analysed 39 duplicated samples as blind controls. The mismatch error rate per primer–enzyme combination was calculated as the percentage of loci with different scores between these blind controls and their respective original samples after locus selection and phenotype identification (Bonin *et al.* 2004; Pompanon *et al.* 2005; Herrmann *et al.* 2010; Table 1). Loci were considered as polymorphic when at least three individuals had a different score than all other individuals.

Linkage disequilibrium among AFLP loci was tested using successive logistic regressions between each locus as the response variable and the others as explanatory variables. We also applied a principal component analysis to investigate correlations among the occurrences of AFLP loci within sites. These analyses were performed using the R software (R Development Core Team, 2007).

Population structure, historical and spatial effects

Here, our main objective was to investigate whether the French and the Swiss Alps could be considered as two independent study regions with no admixture due to secondary contact. We conducted an individual-level analysis to check the population structure of our dataset using STRUCTURE 2.2 adapted for dominant markers (Falush *et al.* 2007) with and without an admixture model. We determined the optimal model by testing a range of values for the number of populations (K). For each value of K, seven runs were conducted. The number of populations K was approximated as the highest estimated log probability of the data $Pr(X|K)$ (Pritchard *et al.* 2000). Because $Pr(X|K)$ often reaches a plateau for maximal K values and to be more discriminatory, we also used the approach introduced by Evanno *et al.* (2005) based on the computation of the ΔK. ΔK is defined as $|m_L(K+1) - 2m_L(K) + m_L(K-1)|/\mathrm{sd}_L(K)$ with m_L and sd_L being the mean and the standard deviation of the likelihood value across the runs. As recommended by these authors, we used the number of populations of the modal value of the ΔK distribution as the correct estimation of K. The STRUCTURE 2.2 runs were conducted for each value of K ranging between 1 and 8 using 50 000 burn in periods followed by 50 000 Markov chain Monte Carlo (MCMC) repetitions. In both the admixture and the nonadmixture models, the correlated allele frequency option was used.

Genetic differentiation between the two sampling regions was estimated with AFLP-SURV (Vekemans *et al.* 2002) using the Bayesian method with nonuniform prior distribution and assuming Hardy–Weinberg equilibrium ($F_{IS} = 0$). Although this last assumption is

© 2010 Blackwell Publishing Ltd

expected to be violated (preliminary not shown results), Ansell *et al.* (2008) and Bonin *et al.* (2007) have shown that multilocus estimates of F_{ST} were robust to change in assumed F_{IS}. Genetic differentiation between sampling locations was estimated (1000 permutations).

To test for possible spatial effects, we calculated isolation by distance (IBD) estimates. For each region, pairwise kinship coefficients (F_{ij}) for all sample pairs, adjusted for dominant loci and assuming no inbreeding (Hardy 2003), were estimated using SPAGeDi (Hardy & Vekemans 2002). Kinship under IBD is expected to decrease in line with geographic distance. The correlations between pairwise F_{ij} and both untransformed and log-transformed Euclidian distances were inferred (999 permutations) and plotted for 1000 classes of geographic distances.

Detection of ecologically relevant loci using GEE

To detect alleles correlated to environmental variables, we used generalized estimating equations (GEE), which are an extension of generalized linear models (Diggle & Elliott 1995; Carl & Kuhn 2007). GEE consider spatial autocorrelation between individuals collected at the same sampling location by including an additional variance component to accommodate correlated data. Individuals collected in the same location were considered to belong to the same cluster in the model. Therefore, each cluster usually consisted of three (rarely nine) plants, with 99 clusters in one region (French Alps) and 109 in the other region (Swiss Alps). This method makes it possible to take into account the fact that neighbouring individuals within locations (i.e. at distances of only some metres) are genetically more similar than distant individuals between locations (inter-location distances: 23–163 km). As we dealt with binary data, we used a logit-link and binomial error distribution to correlate allele occurrence for each AFLP locus per sampling location to quadratic polynomials of environmental variables. Using quadratic polynomials makes it possible to account for the variety of response curve shapes other than a linear response (Legendre & Legendre 1998). GEE are nonlikelihood-based models and cannot be compared using a maximum likelihood methods such as Akaike's information criterion (AIC). Instead, we used the quasi-likelihood information criterion (QIC) adapted for the selection of GEE models by Pan (2001). The best model is that with the lowest QIC. All combinations of variables were investigated for each locus. Finally, we tested 255 GEE (from the null model to the full model with four variables in second-order polynomials) separately in the French data set and in the Swiss data set. All GEE models were calculated using the R package *geepack* (Yan & Fine 2004), and we implemented the QIC calculation in R for this study (R Development Core Team 2007).

For each selected model, we also tested each regression coefficient using the Wald test. To avoid a high rate of false positives due to multiple tests, significance levels were calculated for each regional data set by minimizing the false discovery rate (FDR; Storey 2002). *P*-value rejection thresholds were adjusted so as to have less than one false positive locus per regional data set and were estimated using the R package *qvalue*.

Results

Locus polymorphisms

A total of 679 individuals from 99 sampling locations in the French Alps and 109 in the Swiss Alps (1 to 9 individuals per location; Table S1, Supporting information) were reliably scored for the presence or absence of AFLP fragments. Overall, 1731 reliable loci were identified from the 19 selected AFLP primer–enzyme combinations, and 825 (47.7%) of them were polymorphic (Table 1) ranging in size from 52 to 493 bp. Error rates were between 0.7% and 2.3% per primer–enzyme combination, with a mean of 1.2%. All individuals studied exhibited unique AFLP profiles. The final data set of the French samples consisted of 411 polymorphic loci in 299 individuals, whilst the data set of the Swiss samples consisted of 354 polymorphic loci in 322 individuals. We identified 254 (31%) common polymorphic loci between the two regions. Logistic regressions used to estimate independence between loci were significant for only 3.5% of all comparisons at $P = 10^{-7}$, defined by using FDR.

Population structure, historical and spatial effects

STRUCTURE estimated the maximal log-likelihood for all K values in the admixture model (Fig. 2a), but determining K only from the highest log probability of the data given K, $Pr(X \mid K)$, was difficult. The ΔK criterion (Evanno *et al.* 2005) reached its modal value for $K = 2$, suggesting that the uppermost level of genetic structure has two distinct clusters: the French and the Swiss Alps (Fig. 2b). The results for $K = 2$ to $K = 4$ (Fig. 2c) are presented to illustrate the formation of populations. No secondary contact zone between sampling regions was detected in our study (Fig. 2b). French and Swiss samples of *A. alpina* were significantly differentiated ($F_{ST} = 0.1652$, $P < 0.0001$). Based on these results, we considered the French and the Swiss Alps as two independent study regions, representing true replicates. The correlation between pairwise kinship coefficients and untransformed and log-transformed Euclidian distances between locations were significant for the French Alps ($r^2 = 0.33$, $P < 0.0001$

© 2010 Blackwell Publishing Ltd

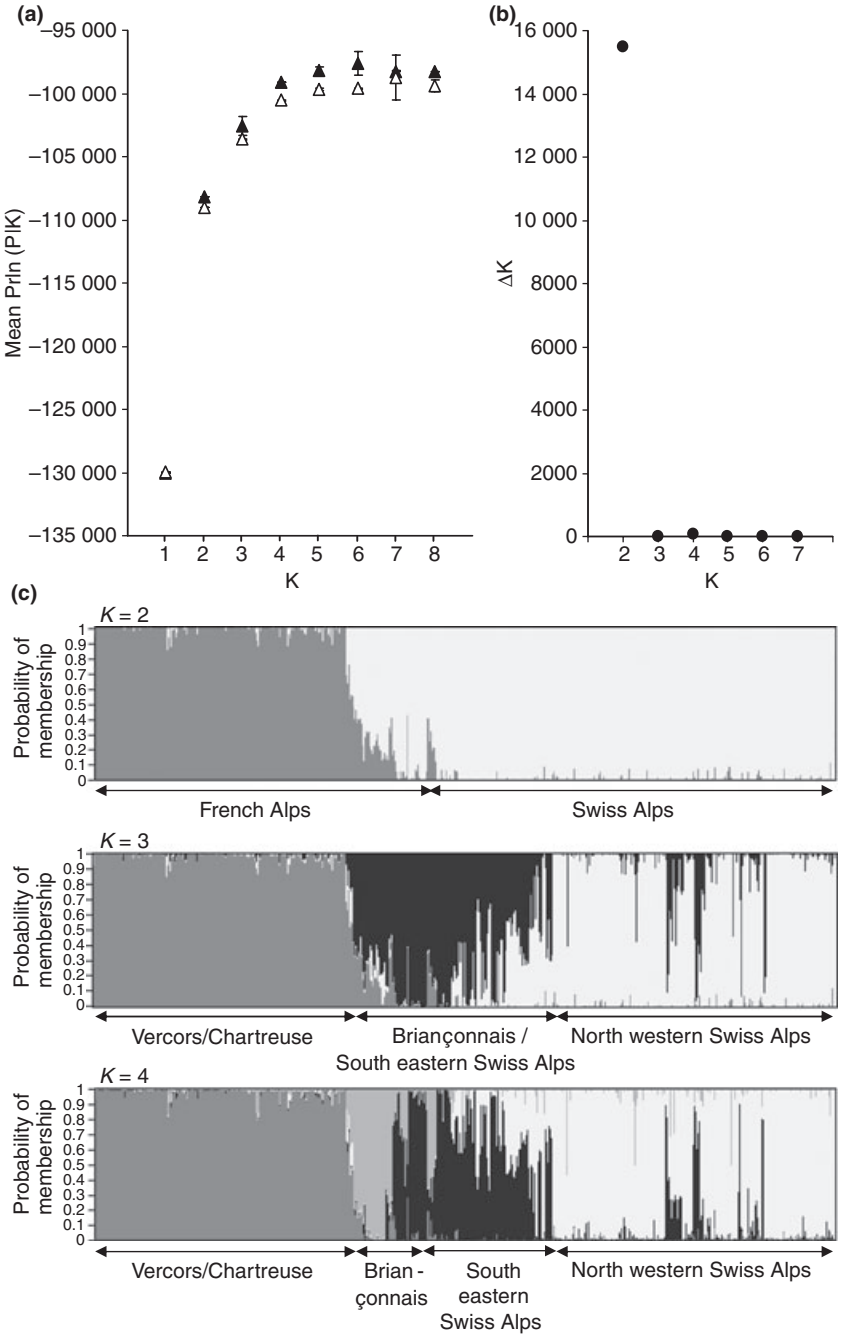

Fig. 2 Population structure of the alpine plant *Arabis alpina*, inferred using STRUCTURE (Pritchard *et al*. 2000; Falush *et al*. 2007), from 99 French and 109 Swiss sampling locations. Seven runs were conducted for each value of *K*. (a) Mean maximum likelihood probability of the data (±SD) represented as a function of the number of clusters (*K*) using admixture (black triangles) and no admixture models (white triangles). (b) Δ*K* as calculated by Evanno *et al*. (2005) for each *K* using admixture model. (c) Individual assignment to each cluster for *K* = 2, 3 and 4. Each individual is represented by a thin vertical line partitioned into 2, 3 or 4 shaded segments proportional to its membership to the corresponding genetic cluster.

and $r^2 = 0.39$, $P < 0.0001$, respectively) and for the Swiss Alps ($r^2 = 0.15$, $P < 0.0001$ and $r^2 = 0.16$, $P < 0.0001$ respectively; Fig. 3). The slope of the regression was steeper, indicating a stronger isolation by distance pattern in the French than in the Swiss Alps.

Detection of ecologically relevant loci using GEE

GEE were used to separately correlate allele frequencies in each of the two independent regions, the French and the Swiss Alps, to environmental variables. Thresholds

© 2010 Blackwell Publishing Ltd

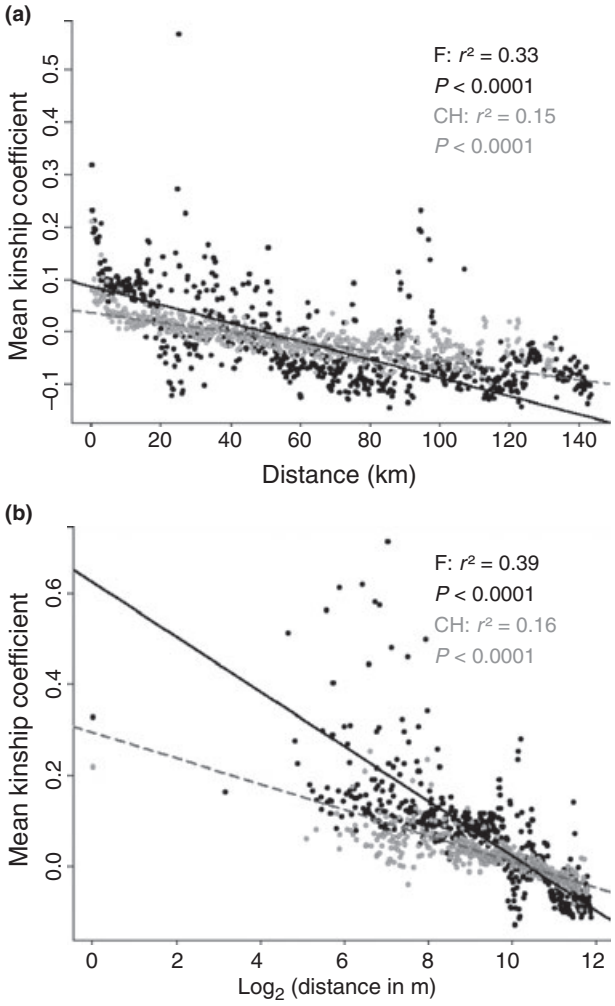

Fig. 3 Correlation between pairwise kinship (F_{ij}) coefficients of *Arabis alpina* individuals from 99 French (F, black line, black points) and 109 Swiss sampling locations (CH, grey line, grey points): (a) untransformed geographical distances and (b) log-transformed geographical distances.

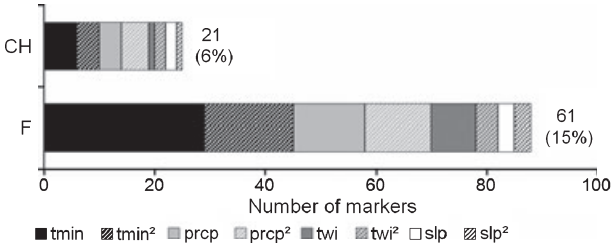

Fig. 4 Number of amplified fragment length polymorphism (AFLP) loci significantly correlated with environmental variables (quadratic polynomials) in each of the two study regions in the generalized estimating equations (GEE) analysis of *Arabis alpina* (F: French Alps; CH: Swiss Alps). Environmental variables are, from left to right, annual mean minimal temperature (tmin), monthly mean precipitation between March and May (prcp), soil wetness index (twi) and slope index (slp). The respective second order order variables are represented by the same shading, but hatched. The numbers at the end of the bars are the numbers of ecologically relevant loci and the corresponding percentage of these loci in relation to all loci (in brackets).

Table 2 Generalized estimating equation (GEE) analyses for four amplified fragment length polymorphism (AFLP) loci in *Arabis alpina* responding to environmental variables in both the French (F) and the Swiss (CH) Alps. QIC is the quasi-likelihood under the independence model criterion for GEE. These best models were of the kind $\beta_0 + \beta_1$ Variable$_1$

Locus	Region	QIC	β_0	Variable$_1$	β_1	P-value
PM251.7	F	95.1	−13.7	tmin	0.0374	8.98×10^{-6}
	CH	256.3	−6.2	tmin	0.0305	1.33×10^{-5}
PM342.7	F	121.4	−15.4	twi	2.6050	4.18×10^{-5}
	CH	336.4	2.9	prcp	−0.0867	1.08×10^{-5}
PM212.8	F	92.0	−25.2	twi	2.8016	1.45×10^{-4}
	CH	223.2	0.7	prcp^2	−0.0016	1.46×10^{-6}
EM125.1	F	60.7	−60.5	twi^2	−1.4342	4.11×10^{-4}
	CH	305.9	−4.8	tmin^2	7.62×10^{-5}	9.72×10^{-5}

for significant P values were estimated at 0.00042 for the French and 0.00012 for the Swiss samples. For each locus, the best GEE model with at least one significant regression coefficient is given in Table S2 (Supporting information). In the French data set, 61 loci (15% of all loci) were significantly correlated with one, two or three environmental variables (Table S2, Supporting information), with 10.5% of these 61 loci showing non-independence, i.e. linkage disequilibrium (results from logistic regressions, see above). Among these 61 ERLs, tmin was correlated with 43 loci (51% calculated in relation to the total number of significant correlations) while prcp was correlated with 21 loci (28%), twi with 12 loci (14%), and slp with six loci (7%; Fig. 4). In the Swiss data set, 21 loci (6% of all loci) were significantly correlated with one or two environmental variables (Table S2, Supporting information), with 14.3% of these 21 loci showing non-independence. Of the 21 ERLs identified in this second region, tmin was correlated with 10 loci (40%), prcp with eight loci (36%), twi with three loci and slp with two loci (12%; Fig. 4). Second-order polynomials of environmental variables were involved in 32 loci in the French data set and in 12 loci in the Swiss data set (Fig. 4).

Four loci were jointly identified as ERLs in both regions studied (Table 2). Locus PM251.7 was correlated with tmin both in the French and the Swiss Alps. The frequency of this locus increased with tmin in the same order of magnitude in both regions as indicated by the positive value of the regression coefficients (β_1 = 0.0374 and 0.0305, respectively; Table 2) and by the spatial distribution of tmin in the study area

© 2010 Blackwell Publishing Ltd

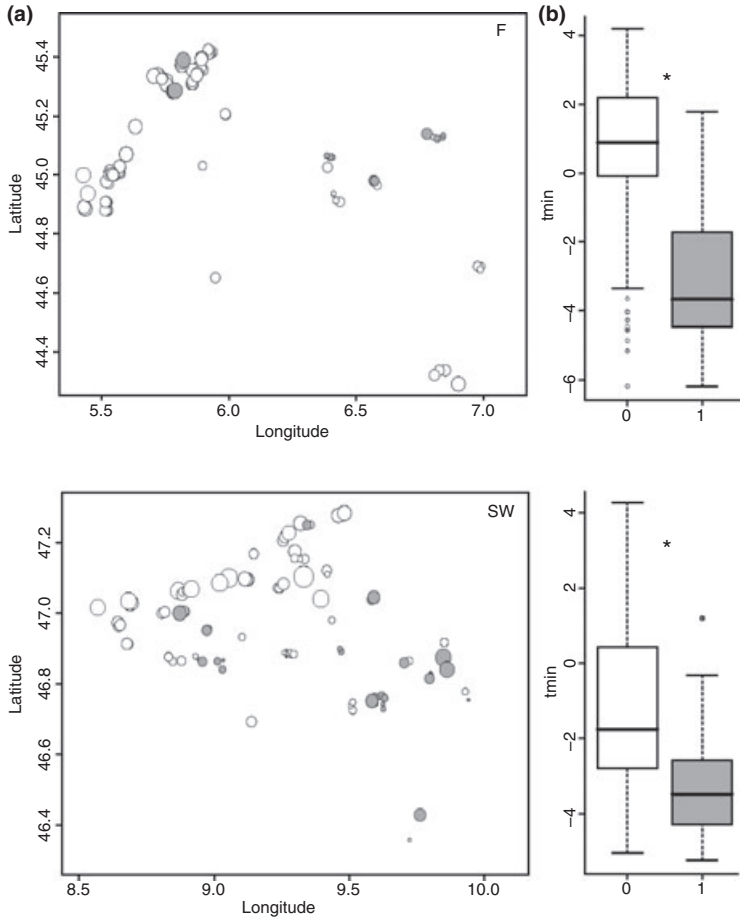

Fig. 5 Correlation between alleles at the ecologically relevant locus PM251.7 of *Arabis alpina* and annual mean minimal temperature (tmin). (a) Spatial variation of tmin overlaid with the geographic distributions of this locus in the French (F) and the Swiss Alps (CH). Grey shading indicates the mean fragment presence at this locus while white refers to mean fragment absence per location. The diameter of the circle is proportional to the tmin value. (b) Box plot of tmin (degrees) in sampling locations where AFLP fragments were absent (0) or present (1), respectively, at this locus in the French (F) and Swiss Alps (CH). Stars indicate the significance of *t*-tests.

(Fig. 5a). For this locus, tmin was significantly smaller when the restriction fragment was present than when it was absent (*t* test, $t = 6.7279$, df = 119, $P = 1.899 \times 10^{-6}$ in the French Alps and $t = 8.6567$, df = 124, $P = 2.158 \times 10^{-14}$ in the Swiss Alps; Fig. 5b). PM342.7 and PM212.8 were both correlated with twi and prcp. The fourth locus, EM125.1, was correlated with twi and tmin. Pairwise correlations between the distributions of these four ERLs showed that the distribution of locus PM251.7 was significantly correlated to the distribution of locus EM125.1 (logistic regression, $P = 3 \times 10^{-4}$). This association also holds true for the two loci PM342.7 and PM212.8 (logistic regression, $P = 2 \times 10^{-7}$).

Discussion

Divergent selection in relation to environmental variation is a key process in the evolution of populations. As a result of recent technological advances in molecular methods, studying local adaptation using genome scans has received considerable attention, which also helps to better understand the responses of organisms to global change.

In this study, we identified 78 ERLs whose distribution was significantly related to the environmental variables tested, in particular to temperature (43 loci in the French Alps and 10 in the Swiss Alps) and precipitation (21 loci in the French Alps and eight in the Swiss Alps). Four of these loci were identified in both regions considered as independent spatial replicates. Clinal patterns of genetic variation observed in loci along environmental gradients have usually been interpreted as being caused by natural selection (Endler 1977; Schmidt *et al.* 2008). Therefore, such loci have been interpreted as being located in or being linked to candidate genes. We have strong confidence in the four loci

© 2010 Blackwell Publishing Ltd

identified in our study, because they were found to be of ecological relevance in two independent regions. However, we lack the experimental evidence to securely interpret these loci as being genomic regions under selection, although the use of an appropriate statistical approach gives strong support to this conclusion. Here, we will first discuss the innovative character and the reliability of our approach, but also its limitations, before interpreting our ERLs in an ecological context.

Previous studies on ERLs in plant species found similar patterns of genetic variation in allozyme frequencies along environmental gradients (Allard *et al.* 1993; Linhart & Grant 1996; Prentice *et al.* 2000). For example, Hirao & Kudo (2008) found a correlation between allozyme frequency and a flowering time trait in *Primula cuneifolia* along a snowmelt gradient. Shimono *et al.* (2009) identified morphological traits covarying with allozyme frequencies of fellfield and snowbed populations in *Potentilla matsumurae*, suggesting that the timing of snowmelt applies selective pressure, thus driving local adaptation in this alpine plant species. Recent technological advances have made large-scale genome scans available, optimizing the identification of loci potentially of ecological relevance. This is particularly important for our study, in which we tested the excessively large number of 825, mostly unlinked AFLP loci for their relationship with environmental variation.

We analysed our data based on individuals (typically three individuals per site), since we aimed at including many locations and a broad range of environmental conditions. In consequence, we did not use recent population genetic approaches searching for correlations between allele frequencies and climate variables (Yu *et al.* 2006; Hancock *et al.* 2008; Manel *et al.* 2010b), which are able to correct for population structure. Instead, we applied a correlative individual-based approach, GEE, including the possibility to detect false positive signatures of selection due to historical or spatial effects. Therefore, we used additional biological criteria to separate true selection signatures from those of population structure or history. Firstly, the population structure analysis showed that the French and Swiss regions can be considered as independent from a historical perspective, i.e. genetic drift exceeds migration. Both regions were characterized by significant isolation by distance, with gene flow occurring over small distances and decreasing over larger geographic distances. Identifying joint ERLs in the two independent and well differentiated regions (F_{ST} = 0.1652) is a way of preventing interference from purely historical effects (Endler 1986; Vasemägi 2006; Schmidt *et al.* 2008); an ideal situation (genome-wide vs. locus-specific clines) for detecting adaptive genetic variation. Secondly, our study regions did not represent admixture zones according to STRUCTURE analysis (Fig. 2), although previous phylogeographic analyses of *A. alpina* showed that the Swiss sampling area was in the vicinity of a contact zone of two evolutionary lineages in the central Alps (Assefa *et al.* 2007; Ehrich *et al.* 2007; Ansell *et al.* 2008). Finally, the isolation by distance patterns detected in both regions (Fig. 3) suggested spatial autocorrelation between sampling locations, for which GEE does not correct. The consequence of spatial autocorrelation between sampling locations in regression models has recently been discussed by Diniz-Filho *et al.* (2009) in the context of genetic data: interdependence among samples disturbs the significance of tests and reduces the power of the analysis. In our case, however, spatial autocorrelation decreased rapidly and was not relevant for comparisons of distant locations, reflecting recurrent gene flow only between close locations.

The choice, scale and precision of the environmental variables to be included are crucial criteria in landscape genomic studies. In our case, only variables showing a substantial gradient, or at least large variation in space, were included in the analysis to detect adaptive genetic variation (Etterson 2004). As a result, we did not consider different substrate types, which have been shown to be an important driver of genetic patterns in alpine plants (Choler *et al.* 2004; Alvarez *et al.* 2009); in any case *Arabis alpina* mainly occurs on calcareous substrate. On small spatial scales, it is appropriate to use finer measurements of environmental variables (e.g. Linhart & Grant 1996), but it is not always possible to obtain local measurements over extended areas. Therefore, climatic and topographic variables obtained from GIS databases are relevant alternatives (Joost *et al.* 2007; Manel *et al.* in revision).

Out of the 78 ERLs identified in the two study regions, four loci were correlated with at least one environmental variable in both the French and the Swiss Alps. One of them (locus PM251.7) was correlated with tmin in both study regions and showed higher allele occurrence when minimum temperatures were smaller (Fig. 5b). The spatial distribution of allele presence at this locus and of minimum temperatures in the French and the Swiss Alps showed that in France, the tmin effect was restricted to the Chartreuse, Cerces and Thabord areas (Fig. 5a), suggesting local adaptation on the massif scale, whereas the restriction fragment was present at the coldest sites, i.e. at the highest altitudes in Switzerland. The scale effect of adaptive genetic variation (i.e. local versus large scale) is discussed in a recent analysis on *A. alpina* (Manel *et al.* 2010a). The three other ERLs were not correlated with the same environmental variable in both study regions. As adaptation may include complex genetic responses, these markers might nevertheless show signs of selection. Linkage

© 2010 Blackwell Publishing Ltd

disequilibrium between some pairs of the four loci suggested that loci PM251.7 and EM125.1 and loci PM342.7 and PM212.8 might be located on the same linkage groups, respectively, being associated with the same environmental variables except for EM125.1 linked to twi in the French Alps.

In our study, the environmental variable that explained most of the allele distributions was tmin (Table S2, Supporting information). Minimal temperatures are a well-known selective pressure in mountain and alpine plants. Indeed, freezing temperatures constitute a major limitation to plant growth (Körner 2003), productivity and distribution (Boyer 1982). Freezing tolerance is a complex process involving a number of biochemical and physiological adaptations under genetic control (Hughes & Dunn 1996). Precipitation was also often correlated with allele occurrence in our study, but a direct link of the identified markers of ecological relevance to drought is not obvious, as adaptation to drought involves hundreds of genes (Chaves *et al.* 2003) and as moisture availability is considered to be a complex environmental trait along altitudinal gradients (Körner 2007).

The next step would be to link the identified ERLs to traits potentially involved in adaptation to the corresponding selective pressure. For alpine plants, the physiological traits involved in adaptation should mainly be linked to key growth-related traits (Chapin *et al.* 1993), flowering time (Hirao & Kudo 2004) and vernalization (Hopkins *et al.* 2008). Genes under natural selection are studied in the related model plant *Arabidopsis thaliana* (Mitchell-Olds & Schmitt 2006; Chiang *et al.* 2009; Reininga *et al.* 2009) and have started to be investigated in *A. alpina* (Amasino 2009; Wang *et al.* 2009). Some functional traits have been studied in a landscape perspective, especially with regard to the genetic control of flowering and the perennial life history trait (Johanson *et al.* 2000; Wang *et al.* 2009). Perpetual flowering (*PEP1*) genes contribute to three perennial traits because they limit the duration of flowering, facilitate a return to vegetative growth, prevent branches from undergoing transition to flower production, allow polycarpic growth and confer an adaptive response to low winter temperatures, which restricts flowering to spring. Such genes might well be involved in the adaptive response of *A. alpina*, as we found ERLs correlated to minimum temperatures, which in turn relate to the length of the flowering period in alpine environments (Körner 2007). Further studies should therefore attempt to identify the genomic regions of *A. alpina* containing the ERLs by sequencing the corresponding AFLP fragments and comparing these sequences with the *A. thaliana* genome (The Arabidopsis Genome Initiative 2000). This could give a first indication of the genes involved as well as their functionality (Nielsen 2005; Stinchcombe & Hoekstra 2008). Validation is necessary to determine if the identified ERLs are truly under divergent selection or only false positives, and we are currently developing experiments to investigate the influence of minimal temperatures in the adaptation of *Arabis alpina*.

In conclusion, our results demonstrate the potential of using genome-wide marker surveys in conjunction with allele distribution models to reveal signatures of natural selection along environmental gradients. However, combining the information from genome scans with other approaches (fitness variation along gradients, genomic characterization of ERLs, transplant experiments; Feder & Mitchell-Olds 2003; Kawecki & Ebert 2004; Holderegger *et al.* 2008) may substantiate the genetic and functional architecture behind complex adaptive traits.

Acknowledgements

We would like to thank Rolland Douzet, Christian Miquel, Claire Redjadj, Sabine Brodbeck, Annina Bürgi, Nathalie Baumgartner, Fabio Rimensberger, Conny Thiel-Egenter and René Graf for helping with sampling and Irène Till-Bottraud, Laurence Després, Aurélie Bonin, Philippe Choler and two anonymous referees for their constructive comments on the manuscript. B.N. Poncet was funded by the French Ministry of Research. The contributions of F. Gugerli and R. Holderegger were associated to the CCES-BIOCHANGE project of the ETH domain.

References

Allard RW, Garcia P, Saenz-de-Mierat LE, Pérez de la Vega M (1993) Evolution of multilocus genetic structure in *Avena hirtula* and *Avena barbata*. *Genetics*, **135**, 1125–1139.

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.

Amasino R (2009) Floral induction and monocarpic versus polycarpic life histories. *Genome Biology*, **10**, 228.

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.

Assefa A, Ehrich D, Taberlet P, Nemomissa S, Brochmann C (2007) Pleistocene colonization of afro-alpine 'sky islands' by the arctic-alpine *Arabis alpina*. *Heredity*, **99**, 133–142.

Beven K, Kirkby MJ (1979) A physically-based variable contribution area model of catchment hydrology. *Hydrology Science Bulletin*, **24**, 43–69.

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441–469.

Bonin A, Bellemain E, Eidesen PB *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a

© 2010 Blackwell Publishing Ltd

gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

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.

Boyer JS (1982) Plant productivity and environment. *Science*, **218**, 443–448.

Carl G, Kuhn I (2007) Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *Ecological Modelling*, **207**, 159–170.

Chapin FS, Autumn K, Pugnaire F (1993) Evolution of suites of traits in response to environmental stress. *American Naturalist*, **142**, S78–S92.

Chaves MM, Maroco JP, Pereira JS (2003) Understanding plant responses to drought—from genes to the whole plant. *Functional Plant Biology*, **30**, 239–264.

Chiang GCK, Barua D, Kramer EM, Amasino RM, Donohue K (2009) Major flowering time gene, FLOWERING LOCUS C, regulates seed germination in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA*, **106**, 11661–11666.

Choler P, Erschbamer B, Tribsch A, Gielly L, Taberlet P (2004) Genetic introgression as a potential to widen a species' niche: insights from alpine *Carex curvula*. *Proceedings of the National Academy of Sciences, USA*, **101**, 171–176.

Diggle P, Elliott P (1995) Disease risk near point sources: statistical issues for analyses using individual or spatially aggregated data. *Journal of Epidemiology and Community Health*, **49**, S20–S27.

Diniz-Filho JAF, Nabout JC, de Campos Telles MP *et al.* (2009) A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genetics and Molecular Biology*, **32**, 203–211.

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.

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton.

Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton.

Etterson JR (2004) Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. 1. Clinal patterns of selection along an environmental gradient in the Great Plains. *Evolution*, **58**, 1446–1458.

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

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature Reviews Genetics*, **4**, 649–655.

Hancock AM, Witonsky DB, Gordon AS *et al.* (2008) Adaptation to climate in candidate genes for common metabolic disorders. *PLoS Genetics*, **4**, e32.

Hardy O (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology*, **12**, 1577–1588.

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

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

Hirao AS, Kudo G (2008) The effect of segregation of flowering time on fine-scale spatial genetic structure in an alpine-snowbed herb *Primula cuneifolia*. *Heredity*, **100**, 424–430.

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.

Hopkins R, Schmitt J, Stinchcombe JR (2008) A latitudinal cline and response to vernalization in leaf angle and morphology in *Arabidopsis thaliana* (Brassicaceae). *New Phytologist*, **179**, 155–164.

Hughes MA, Dunn MA (1996) The molecular biology of plant acclimation to low temperature. *Journal of Experimental Botany*, **47**, 291–305.

Johanson U, West J, Lister C *et al.* (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. *Science*, **290**, 344–347.

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.

Kawecki T (2008) Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 321–342.

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

Körner C (2003) *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems*. Springer, Heidelberg.

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

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

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

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Manel S, Poncet BN, Legendre P *et al.* (2010a) Common factors drive genetic variation of adaptive relevance at different spatial scales in *Arabis alpina*. *Molecular Ecology*, doi: 10.1111/j.1365-294x.2010.04716.x.

Manel S, Joost S, Epperson BK *et al.* (2010b) Perspectives on the use of geo-environmental data in landscape genetics in the face of global change. *Molecular Ecology*, doi: 10.1111/j.1365-294x.2010.04717.x.

© 2010 Blackwell Publishing Ltd

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

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature*, **441**, 947–952.

Namroud MC, 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. *Molecular Ecology*, **17**, 3599–3613.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–127.

Pan W (2001) Model selection in estimating equations. *Biometrics*, **57**, 529–534.

Parisod C, Christin PA (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist*, **178**, 436–447.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.

Prentice HC, Lönn M, Lager H, Rosén E, van der Maarel E (2000) Changes in allozyme frequencies in *Festuca ovina* populations after a 9-year nutrient/water experiment. *Journal of Ecology*, **88**, 331–347.

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

R Development Core Team (2007) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Reininga JM, Nielsen D, Purugganan MD (2009) Functional and geographical differentiation of candidate balanced polymorphisms in *Arabidopsis thaliana*. *Molecular Ecology*, **18**, 2844–2855.

Reusch TBH, Wood TE (2007) Molecular ecology of global change. *Molecular Ecology*, **16**, 3973–3992.

Savolainen V, Anstett M-C, Lexer C *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature*, **441**, 210–213.

Schlötterer C (2003) Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends in Genetics*, **19**, 32–38.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences, USA*, **106**, 9955–9962.

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

Scotti-Saintagne C, Mariette S, Porth I *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species (*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.). *Genetics*, **168**, 1615–1626.

Shimono Y, Watanabe M, Hirao AS, Wada N, Kudo G (2009) Morphological and genetic variations of *Potentilla matsumurae* (Rosaceae) between fellfield and snowbed populations. *American Journal of Botany*, **96**, 728–737.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.

Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **64**, 479–498.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **534**, 796–815.

Vasemägi A (2006) The adaptive hypothesis of clinal variation revisited: single-locus clines as a result of spatially restricted gene flow. *Genetics*, **173**, 2411–2414.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

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

Wang RH, Farrona S, Vincent C *et al.* (2009) PEP1 regulates perennial flowering in *Arabis alpina*. *Nature*, **459**, 423–U138.

Wark AR, Peichel CL (2010) Lateral line diversity among ecologically divergent threespine stickleback populations. *Journal of Experimental Biology*, **213**, 108–117.

Wright S (1938) Size of population and breeding structure in relation to evolution. *Science*, **87**, 430–431.

Yan J, Fine JP (2004) Estimating equations for association structures. *Statistics in Medicine*, **23**, 859–880.

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

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

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Description of *Arabis alpina* sampling locations in the French and the Swiss Alps

Table S2 Results of GEE analyses in *Arabis alpina* from the French and the Swiss Alps

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2010 Blackwell Publishing Ltd