Review

Cell PRESS

Landscape genetics of plants

Rolf Holderegger[1,2], Dominique Buehler[1,2], Felix Gugerli[1] and Stéphanie Manel[3,4]

[1] WSL Swiss Federal Research Institute, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
[2] ETH Zürich, Department of Environmental Sciences, Universitätsstrasse 16, CH-8092 Zürich, Switzerland
[3] Laboratoire Population Environnement Développement, UMR 151 UP/IRD, Université de Provence, 3 place Victor Hugo, 13331 Marseille Cedex 03, France
[4] Laboratoire d'Écologie Alpine (LECA), CNRS UMR 5553, Université Joseph Fourier, BP 53, 2233 Rue de la Piscine, 38041 Grenoble Cedex 09, France

Landscape genetics is the amalgamation of landscape ecology and population genetics to help with understanding microevolutionary processes such as gene flow and adaptation. In this review, we examine why landscape genetics of plants lags behind that of animals, both in number of studies and consideration of landscape elements. The classical landscape distance/resistance approach to study gene flow is challenging in plants, whereas boundary detection and the assessment of contemporary gene flow are more feasible. By contrast, the new field of landscape genetics of adaptive genetic variation, establishing the relationship between adaptive genomic regions and environmental factors in natural populations, is prominent in plant studies. Landscape genetics is ideally suited to study processes such as migration and adaptation under global change.

Glossary

AFLPs: amplified fragment length polymorphisms. Dominant and anonymous DNA fingerprints.
Allele distribution model: a (usually non-spatial) statistical description of how allele frequencies at loci linked to genes under selection are influenced by distinct environmental factors.
Assignment test: statistical approach that assigns an individual to that sampled population from which its multilocus genotype is most likely to be derived.
Bayesian inference: statistical approach using prior data or information to estimate posterior probabilities of a hypothesis to be correct. For instance, prior information (or just a guess) on migration rates can be used in Bayesian assignment tests to infer contemporary gene flow.
Chord distance (d_c): a measure of genetic similarity between individuals or populations based on allele frequencies and located on a sphere. Taking values between 0 and 1.
Gene flow: exchange of genes among populations or individuals.
Genome scan: genotyping of many samples at a large number of (potentially anonymous) molecular markers across the genome, used in outlier detection and the landscape genetic analysis of adaptive genetic variation.
Hardy-Weinberg equilibrium: population genetic principle stating that allele and genotype frequencies reach equilibrium and stay constant in random mating populations assuming large population size, no selection, no migration and no mutation.
Isolation by distance: spatial pattern describing decreasing genetic relatedness of populations or individuals with increasing geographic distance.
Kriging: geostatistical technique to interpolate the value of a parameter of interest at an unobserved geographic location from observations of this value at nearby locations.
Landscape: an area spatially heterogeneous in one or more biotic and abiotic factors of interest. From the human perspective, a landscape is perceived as a kilometers-wide environmental mosaic.
Landscape distance: distance-like measurements parameterizing the landscape between two localities, for example, geographical distance along a river, number of roads to be crossed, or percentage of forest cover in a corridor strip connecting two localities.
Landscape resistance: permeability values of different landscape elements, describing their resistance to migration and dispersal. From these resistance values, different types of cumulative resistances between localities can be calculated, for example, least cost paths.
Least cost path: length of a path minimizing the cumulative landscape resistance between two localities.
Mantel test: a permutation-based statistical test describing the correlation between two distance matrices. A partial Mantel simultaneously accounts for the effects of a third (or several) distance matrix (matrices).
Monmonier's algorithm: detects genetic boundaries by finding the path exhibiting the largest genetic distances among neighboring populations.
Nei's genetic distance (d_n): measure of genetic similarity based on the probability that two randomly chosen alleles from different populations or individuals are identical. Taking values between 0 and 1.
Neutral molecular marker: molecular markers not affected by natural selection.
PCNMs: principal coordinates of neighbor matrices are a spectral representation of all spatial relationships among sampling locations and describe all spatial scales that can be accommodated by the sampling design. PCNMs are calculated from principal coordinate analysis (PCoA). In landscape genetics, they are used to account for spatial relationships among sampling locations and for unaccounted environmental factors.
Permutation test: type of statistical tests that rely on resampling of data for significance testing and not on theoretical probability distributions as in classical statistics. For instance, permutation tests are used to account for the non-independence of genetic, geographic and landscape distances among sampling locations in landscape genetics.

The booming field of landscape genetics

Landscape genetics is the amalgamation of landscape ecology and population genetics to help with understanding microevolutionary processes such as gene flow and adaptation on the scale of natural landscapes [1]. This field investigates how landscape elements and environmental factors influence the spatial distribution of genetic variation. For instance, landscape genetics assesses how landscape elements such as forests or open fields affect gene flow (see Glossary) in species inhabiting semi-natural habitat remnants in an otherwise unsuitable and intensively used landscape; a question of interest in conservation management. Similarly, landscape genetics analyzes whether non-crop strips provide effective barriers to gene flow between organic and genetically modified crops; a question of practical importance in agriculture. Landscape genetics also deals with how environmental factors such as temperature or precipitation affect adaptive genetic variation; relevant information in the context of climate change [2,3]. Landscape genetics still suffers from a lack of theoretical foundations and expectations [4–6], but empirical studies are typically characterized by including geo-referenced individuals or populations genotyped at multiple loci and at least one landscape or environmental variable of interest (in addition to geographical distance) measured at or in between sampling locations [7]. Depending on the landscape or environmental variables assessed, the study area might be large (thousands of km^2), for example, when studying the adaptive response of a dominant forest tree to

---

Corresponding author: Holderegger, R. (rolf.holderegger@wsl.ch).

doi:10.1016/j.tplants.2010.09.002 Trends in Plant Science, December 2010, Vol. 15, No. 12

Population differentiation (F_{ST}): different measurements of the amount of genetic variation found between populations. Most often used is Wright's F_{ST}, taking values between 0 and 1. Similar measurements can be calculated for genetic differentiation among groups of populations (F_{CT}).

SNPs: single-nucleotide polymorphisms. Bi-allelic, co-dominant molecular markers of known position in the genome. Increasingly used in landscape genetics.

Wombling: approach to search for areas across an interpolated allele frequency surface where the slopes of the surface are particularly steep. Used to infer genetic breaks or discontinuities.

latitude, or small (hundreds of m^2), for example, when assessing pollen dispersal in a forest herb.

Given its appeal to both basic and applied sciences, landscape genetics has received increasing attention in recent years. The field currently deals with two main topics. First, landscape and environment are evaluated by considering their effects on migration, dispersal and gene flow, which are measured in terms of neutral genetic variation [7]. Second, landscape genetics has started to explore the interaction between environment and adaptive genetic variation in natural populations and individuals, a new field often referred to as landscape genomics [8–10].

Landscape genetics of plants

A recent survey [7] showed that the majority of landscape genetic studies have dealt with animals and only seldom with plants (~18%). Furthermore, studies on animals and plants differ in study design and analytical approaches. Why are there such differences in landscape genetic studies on plants and animals?

When dealing with gene flow, landscape genetics considers the landscape between sampling locations. When studying adaptive genetic variation, landscape genetics deals with the particular environment at sampling locations [11]. These two situations make plants either less or more amenable to landscape genetic analysis than animals. Because sessile plants directly respond to the environment at their growing site, the study of adaptive genetic variation in plants is straightforward. By contrast, corresponding animal studies must account for the environment of the entire home range or of all resource sites of individuals or populations. However, when assessing migration, dispersal and gene flow, studying plants is more complex than investigating animals. In animals, individuals disperse to other locations and provide gene flow when mating. In vascular plants, gene flow mainly happens through two processes, the dispersal of diploid embryos in seeds and of haploid male gametes in pollen. In wind-pollinated and -dispersed plants, an abiotic factor acts as the primary pollen and seed vector. In insect-pollinated and animal-dispersed plants, animals act as dispersal vectors of propagules. Here, it is a moving animal that reacts to the landscape and not the plant itself. Landscape genetic studies investigating gene flow in plants thus deal with the problem that two propagule types are dispersed by particular vectors. For instance, the fragmentation effect of roads on plants is not direct, but induced by the indirect effects of roads on pollinators or animal seed dispersers, whereas wind-pollinated and -dispersed plants might not be affected by roads [12].

Published reviews on landscape genetics have focused on animals, and many plant studies have not included landscape elements, apart from geographical distance. Given the characteristics of plants and the present shortage of empirical studies, it seems relevant to provide an overview focusing on plants and stressing associated benefits and limitations of common landscape genetic approaches. In this review, we will first summarize and discuss the main approaches currently used to study gene flow at the landscape scale: the landscape distance/resistance approach, the overlay technique and the assessment of contemporary gene flow. We will then examine landscape genetic approaches exploring adaptively relevant genetic variation [13,14].

Gene flow on the landscape scale

Landscape distance/resistance

This classical approach correlates a matrix of genetic distances as indirect measurements of gene flow [15] with matrices of landscape distances/resistances and geographic distances (Figure 1a). The genetic matrix consists of pairwise genetic distances among all pairs of individuals or populations studied. Various estimators can be used for this purpose, for example, genetic Chord distance (d_c), Nei's distance (d_n) or population differentiation (F_{ST}) [7,16]. The geographic distance matrix contains the straight line distances among all sampling locations. By contrast, the landscape distance matrix varies and can contain the length, area or percentage of cover of landscape elements, such as ditches, wetlands or woodlands in corridors of a certain width between pairs of sampling locations. It can also be a 0/1 matrix, for example, when some sampling locations are separated by a river and others are not. In a more sophisticated design, land cover/land use and topography are taken from existing geographic information system (GIS) data or from field surveys, and a level of resistance to movement is given to each raster cell containing a particular landscape feature. For instance, forests might hinder gene flow whereas open fields have no effect, and a high resistance is thus assigned to forested grid cells, whereas open-field cells receive low resistance values. Using GIS technology, the length of the shortest path connecting two sampling locations is determined by maximizing movement through low-resistance cells [17]. Such least-cost paths form the entries for a landscape resistance matrix. Resistance assignment relies on expert knowledge or *a priori* ecological information on study organisms [18], and several alternative landscape resistance models are generally tested per study.

For statistical analysis, the genetic distance matrix is either correlated to the geographic distance and landscape distance/resistance matrices separately using a Mantel test, or the effect of geographic distance is first partialed out before estimating correlation with landscape distance/resistance in a partial Mantel test [19] (Figure 1a). This procedure enables the effects of geographic distance (i.e. isolation by distance [20]) and landscape elements on gene flow to be disentangled [21]. However, partial Mantel tests have been criticized because of their permutation procedure (e.g. [22]), and a variety of alternative approaches have been suggested [23,24].

The landscape resistance approach and especially least cost path analysis have been popular in animal studies, but there are few such studies in plants [7] (but see [25,26]).

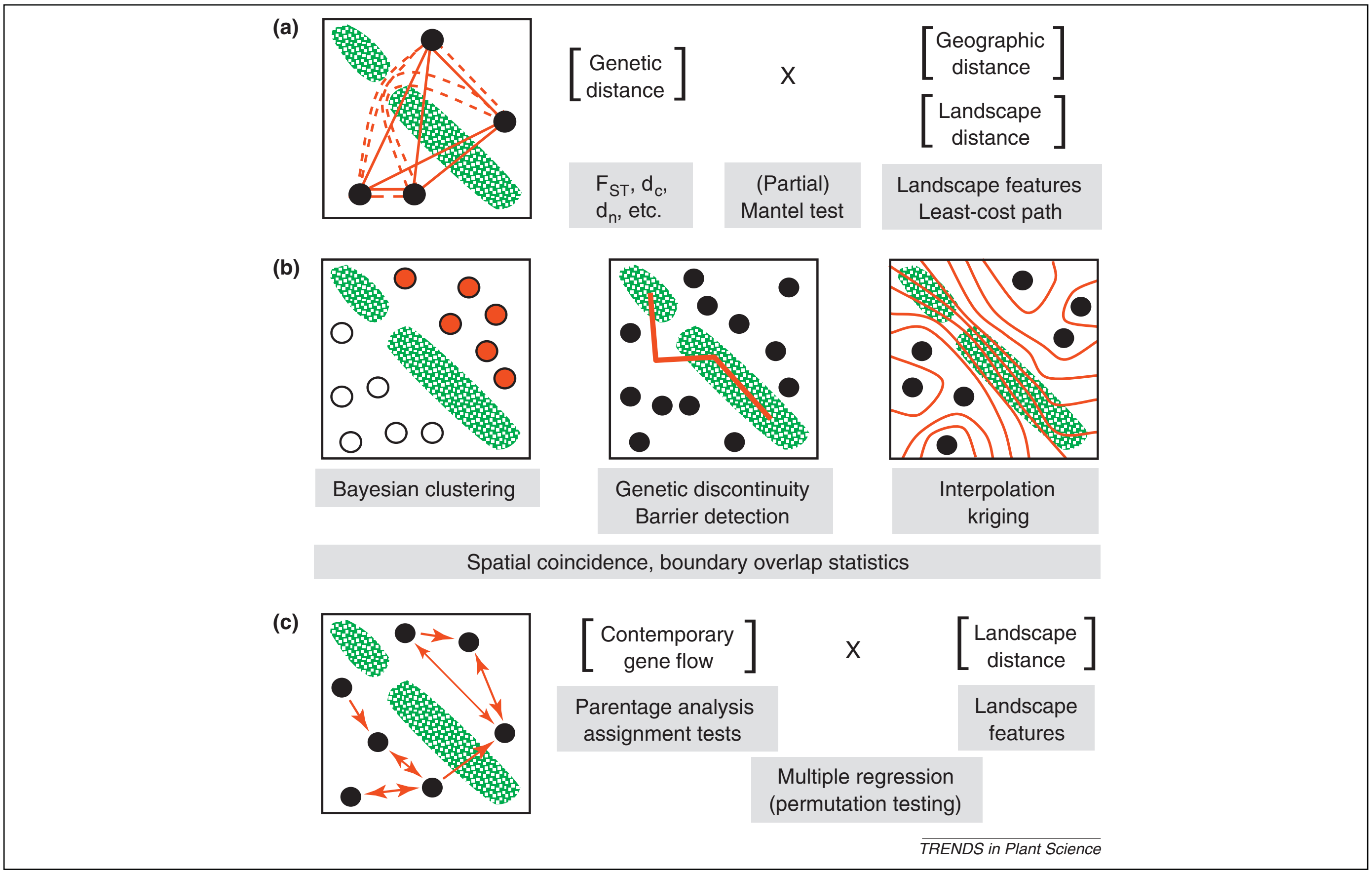

Figure 1. Schematic summary of the three major analytical approaches that are currently used by landscape geneticists to study gene flow. **(a–c)** Individuals or populations (circles) are studied in a simple landscape consisting of two land cover types: meadows (open) and forests (hatched green). **(a)** In the landscape distance/resistance approach, a matrix of genetic distances between all pairs of individuals (e.g. Nei's genetic distance d_n, Chord distance d_c [17]) or between all pairs of populations (e.g. genetic differentiation F_{ST} [16]) is correlated with a matrix of geographic distances between sampling locations (solid red straight lines) and landscape distances (landscape features such as percentage of forest between sampling locations) or landscape resistance (e.g. length of least-cost path [17] assuming a high resistance value of forests to gene flow; broken red lines) in Mantel or partial Mantel tests [7]. **(b)** The overlay technique uses several methods to cluster individuals into groups (e.g. Bayesian clustering; groups indicated by open and red circles [5]), to detect genetic discontinuities or barriers (thick red line [36]) or to interpolate genetic distances among individuals or populations (e.g. kriging resulting in genetic isolines given in red [1]). These genetic groups, barriers, genetic discontinuities or isolines are overlaid on topographical or land cover/land use maps to search for spatial coincidences of these genetic structures with landscape elements. **(c)** Contemporary gene flow events or migrants (red arrows indicating direction and abundance) can be assessed by either parentage analysis (paternity or maternity analysis [58]) or assignment tests [36]. Any number of matrices of landscape distances between sampling locations connected by migration or gene flow is correlated with a matrix of the frequency of contemporary gene flow in multiple linear regression using permutation for significance testing. Note that in all of the above examples, forests are hindering gene flow.

The approach is intuitively appealing in animals, but less so in plants because pollen and seed dispersal mechanisms depend on (multiple) biotic or abiotic vectors. In addition, genetic distances in plants do not account for differences in seed and pollen dispersal [16].

Given these shortcomings of the landscape distance/resistance approach in plants, it is not surprising that it has mainly been applied in special situations. Mantel and partial Mantel tests of genetic distances with geographic distances along coasts or streams have been used in studies on wild sugar beet (*Beta vulgaris* ssp. *maritima* [27]) or Japanese primrose (*Primula sieboldii* [28]). Similarly, the effect of differences in flowering time among individuals on genetic distances has been studied with partial Mantel tests in several alpine snowbed plants [29,30]. In fact, researchers can still profit from the landscape distance/resistance approach in plants when analyzing the influence of major landscape elements on gene flow in a general way (Box 1). For instance, forests might act as barriers to gene flow in a wind-pollinated and wind-dispersed meadow herb. The occurrence of large forests between populations or individuals should generally exhibit a negative effect on gene flow and effectively increase genetic distances. Researchers could also simultaneously determine the effects of several landscape distances between pairs of sampling locations (including geographic distance) on genetic distances in multiple linear regression with permutation-based significance testing using software such as PERMUTE [31] or BLOSSOM [32]. When testing several models of landscape distance/resistance, significance values should be adjusted because of multiple testing, and model performance has to be evaluated (Box 2) [33], which has rarely been done in landscape genetics (but see [34,35]).

Overlay technique

The overlay approach of landscape genetics [7] identifies population groups, barriers, genetic discontinuities or isolines. These genetic structures are overlaid onto maps of selected landscape elements such as topography or land

Box 1. Five current hot topics in landscape genetics of plants

(i) Perform landscape distance/resistance analysis to detect the influence of major landscape elements on gene flow by seed and pollen, that is correlate genetic distance with landscape elements (e.g. forests, mountain ridges) or abiotic factors (e.g. wind direction), and make use of statistical methods alternative to (partial) Mantel tests.
(ii) Expand the use of overlays in plants and incorporate boundary overlap statistics.
(iii) Combine estimates of contemporary gene flow by pollen or seed with landscape data in multiple regression analysis with permutation testing.
(iv) Evaluate the influence of spatial genetic structure and population history in outlier detection or allele distribution analyses.
(v) Describe and prove molecular function of identified outlier genomic region and provide empirical tests of the selective relevance of identified adaptive genetic markers in plants (e.g. transplant experiments).

Box 2. Landscape genetic approach to identify molecular markers bearing the signature of natural selection

Landscape genetics tries to identify molecular markers whose changes in allele frequencies are correlated with environmental factors potentially acting as selective pressures and enforcing directional natural selection (Figure 2a). Usually, many samples are taken along environmental gradients or in environmentally heterogeneous situations, and large genome scans with hundreds to thousands of co-dominant or dominant markers (AFLPs or SNPs [16]) are performed. Finally, allele presence/absence (individual-based analysis) or allele frequencies (population-based analysis) are correlated with environmental variables taken from geo-referenced databases or from field surveys [10]. Markers significantly correlated with environmental factors are considered to be linked to or to be located within genomic regions under selection, whereas uncorrelated molecular markers are considered as neutral, at least with respect to the particular set of environmental variables tested.

Various statistical methods are used to establish allele distribution models [9]. First, logistic regression relates allele occurrences with environmental variables. The spatial analysis method (SAM [87]) offers a user-friendly framework to perform logistic regression in a landscape genetic approach. SAM has successfully been used in several animal studies [14,97], but we are aware of only a single plant study that has applied SAM. Parisod and Joost [98] examined patterns of selection in populations of buckler mustard (*Biscutella laevigata*) characterized by different population histories. Logistic regression has also been used to correlate allele frequencies at an outlier locus associated with altitude and temperature in common beech (*Fagus sylvatica*) [99]. The logistic regression approach can easily be extended to more sophisticated generalized linear models [92]. Recently, we successfully applied multiple linear regression to identify AFLP fragments correlated with temperature and precipitation in the alpine rock cress (*Arabis alpina*) [13]. Polynomial transformation of environmental variables can be included in multiple linear regression, thus also tracking non-linear adaptive responses.

The correlative landscape genetic approach to adaptive genetic variation has several shortcomings. (i) Researchers can only identify correlations of molecular markers with those environmental factors that were included in the analysis. Molecular markers not correlated with these environmental factors are therefore not necessarily neutral in a general sense. For this reason, researchers might also wish to apply classical outlier locus detection to their data sets if population-based sampling is available [97]. (ii) Given that a potentially large number of statistical tests is applied in explorative data analysis, it is important to adjust significance values for type I error inflation due to multiple testing and/or using model evaluation criteria such as adjusted R^2 or the Akaike information criterion (AIC) values [33]. The use of adjusted R^2 and AIC is currently explored in landscape genetic studies of plants.

cover/land use to search for geographical coincidences of group boundaries, barriers, genetic discontinuities or isolines with landscape elements (Figure 1b).

Various statistical approaches can be used to form genetic groups of populations or individuals to be used in overlays. Bayesian clustering [36] based on Hardy-Weinberg and linkage equilibrium as implemented in STRUCTURE [37], BAPS-5 [38], TESS [39] or GENELAND [40] are widely used for this purpose. These programs can also consider coordinates of sampling locations [38–40]. Alternatively, non-Bayesian S-AMOVA can be applied to form population groups. This method maximizes genetic differentiation (F_{CT}) among groups [41]. For instance, no or only weak clustering was detected in the alpine blue thistle (*Eryngium alpinum*) [42] and the rainforest Anguama tree (*Aucoumea klaineana*) [43] despite spatial distribution gaps, and grouping according to river catchments was inferred in the Chinese maidenhair fern (*Adiantum reniforme* var. *sinensis*) [44]. Other methods search for areas of strong changes in allele frequencies, such as the Monmonier algorithm implemented in BARRIER [45] or ALLELES IN SPACE [46] to detect genetic barriers among populations (e.g. [26]). Individual- or population-based wombling also identifies genetic discontinuities [1,47,48]. Finally, interpolation such as kriging from principal component analysis (PCA) axis loadings determines genetic isolines similar to contour lines in topographical maps (Figure 1b). For instance, kriging helped in visualizing the small-scale genetic structure in snapdragon (*Antirrhinum microphyllum*) [49]. Membership coefficients estimated from Bayesian clustering methods can also be interpolated, providing clustering surface maps (e.g. [25]).

The major drawback of the common overlay approach is that spatial coincidence of landscape elements with genetic discontinuities, barriers or isolines is simply based on subjective visual inspection. No statistical procedure is usually involved in this step. Overlays are thus of exploratory nature and prone to false inference. However, subjectivity in analysis could be avoided by applying boundary overlap statistics [50,51]. Such statistics have, for example, been used to study the spatial coincidence between heterogeneity of forest structures and boundaries of bird territories [52]. To our knowledge, boundary overlap statistics have rarely been used in landscape genetics (Box 1).

Despite its appeal for studying sessile organisms, the overlay technique has not been popular in plants and should therefore be further explored, especially in connection with appropriate significance testing. In fact, most plant studies have dealt with phylogeographic patterns [53], and only a few have overlaid inferred genetic patterns onto land-cover/land-use or topographical maps. By contrast, obvious effects of various landscape elements on gene flow have been identified with overlays in animal studies (e.g. [54,55]).

Assessment of contemporary gene flow

Landscapes throughout the world are changing at an unprecedented speed [2]. This poses a problem to landscape genetics because the landscape distance/resistance and (most) overlay approaches rely on historical measurements of gene flow (i.e. genetic distance and differentiation [15]). However, when landscapes are changing rapidly, historical gene flow measures tend to reflect the historical rather than the contemporary landscape [56,57]. In such situations, researchers would like to assess contemporary migration and gene flow [6]. To do so, two main approaches are currently available: parentage analysis [58] and assignment tests [36] (Figure 1c).

Parentage analysis, and its variants maternity and paternity analysis, has widely been used in plants, especially trees [59]. Paternity analysis infers contemporary pollen flow from the genetic analysis of open-pollinated offspring of known mothers by using maximum likelihood methods (e.g. CERVUS [60]) or Bayesian inference (MASTERBAYES [61]) to identify the most probable fathers [58]. Similarly, maternity analysis uses maternal seed coat tissue [62] or uniparentally inherited organellar DNA markers (e.g. chloroplast DNA [16]) to detect the most probable mothers of seedlings or trapped seed. Parentage analysis asks for complete sampling of all potential parents in a study area, which logistically limits the study range. This is a major shortcoming of studies using parentage analysis, and therefore landscape-scale studies are rare (but see [63]).

From studies on contemporary gene flow, we have learned that pollen and seed dispersal are more frequent and occur over greater distances than expected from ecological investigations [6]. However, the effect of landscapes has rarely been considered, and most studies simply compared contemporary gene flow patterns in fragmented versus non-fragmented situations (e.g. [64–66]). Researchers have started to assess landscape effects by applying multiple linear regression with permutation testing of the frequency of mating events among pairs of individuals and various landscape elements [31,32] (Figure 1c). We have also compared realized mating patterns with a null model of saturated mating among all individuals studied in a regional population of the service tree (*Sorbus domestica* [67]). One problem with the landscape genetic analysis of data from parentage analysis is that no mating between particular individuals either reflects a real lack of mating or simply a low detection probability because of insufficient sample size in terms of parents and offspring studied.

Alternatively, contemporary migration and gene flow are inferred through assignment tests, which are often used in animals [36]. Assignment tests can discriminate between first generation migrants, migrants during the last generation and recent migrants (during the last few generations) using Bayesian software such as GENECLASS [68] or BAYESASS [69]. Assignment tests on plants are rare, probably because it is difficult to disentangle gene flow by seed and pollen. However, by using particular settings in GENECLASS [68], He *et al.* [70] restricted assignment to contemporary seed dispersal. Two major caveats to be considered when using assignment tests are that although assignment tests enable individual migrants to be identified, migration rates and populations of origin can only be appropriately identified if all populations in a landscape were included in sampling and analysis.

Several studies have detected that contemporary seed dispersal occurs over large stretches of unsuitable habitat, across inhospitable mountain ridges or within river catchments [70–72]. Again, landscape effects have seldom been combined with assignment tests, although dedicated software (BMIR [73]) is available for correlating directional contemporary migration rates based on Bayesian inference with landscape distance data in a multiple linear regression framework. It is evident that additional statistical tools have to be developed to analyze the full breadth of contemporary gene flow and migration in a landscape context (Box 1).

Landscape genetics of adaptive genetic variation

A popular route in landscape genetics of plants has been to correlate population genetic diversity with environmental factors at habitat patches. Researchers have studied whether genetic diversity was related to local soil type, humidity, vegetation structure or management type (e.g. [74–76]). These studies were based on the analysis of neutral markers. However, neutral genetic diversity does not directly relate to adaptive genetic variation [77,78] and mainly reflects local population size [79]. Neutral genetic diversity is only indirectly affected by local environmental factors if these factors influence processes such as gene flow or mating [3]. Therefore, if researchers want to establish the relationship between genetic diversity and environmental factors, they should specifically assess adaptive genetic variation [2].

The first step in analyzing adaptive genetic variation is to identify genomic regions bearing signs of selection. Among various methodological approaches [80], genome scans have been the preferred method in non-model organisms for about a decade. In this analysis, loci showing a higher genetic differentiation among populations (F_{ST}) than expected under neutrality (i.e. outlier loci) are identified out of a large number of loci studied across a genome [81–89]. This population genomic approach has also been applied to plants (e.g. [84,85]). Selected environmental factors are only *a posteriori* correlated with allele frequencies at outlier loci to infer potential selective pressures [9].

By contrast, the landscape genetic approach [8] directly uses environmental data to pinpoint molecular markers linked to or located within genomic regions under selection [9]. Accordingly, many samples are collected along environmental gradients (Figure 2), and large genome scans with hundreds to thousands of amplified fragment polymorphisms (AFLPs) or single-nucleotide polymorphisms (SNPs) are performed [16,86]. Allele occurrence in individuals or allele frequency in populations is subsequently correlated with local environmental conditions, for example, estimates of temperature, precipitation, slope, altitude or habitat type [9,10]. Several statistical methods are used for this purpose, such as simple linear regression [13] or generalized linear models, for example the logistic regression implemented in the spatial analysis method (SAM [87]) (Box 2). Molecular markers significantly correlated with environmental factors having substantial effect size are seen as linked to genomic regions influenced by these

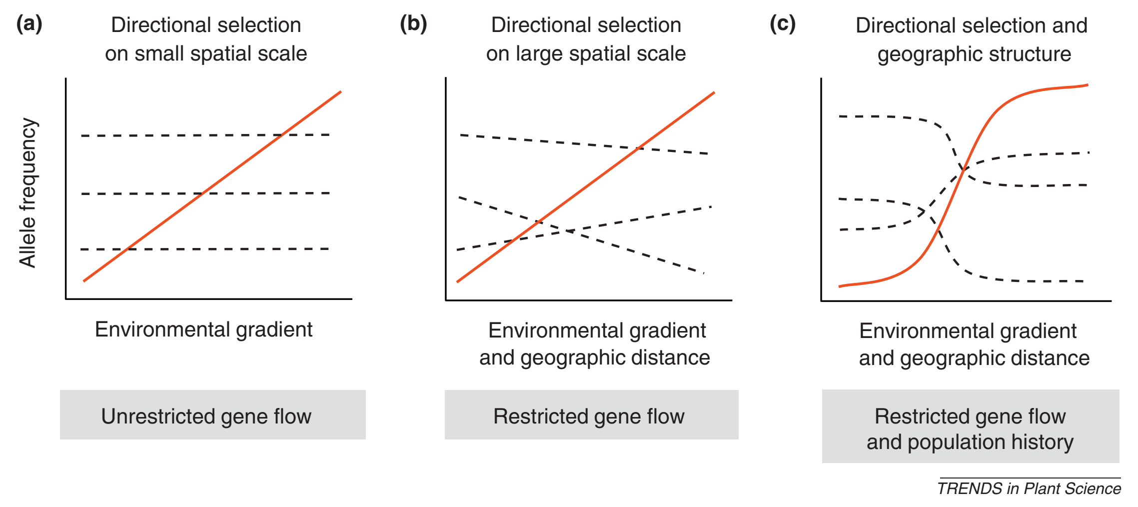

Figure 2. Conceptual characteristics of landscape genetic studies assessing adaptive genetic variation. **(a)** A molecular marker linked to a genomic region under directional selection shows a change in allele frequency along an environmental gradient (solid red line). By contrast, neutral molecular markers show no such change on a small-spatial scale (hence, unrestricted gene flow; broken lines) because they are not affected by natural selection (modified from [96]). **(b)** If directional selection occurs on a larger spatial scale, geographic distance and environmental gradients often co-vary. In consequence, allele frequencies at neutral loci will randomly change with distance and, indirectly, along the environmental gradient because of restricted gene flow and genetic drift (i.e. isolation by distance [7]). Therefore, landscape genetic studies on adaptive genetic variation might falsely identify some neutral markers as bearing the signature of adaptive evolution, that is, spatial genetic structure is a nuisance factor in analysis (Box 3). **(c)** The effect presented in (b) is most prominent if spatial genetic structure due to phylogeography or population history has lead to strong changes in allele frequencies. Here, changes in allele frequencies of markers showing signs of adaptive evolution are expected to be more pronounced than those of neutral loci.

factors. After molecular characterization, identified markers can be used for cross-validation in other landscapes or in experiments aiming at verifying their adaptive or molecular functionality (Boxes 1 and 2) [2,9]. Plants are particularly more amenable to corresponding experimentation than (most) animals. In the model plant thale cress (*Arabidopsis thaliana*), a large SNP genome scan and an in detail genomic analysis found that one particular allele at locus *ACD6* underpins resistance to microbial infection and herbivory in natural populations and therefore provides large fitness advantages under high pathogen and herbivory pressure, despite severely reducing vegetative growth [88].

Only a few landscape genetic studies on the adaptive genetic variation of plants are available so far (Box 2), and the methodological foundations of the field are not yet fully explored. In particular, researchers have only started to deal with the problem of spatial genetic structure [10,89], potentially interfering with landscape genetic analysis and leading to the detection of molecular markers falsely found to be linked to genomic regions under selection (Boxes 1 and 3). Likewise, most studies still lack any indication of the functionality of the outlier loci identified, be it based on experimental or molecular evidence.

Perspectives

Landscape genetics of plants is a largely under-explored field. One reason for this is that applying the classical landscape distance/resistance approach to infer landscape effects on gene flow is less amenable to studies of plants than to animals, owing to the particular means of gene dispersal through pollen and seed. Also the overlay technique has rarely been used in plants in a true landscape genetic setting with sound statistical analysis. Many studies on contemporary gene flow were small-scale and did not consider landscape effects. Therefore, significant progress can be achieved by applying existing methodology to plants on adequate spatial scales. However, it is apparent that the development of new statistical tools is necessary to analyze genetic data in concert with landscape and environmental data [90]. Interesting recent advances include the incorporation of graph theory in landscape genetics [91], which

Box 3. Spatial genetic structure as a nuisance factor in landscape genetic studies on adaptive genetic variation

Spatial genetic structure caused by (i) restricted gene flow and leading to isolation by distance patterns (Figure 2b) or (ii) phylogeographic history, range expansion or population demography (e.g. bottlenecks; Figure 2c) might substantially interfere with both population genomic outlier analysis and landscape genetic approaches. Pronounced genetic structure is thus a nuisance parameter in landscape genetics of adaptive genetic variation [9,89]. Excoffier *et al.* [89] showed that ignoring hierarchical spatial genetic structure in classical outlier detection analysis results in the identification of numerous false outlier loci. This study highlights the need for adequate treatment of spatial genetic structure when searching for molecular markers linked to genes under selection.

The consideration of spatial genetic structure in landscape genetics of adaptive genetic variation has only just begun. So far, researchers have tended to apply Bayesian clustering to their samples and then use landscape genetic analysis with logistic or linear regression [13,84] within each genetic cluster separately. Recently, more sophisticated approaches to account for spatial genetic structure have been introduced. For instance, mixed linear models allow controlling for population structure when populations are known [100]. Accordingly, we used generalized estimating equations (GEE), which take small-scale autocorrelation of samples into consideration [92], and we [13] applied principal coordinates of neighbor matrices (PCNMs [101]) to landscape genetic analysis using R [102]. PCNM values on large- and small-spatial scales can be introduced as additional factors in linear regression and account for the effects of different spatial scales and for the effects of un-accounted environmental factors. Despite such promising new tools, it is obvious that the issue of spatial genetic structure in landscape genomic research needs more attention in future analyses as well as the development of appropriate statistical methods.

considers landscape and environmental data at and between sampling locations. Various types of hierarchical Bayesian models are also gaining increasing popularity in ecology and genetics [11]. A conceptual shortcoming of many landscape genetic studies is the virtual lack of replication at the landscape level (but see [75,92]) and of multi-species studies (but see [91,93]). Replicated landscape genetic analyses require particularly large genetic sample sizes and hence should profit from the current exponential increase in sequencing and genotyping capacity [94]. To deal with the corresponding huge genetic data sets, bioinformatics already relies on machine learning techniques. These techniques enable the screening of large numbers of genetic markers such as SNPs, making them particularly relevant in genomics and the study of adaptive genetic variation [95].

Global change, that is the world-wide alteration of natural and traditionally used landscapes and the rapidly changing climate, demands profound knowledge about the migration ability of species as well as their potential to adapt to new, human-altered environments. Landscape genetics is ideally suited to provide such relevant real-world data [2,9]. For this task to be achieved, researchers have to make full use of existing and newly developed methodological landscape genetic approaches, especially so in plants.

Disclosure statement

The authors declare no conflict of interest.

Acknowledgements

R.H., D.B. and F.G. thank the CCES ENHANCE and BIOCHANGE projects of the ETH domain, the AVE project of the Swiss National Science Foundation (CRSI33 127155/1) and the European ECOCHANGE project (FP6-036866) for financial support. S.M. was supported by the Institut Universitaire de France as a junior member. R.H. and S.M. also acknowledge support by the National Center for Ecological Analysis and Synthesis, funded by NSF (Grant DEB-0553768), the University of California at Santa Barbara, and the State of California. We also thank an anonymous referee for valuable comments on the manuscript.

References

1 Manel, S. *et al.* (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18, 189–197
2 Reusch, T.B.H. and Wood, T.E. (2007) Molecular ecology of global change. *Mol. Ecol.* 16, 3973–3992
3 Ouburg, N.J. *et al.* (2010) An essay on the necessity and feasibility of conservation genomics. *Conserv. Genet.* 11, 643–653
4 Holderegger, R. and Wagner, H.H. (2006) A brief guide to landscape genetics. *Landsc. Ecol.* 21, 793–796
5 Storfer, A. *et al.* (2006) Putting the "landscape" in landscape genetics. *Heredity* 98, 128–142
6 Holderegger, R. and Wagner, H.H. (2008) Landscape genetics. *BioScience* 58, 199–207
7 Storfer, A. *et al.* (2010) Landscape genetics: where are we now? *Mol. Ecol.* 19, 3496–3514
8 Lowry, D.B. (2010) Landscape evolutionary genomics. *Biol. Lett.* 6, 502–504
9 Holderegger, R. *et al.* (2008) Land ahead: using genome scans to identify molecular markers of adaptive relevance. *Plant Ecol. Div.* 1, 273–283
10 Manel, S. *et al.* (2010) Perspectives on the use of landscape genetics to detect adaptive genetic variation in the field. *Mol. Ecol.* 19, 3760–3772
11 Segelbacher, G. *et al.* (2010) Landscape genetics: concepts and challenges in a conservation context. *Conserv. Genet.* 11, 375–385
12 Holderegger, R. and Di Giulio, M. (2010) The genetic effects of roads: a review of empirical evidence. *Basic Appl. Ecol.* 11, 522–531
13 Manel, S. *et al.* (2010) Common factors drive genetic variation of adaptive relevance on different spatial scales in Arabis alpina. *Mol. Ecol.* 19, 3824–3835
14 Joost, S. *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* 16, 3955–3969
15 Whitlock, M.C. and McCauley, D.E. (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity* 82, 117–125
16 Lowe, A.J. *et al.* (2004) *Ecological Genetics. Design, Analysis and Application*, Blackwell
17 Adriaensen, F. *et al.* (2003) The application of "least-cost" modeling as a functional landscape model. *Landsc. Urban Plan.* 64, 233–247
18 Baguette, M. and Van Dyck, H. (2007) Landscape connectivity and animal behaviour: functional grain as a key determinant for dispersal. *Landsc. Ecol.* 22, 117–1129
19 Smouse, P.E. *et al.* (1986) Multiple regression and correlation extensions of the Mantel tests of matrix correspondence. *Syst. Zool.* 35, 627–632
20 Wright, S. (1943) Isolation by distance. *Genetics* 28, 114–138
21 Cushman, S.A. *et al.* (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *Am. Nat.* 168, 486–499
22 Raufaste, N. and Rousset, F. (2001) Are partial Mantel tests adequate? *Evolution* 55, 1703–1705
23 Balkenhol, N. *et al.* (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* 32, 818–830
24 Legendre, P. and Fortin, M-J. (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol. Ecol. Res.* 10, 831–844
25 McRae, B.H. (2006) Isolation by resistance. *Evolution* 60, 1551–1561
26 Trénel, P. *et al.* (2008) Landscape genetics, historical isolation and cross-Andean gene flow in the wax palm, *Ceroxylon echinulatum* (Arecaceae). *Mol. Ecol.* 17, 3528–3540
27 Fievet, V. *et al.* (2007) Spatial analysis of nuclear and cytoplasmic DNA diversity in wild sea beet (*Beta vulgaris* ssp. *maritima*) populations: do marine currents shape genetic structure? *Mol. Ecol.* 16, 1847–1864
28 Kitamoto, N. *et al.* (2005) Spatial genetic structure among and within populations of *Primula sieboldii* growing on separate streams. *Mol. Ecol.* 14, 149–157
29 Hirao, A.S. and Kudo, G. (2004) Landscape genetics of alpine-snowbed plants: comparisons along geographic and snowmelt gradients. *Heredity* 93, 290–298
30 Hirao, A.S. and 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
31 Legendre, P. *et al.* (1994) Modeling brain evolution from behaviour: a permutational regression approach. *Evolution* 48, 1487–1499
32 Cade, B.S. and Richards, J.D. (2005) *User Manual BLOSSOM Statistical Software*, US Geological Survey
33 Johnson, J.B. *et al.* (2004) Model selection in ecology and evolution. *Trends Ecol. Evol.* 19, 101–108
34 Parisod, C. and Christin, P-A. (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytol.* 178, 436–447
35 Helm, A. *et al.* (2009) Human influence lowers plant genetic diversity in communities with extinction debt. *J. Ecol.* 97, 1329–1336
36 Manel, S. *et al.* (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol. Evol.* 20, 136–142
37 Pritchard, J.K. *et al.* (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959
38 Corander, J. *et al.* (2008) Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9, 539
39 François, O. *et al.* (2006) Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* 174, 805–816
40 Guillot, G. *et al.* (2005) GENELAND: a computer package for landscape genetics. *Mol. Ecol. Notes* 5, 712–715
41 Dupanloup, I. *et al.* (2002) A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* 11, 2571–2581
42 Gaudeul, M. and Till-Bottraud, I. (2008) Genetic structure of the endangered perennial plant *Eryngium alpinum* (Apiaceae) in an alpine valley. *Biol. J. Linn. Soc.* 93, 667–677

43 Born, C. *et al.* (2008) Small-scale spatial genetic structure in the Central African rainforest tree species *Aucomea klaineana*: a stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. *Mol. Ecol.* 17, 2041–2050
44 Kang, M. *et al.* (2008) Understanding population structure and historical demography in a conservation context: population genetics of an endangered fern. *Divers. Distrib.* 14, 799–807
45 Manni, F. *et al.* (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Hum. Biol.* 76, 173–190
46 Miller, M.P. (2005) ALLELES IN SPACE (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *J. Hered.* 96, 722–724
47 Manel, S. *et al.* (2007) A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Mol. Ecol.* 16, 2031–2043
48 Soares, T.N. *et al.* (2008) Landscape conservation genetics of *Dipetryx alata* ("baru" tree: Fabaceae) from Cerrado region of Central Brazil. *Genetica* 132, 9–19
49 Torres, E. *et al.* (2003) Analysis of within-population spatial genetic structure in *Antirrhinum microphyllum* (Scrophulariaceae). *Am. J. Bot.* 90, 1688–1695
50 Fortin, M-J. and Dale, M.R.T. (2005) *Spatial Analysis. A Guide for Ecologists*, Cambridge University Press
51 Fortin, M-J. *et al.* (2005) Species' geographic ranges and distribution limits: pattern analysis and statistical issues. *Oikos* 108, 7–17
52 St. Louis, V. *et al.* (2004) Spatial association between forest heterogeneity and breeding territory boundaries of two forest songbirds. *Landsc. Ecol.* 19, 591–601
53 Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*, Chapman and Hall
54 Coulon, A. *et al.* (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Mol. Ecol.* 15, 1669–1679
55 Kuehn, R. *et al.* (2007) Genetic effect of transportation infrastructure on roe deer populations (*Capreolus capreolus*). *J. Hered.* 98, 13–22
56 Balkenhol, N. *et al.* (2009) Identifying future research needs in landscape genetics: where to from here? *Landsc. Ecol.* 24, 455–463
57 Anderson, C.D. *et al.* (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol. Ecol.* 19, 3565–3575
58 Jones, A.G. *et al.* (2010) A practical guide to methods of parentage analysis. *Mol. Ecol. Res.* 10, 6–30
59 Sork, V.L. and Smouse, P.E. (2006) Genetic analysis of landscape connectivity in tree populations. *Landsc. Ecol.* 21, 821–836
60 Marshall, T.C. *et al.* (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7, 639–655
61 Walling, C.A. *et al.* (2010) Comparing parentage inference software: reanalysis of a red deer pedigree. *Mol. Ecol.* 19, 1914–1928
62 Godoy, J.A. and Jordano, P. (2001) Seed dispersal by animals: exact identification of source trees with endocarp DNA microsatellites. *Mol. Ecol.* 10, 2275–2283
63 Kamm, U. *et al.* (2009) Frequent long-distance gene flow in a rare temperate forest tree (*Sorbus domestica*) on the landscape scale. *Heredity* 103, 476–482
64 White, G.M. *et al.* (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: an example from *Swietenia humilis* Zuccarini. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2038–2042
65 Bacles, C.F.E. and Ennos, R.A. (2008) Paternity analysis of pollen-mediated gene flow for *Fraxinus excelsior* L. in a chronically fragmented landscape. *Heredity* 101, 368–380
66 Hoebee, S.E. *et al.* (2007) Mating patterns and contemporary gene flow by pollen in a large continuous and a small isolated population of the scattered forest tree *Sorbus torminalis*. *Heredity* 99, 47–55
67 Kamm, U. *et al.* (2010) Open areas in a landscape enhance pollen-mediated gene flow of a tree species: evidence from northern Switzerland. *Landsc. Ecol.* 25, 903–911
68 Piry, S. *et al.* (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J. Hered.* 95, 536–539
69 Wilson, G.A. and Rannala, B. (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163, 1177–1191
70 He, T. *et al.* (2009) Long-distance dispersal of seeds in the fire-tolerant shrub *Banksia attenuata*. *Ecography* 32, 571–580
71 Raffl, C. *et al.* (2008) Patterns in genetic diversity of *Trifolium pallescens* populations do not reflect chronosequence on alpine glacier forelands. *Heredity* 100, 526–532
72 Mitsiu, Y. *et al.* (2010) Multiple spatial scale patterns of genetic diversity in riparian populations of *Ainslianaea faurieana* (Asteraceae) on Yakushima Island, Japan. *Am. J. Bot.* 97, 101–110
73 Faubet, P. and Gaggiotti, O.E. (2008) A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics* 178, 1491–1504
74 Honnay, O. *et al.* (2006) Low impact of present and historical landscape configuration on the genetics of fragmented *Anthyllis vulneraria* populations. *Biol. Conserv.* 127, 411–419
75 Schmidt, T. *et al.* (2008) Effects of landscape structure on genetic diversity of *Geum urbanum* L. populations in agricultural landscapes. *Flora* 204, 549–559
76 Vandepitte, K. *et al.* (2007) Landscape genetics of the self-compatible forest herb *Geum urbanum*: effects of habitat age, fragmentation and local environment. *Mol. Ecol.* 16, 4171–4179
77 Merilä, J. and Crnokrak, P. (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* 14, 892–903
78 Reed, D.H. and Frankham, R. (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55, 1095–1103
79 Leimu, R. *et al.* (2006) How general are positive relationships between plant population size, fitness and genetic variation? *J. Ecol.* 94, 942–952
80 Hoffmann, A.A. and Willi, Y. (2008) Detecting genetic responses to environmental change. *Nat. Rev. Genet.* 9, 421–432
81 Storz, J.F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol. Ecol.* 14, 671–688
82 Stinchcombe, J.R. and Hoekstra, H.E. (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100, 158–170
83 Beaumont, M.A. and Balding, D.J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* 13, 969–980
84 Savolainen, V. *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* 441, 210–213
85 Bonin, A. *et al.* (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conserv. Biol.* 21, 697–708
86 Meudt, H.M. and Clarke, A.C. (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends Plant Sci.* 12, 106–117
87 Joost, S. *et al.* (2008) Spatial analysis method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Mol. Ecol. Res.* 8, 957–960
88 Todesco, M. *et al.* (2010) Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465, 632–636
89 Excoffier, L. *et al.* (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103, 285–298
90 Guillot, G. *et al.* (2009) Statistical methods in spatial genetics. *Mol. Ecol.* 18, 4734–4756
91 Fortuna, M.A. *et al.* (2009) Networks of spatial genetic variation across species. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19044–19049
92 Poncet, B.N. *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Mol. Ecol.* 19, 2896–2907
93 Alvarez, N. *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in alpine plants. *Ecol. Lett.* 12, 632–640
94 Nordborg, M. and Weigel, D. (2008) Next-generation genetics in plants. *Nature* 456, 720–723
95 Besenbacher, S. *et al.* (2009) A fast algorithm for genome-wide haplotype pattern mining. *BMC Bioinformatics* 10, S74
96 Schmidt, P.S. *et al.* (2008) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology* 89, S91–S107

97 Pariset, L. *et al.* (2009) Landscape genomics and biased F_{ST} approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics* 10, 7

98 Parisod, C. and Joost, S. (2010) Divergent selection in trailing- *versus* leading-edge populations of *Biscutella laevigata*. *Ann. Bot.* 105, 655–660

99 Jump, A.S. *et al.* (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Mol. Ecol.* 15, 3469–3480

100 Bradbury, P.J. *et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635

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

102 R Development Core Team (2007) *R: a Language and Environment for Statistical Computing* (http://www.R-project.org)