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.

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 km²), for example, when studying the adaptive response of a dominant forest tree to...
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 m²), 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]).
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 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 using permutation for significance testing. Note that in all of the above examples, forests are hindering gene flow.

**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 cover. The overlay approach 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. 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.
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) 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. 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. First, logistic regression relates allele occurrences with environmental variables. The spatial analysis method (SAM) offers a user-friendly framework to perform logistic regression in a landscape genetic approach. SAM has successfully been used in several animal studies, but we are aware of only a single plant study that has applied SAM. Parisod and Joost 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 beechnut (Fagus sylvatica). The logistic regression approach can easily be extended to more sophisticated generalized linear models. Recently, we successfully applied multiple linear regression to identify AFLP fragments correlated with temperature and precipitation in the alpine rock cress (Arabis alpina). 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. (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 R² or the Akaike information criterion (AIC) values. The use of adjusted R² and AIC is currently explored in landscape genetic studies of plants.

between heterogeneity of forest structures and boundaries of bird territories. 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, 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. 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 (MastersBayes [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 GENECCLASS [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 GENECCLASS [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 (FST) 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
trend 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 (PCNM) to landscape genetic analysis using R [102]. PCNM values on large- and small-scale distances 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 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 (CRS133 121755/1) and the European ECOCHANGE projects of the ETH domain, the AVE project of the Swiss National Foundation (DEB-0853768), the University of California at Santa Barbara, and the State of California. We also thank 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
38 Corander, J. et al. (2008) Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9, 539
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