

王天瑞,冯力,杜芳.生态适应研究新方法:从种群遗传学到景观基因组学[J].中国科学:生命科学,2021,51(02):167-178.

Wang T R, Feng L, Du F. K. 2021. New approaches for ecological adaptation study: from population genetics to landscape genomics (in Chinese). *Sci Sin Vitae*, 2021, 51: 167–178, doi: 10.1360/SSV-2020-0265

[Translated by Prof. Myong Gi Chung on 22 March 2023, E-mail: mgchung@gnu.ac.kr]

[Edited by the original authors between 28 March and 10 April 2023]

[MGC and FKD finally edited the English version on 27 April 2023]

Abstract

Ecological adaptation or local adaptation is one of the most important responses of species to the changing environment. Understanding the molecular mechanism of local adaptation provides a theoretical foundation for protecting and utilizing the species. Landscape genomics studies the interaction between the genetic variation of species and landscape characteristics on the genomic level which quantifies the response of genetic variation caused by environmental changes at the genomic level, and can provide new insights for local adaptation research as compared with the classic population genetics methods. In this review, we first summarized a series of methods for local adaptation research, with particular emphasis on the introduction of generalized linear mixed models, multivariate statistical analysis, non-linear models, and calculation tools involved in genotype-environment associations (GEAs), and we also compared the advantages and disadvantages of these methods. Next, we introduced the newly developed method “risk of non-adaptedness (RONA)”, which evaluates the current and predicts future adaptability of species by combining the allele frequency and the changes in environmental factors. Finally, using case studies of tree-local adaptation, we provided suggestions for future study directions on tree-local adaptation.

Keywords: Genetic variation, climate change, genotype-environment interaction, ecological adaptation

摘要 [Abstract in Korean]

생태적 적응 또는 지역 적응은 변화하는 환경에 대한 종의 가장 중요한 반응 중 하나이며, 지역 적응의 분자 메커니즘을 이해하면 종의 보호 및 이용을 위한 이론적 토대를 제공한다. 경관 유전체학은 종의 유전적 변이와 경관 특성 간의 상호작용을 유전체 수준에서 연구하여 환경 변화에 따른 유전적 변이의 반응을 유전체 수준에서 정량화하고, 고전적인 개체군[집단] 유전학 방법과 비교하여 지역 적응 연구에 대한 새로운 통찰력을 제공할 수 있다. 이 리뷰에서 저자들은 먼저 일반 선형 혼합 모델, 다변량 통계 분석, 비선형 모델 및 유전자형-환경 연관성(genotype-environment associations, GEAs)에 관련된 계산 도구의 도입에 중점을 두고 지역 적응 연구를 위한 일련의 방법을 요약했고 이러한 방법의 장점과 단점을 비교하였다. 다음으로 대립유전자 빈도와 환경적 요인의 변화를 조합하여 종의 현재를 평가하고 미래의 적응성을 예측하는 신개발 방법인 “비적응 위험도(the risk of non-adaptedness, RONA)”를 소개한다. 마지막으로 수목의 국지적응 사례연구를 통해 수목국지적응에 대한 향후 연구방향에 대한 제언을 하였다.

핵심어: 유전적 변이, 기후 변화, 유전자형-환경 상호작용, 생태적 적응

Introduction

Climate change is a major factor leading to global biodiversity loss [1,2]. In general, species lacking phenotypic plasticity can adapt to changing environments by dispersing to other suitable habitats [3,4], and can also be retained in the original habitats through adaptive evolution caused by gene flow or standing genetic variation [5,6]. However, for most plants, when the rate of climate change is faster than the rate of plant adaptation, it is difficult for plants to adapt to the rapidly changing climate through migration or dispersal [5,7-10]. Instead, plants will respond to climate change primarily through their standing genetic variation [6,11]. From a researcher's perspective, there may be some shortcomings when using a single method to disentangle different biotic and/or abiotic factors related to ecological adaptation. Thus, it is necessary to use several state-of-the-art methods to reveal and predict the ability of plants to adapt to novel environments, and thus to help people better understand how plants adapt to a rapidly changing climate [12,13].

The classic method for detecting the genetic basis of ecological adaptation relies on population genetics [14]. This approach attempts to compare the genetic differentiation between populations (F_{ST} , the average inbreeding coefficient between populations within a species) and identify the genetic differentiation-specific loci that are obviously different from neutral evolution. These outlier loci are most likely to be affected by natural selection [15,16]. However, this method heavily depends on a large number of samples and does not consider

environmental heterogeneity, and would produce false positives [17,18]. The newly developed landscape genomics method focuses on genotype-environment associations (GEAs) to reveal the molecular mechanism of ecological adaptation [2,19,20]. Landscape genomics can uncover the interaction between species' adaptive genetic variation and landscape characteristics at the genome level by integrating genetic variation and biological spatial models [21,22]. The applications of generalized linear mixed models [23,24], multivariate statistical analysis [25,26], and nonlinear models [27,28] involved in the field of landscape genomics can deeply identify the loci or functional genes associated with environmental factors with higher accuracy than population genetics studies. However, population genetics potentially provides multiple benefits for researchers to understand the genetic distribution pattern and demographic history of a given species. Therefore, combining the principles and methods of population genetics and landscape genomics can deepen our understanding of the molecular mechanisms of species' ecological adaptation and infer the species' adaptive ability under climate change scenarios [20,29]. In this review, we first summarize the identification methods of ecological adaptation based on the outlier tests, generalized linear mixed models, multivariate statistical analysis, and nonlinear models, and then construct the general technical process (Figure 1) by comparing their advantages and disadvantages (Table 1). Secondly, we introduce the newly developed method of "risk of non-adaptedness (RONA)". This method combines allele frequency and environmental factors to evaluate the species' response to the current environment and further predict its adaptability to the future environment, providing insights into estimating the adaptive potential of species to future environmental gradients. Finally, we depict some recent research cases of tree ecological adaptation and then provide some suggestions and/or proposals for future studies on tree ecological adaptation.

1 Method for the detection of ecological adaptation

1.1 Detection of specific loci based on genetic differentiation

The classical approach to uncovering molecular imprints consistent with adaptive evolution is based on F_{ST} -specific loci analysis, aiming to find genetic differentiation-specific loci beyond the expectations of neutral selection [15,16,30]. A large number of methods for detecting ecological adaptation based on F_{ST} -specific loci have been developed [31,32]. This review introduces the two most widely used outlier tests for detecting adaptive signatures [31,33] (Figure 1, Table 1).

One class of methods utilizes the differences in allele frequencies between populations to identify genetic loci subject to natural selection from genetic data. The theoretical basis of this method is that loci subject to directional selection (adaptive selection or positive selection) show higher genetic differentiation than neutral loci, while loci subjected to balancing selection (purifying selection or negative selection) show lower genetic differentiation than neutral loci [31]. This method considers the effective population size and migration rate and can effectively reduce the false positive rate. At first, the F_{ST} at genetic loci is calculated, then prior odds are set for the neutral model and the distribution of the posterior odds (PO) is calculated. The loci that differed significantly from the F_{ST} expected by the neutral theory are identified as the F_{ST} -outliers. However, special attention should be paid to the selection of PO value when applying this method. For data with many genetic loci, setting a lower PO value will lead to false positives, while setting a higher PO value will lead to false negatives. Meanwhile, it should be noted that if the genetic differentiation loci are located on the candidate genes, the number of outliers may be greater than the number of randomly selected loci [34,35]. BayeScan is one of the most representative programs applying the aforementioned principles. It detects specific loci by directly calculating the PO value of each locus [31]. When conducting BayeScan analysis, the users need to convert genetic data into GESTE/BayeScan format and then set the input, output, and iteration numbers. Before the final analysis, a trial run is required to estimate the mean and variance of the parameters during model simulation. Usually, the number of trial runs (pilot runs) is set to 20, and the length of the trial run (pilot runs) run is 5,000. Finally, the genetic loci whose PO values are greater than the set threshold are defined as F_{ST} -outliers.

Another type of method fits the population dynamics of species to a null hypothesis distribution model, and then compares the genetic differentiation under the model with the actual genetic differentiation, and tests its significance. This method considers the influence of heterozygosity on genetic differentiation and sets different thresholds for different heterozygosities to identify specific genetic differentiation loci [33]. The algorithm can be implemented by Fdist2 software. First, an island model is assumed to simulate the historical dynamics of the population and subsequently compared them with real genetic differentiation loci. Then, the real heterozygosity and F_{ST} of each locus were deduced based on the simulated heterozygosity and estimated genetic differentiation. The results may be very sensitive to the historical dynamics of specific populations, so the false positives may be higher than BayeScan [15].

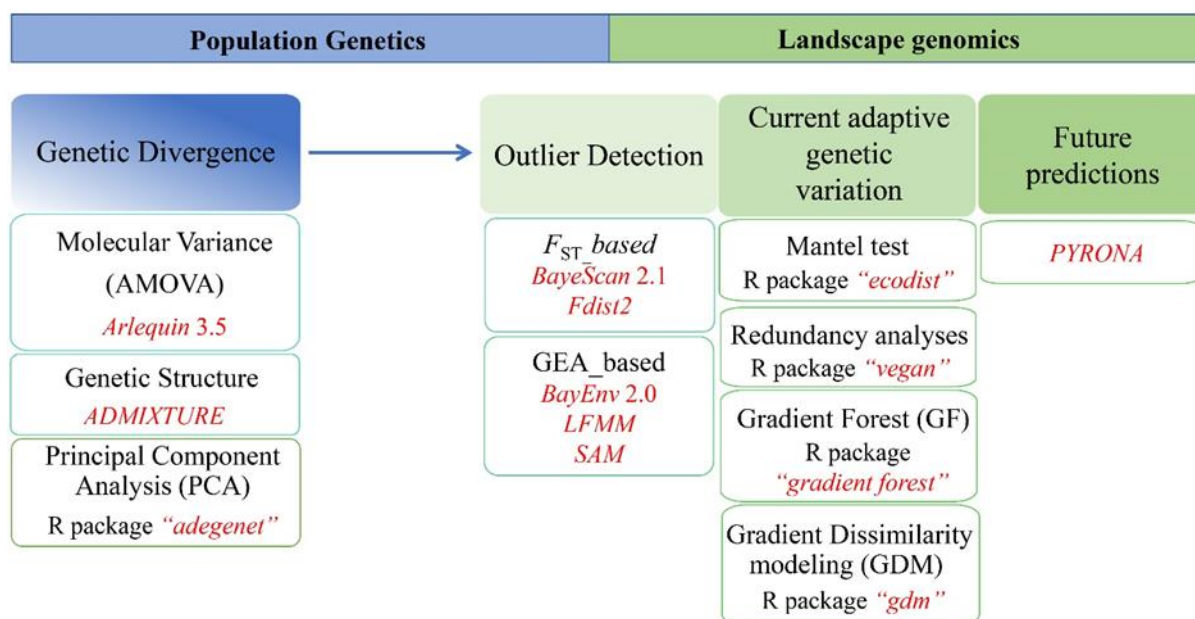


Figure 1 Overview of strategies used for local adaptation analysis. The software and R package used for analysis are shown in italics in red. The detailed information is listed in Table 1.

Table 1 Overview of methods and software available for genetic differentiation and environmental association analysis.

	Method	Association type	Individual /population	Software or R package	Ref.
Genetic differentiation-specific locus detection	Bayesian	Markov Chain	Individual	Fdist2	[31]
		Monte Carlo	Individual	BayeScan 2.1	[33]
	Bayesian mixture model	Linear	Individual and population	Bayenv 2.0	[23]
Generalized linear mixed model	Mixed linear model	Linear	Individual and population	“LEA” R package	[24]
	Multiple logistic regression	Linear	Individual	SAM Matlab®	[37, 38]
Multivariate statistical analysis	Linear regression	Linear	Individual	“geosphere, ecodist” R package	[41, 47]
	Multiple regression lines	Linear	Individual	“vegan” R package	[48]

Nonlinear model	Nonparametric				
	machine learning	Non-linear	Individual	“ <i>gradient forest, extended forest</i> ” R package	[27]
	regression line				
	Permutation matrix regression	Non-linear	Population	“ <i>gdm</i> ” R package	[28]

1.2 Generalized linear mixed model

Outlier tests based on identification of regions of high genetic differentiation among populations as compared to a neutral model usually require large samples from different populations [17], while they do not consider environmental heterogeneity. The detection of GEAs can rely on generalized linear mixed models, which can reveal the adaptive genetic pattern of species by identifying genetic variations associated with environmental gradients [36]. These methods consider the environmental heterogeneity between populations and have advantages over the outlier tests mentioned above. Researchers have developed a series of methods to identify putatively adaptive loci by association analysis between allele frequency and environmental factors. In this review, the most widely used Bayesian mixed model [23], latent factor mixed model [24], and spatial analysis algorithm [37, 38] are taken as examples (Figure 1, Table 1).

The Bayesian mixture model is an empirical model for estimating the covariance of allele frequencies between populations and then uses the model as a null hypothesis for testing genetic loci. The biggest feature of this method is that it establishes a null hypothesis model as a reference via neutral genetic markers, excluding the interference of historical population dynamics [23]. This analysis can be applied by the BayEnv software. Firstly, the covariance matrix of allelic frequencies and environmental variables of populations are established based on neutral genetic loci as the prior distribution model of a null hypothesis. Based on this model, the users then can calculate the posterior distribution P value of the correlation between the allele frequency of each locus and each environmental variable. The correlation strength is identified by the Bayes factor [23]. This method is very effective for detecting loci with congruent responses to the environments among populations, but when analyzing whether loci are affected by differential local adaptation, a separate regional and/or time slice analysis is required [39]. However, this method has some drawbacks, for instance, it assesses correlations

between a large number of single nucleotide polymorphisms (SNPs) and climate variables at a time without taking complex spatial structure into account and might lead to high false positives [40].

The mixed linear model method is based on the Bayesian Markov chain Monte Carlo (MCMC) algorithm [24], which can analyze the association between predictor variables and allele frequencies. This method takes latent factors (such as population structure) into account, it can avoid the interference of model confounding effects caused by demography and geographical isolation (isolation by distance, IBD) [41], and has low false positives [24]. The mixed linear model first uses Bayesian estimated regression coefficients to generate data showing associations between population structure and environmental variables to correct for confounding effects. Then the specific loci associated with environmental variables can be identified as the signal of natural selection [24]. The latent factor mixed models (LFMM) analysis can execute the above process. The LFMM analysis first uses the R package *LEA* [42,43] to convert the genetic data to LFMM format, and the missing data is inputted by using the sparse nonnegative matrix factorization (SNMF) function. Then, the users can set run length parameters (i.e., iterations, burn-in). The simulated results are sensitive to the run length parameters when the data set is relatively small (hundreds of individuals, thousands of loci). In this case, the run length parameters can be increased appropriately [43]. Finally, researchers can set the latent factor (i.e., population structure) and run ten independent calculations by simulating the correlation between the allele frequencies and environmental variables. The correlation between the allele frequencies and environmental variables is evaluated according to the significance level (P) of the hypothesis test. A key benefit of this method is that it can simultaneously evaluate the impacts of neutral population structure and the environmental effects on allele frequencies without having to define a structure or determine putatively neutral loci in advance [39].

Spatial analysis method (SAM) is a method based on spatial analysis combined with geographic information systems (GIS), environmental variables, and molecular data to detect the characteristics of natural selection [37,38]. This method can not only identify loci underlying genetic differentiation but also consider ecological factors. It first constructs a matrix based on genetic loci and corresponding environmental parameters at sampling sites. Then multiple univariate logistic regression is used to detect the association between the allele frequencies and environmental variables. This method can be run by the SAM program developed by Matlab® based on the generalized linear model fitting (GLMfit) [44] function.

The likelihood equation was deduced by entering the number of models to be calculated (i.e., the number of genetic loci and environmental variables). Then, the maximum likelihood estimates for the parameters were determined and the results could be displayed by setting a dynamic table in Excel [37]. However, as this method ignores the population structure, if the studied species has a complex demographic history, it may produce high false positives [24,45]. The recently developed modular SAM β ADA well overcomes the above drawbacks that exist in the initial SAM program and can implement the multivariate analysis. The neutral population structure can be defined as an additional factor when evaluating the correlation between allele frequencies and environmental variables in the SAM β ADA [46]. In addition, the SAM β ADA can quantify the level of spatial autocorrelation of genotypes and can be run on different processors in parallel.

1.3 Multivariate statistical analysis

Multivariate statistical analysis usually combines the environmental gradient and the spatial genetic structure to test the multivariate relationship between them. The most representative methods include the Mantel test (i.e., geographical/environmental isolation analysis) [41,47] and redundancy analyses (RDA) [48] (Figure 1, Table 1).

IBD [41] and isolation by environment (IBE) [47] are the most commonly used methods to evaluate the relationship between geographic/environmental distance and population genetic differentiation (e.g., F_{ST}). The correlation between the matrices of genetic differentiation and the geographical/environmental distance is achieved by Pearson correlation coefficients [49,50]. However, this method should be restricted to the cases in which data can only be expressed as pairwise distances in the taxa. Compared with other alternative methods (i.e., partial Mantel test, multiple matrix regression with randomization (MMRR), and partial RDA), its statistical power is not high, and there are certain false positives [51].

In the IBD/IBE test, to regress the F_{ST} and the pairwise linear geographical/environmental distance with the shortest distance, $F_{ST}/(1-F_{ST})$ can be calculated by linear transformation [52,53]. According to the latitude and longitude of the sampling sites, the R packages *geosphere* [54] and *ecodist* [55] are used to calculate the matrices of geographic distance and environmental distance between sampling sites, respectively. Finally, the IBD/IBE test was performed to detect the correlation between geographical distance, F_{ST} between populations, and climate differences, respectively [56-58]. It should be noted that when analyzing geographical/environmental isolation, environmental/geographical distance

should be controlled and the partial Mantel test [59] and/or multiple regression on distance matrices (MRM) [60] are also needed to execute simultaneously.

The IBD/IBE test (e.g., the Mantel test) can only detect the impacts of a generalized geographical/environmental matrix on genetic differentiation (e.g., F_{ST}) rather than identify the association between F_{ST} and the distance matrices, from specific environmental conditions such as precipitation, temperature, and soil. Redundancy analysis (RDA) overcomes the above shortcomings. RDA is a method of constrained permutation [48] and is a multivariate constrained ordination method that integrates principal component analysis and multiple regression analysis [48]. RDA can profile the distribution of genetic variation on a specific environmental gradient by integrating genetic data with environmental factors for multivariate modeling [22,25,61] with low false positives [53]. RDA can be calculated using the 'rda' function [62] in the R package *vegan*. The allele frequency of each individual at each locus treated as a response variable in RDA can be transformed by the R package *LEA* [43], and the environmental data is treated as explanatory variables. To avoid the influence of geographical (latitude and longitude) variables when analyzing climate variables, partial RDA (*pRDA*) is required. Finally, the climate and genetic data are performed by permutation tests to evaluate the significance levels. Since natural selection usually involves a polygenic process driven by multiple environmental variables, therefore, compared with traditional univariate analysis, multivariate analysis has greater advantages in detecting the ecological adaptation process of species [63]. Similar multivariate statistical analysis methods include distance-based redundancy analysis (dbRDA) [64] and canonical correspondence analysis (CCA) [65].

1.4 Nonlinear model

The influence of environmental factors on genetic differentiation may not conform to the linear model in most cases. Therefore, using linear model-based methods to confirm the association between genetic variation and the environmental gradient might arise biased interpretations. In ecological analysis, researchers have mainly focused on species' responses to environmental gradients and identified the most important environmental variables that can be used to predict biodiversity patterns. The most representative methods based on nonlinear regression include gradient forests (GF) [27] and generalized dissimilarity modeling (GDM) [28] (Figure 1, Table 1).

GF is a community-level extension of the nonparametric, machine-learning regression tree method known as random forests (RF). RF can improve prediction accuracy and identify

gradient abundance changes [66]. GF measures the role of predicted environmental variables by integrating univariate RF analysis and cross-validation of important values, which can directly simulate the compositional changes of genetic variation and explore the nonlinear associations of space, environment, and allelic frequencies. GF also provides a means to examine the response of individual locus to environmental gradient, although it will become computationally limited when using a large SNP-based dataset [67]. The R packages *gradientForest* and *extendedForest* [68,69] can be used to integrate the results of RF of multiple species to determine potential predictor variables and obtain the proportion of genetic variation explained by each predictor. The influence of different environmental variables on genetic variation can be evaluated by calculating the proportion of the corresponding predicted variables changing along the environmental gradient [27,70].

GDM is based on the dissimilarity between sampling sites to simulate the spatial variation of genetic differentiation and then explains the correlations among genetic variation, environment, and geographical distance using a nonlinear model. GDM is a nonlinear extension of permutational matrix regression that models pairwise biological dissimilarity between sites as a nonlinear function of pairwise site differences in environmental and geographic variables. It also utilizes I-spline and percent deviance to determine the compositional turnover of allele frequencies along environmental gradients and explains the degree of model fitting, respectively [28, 70]. The distance-based GDM can explain the spatial patterns of genetic variation caused by demographic processes (i.e., IBD or isolation by ecological resistance) and explore infinite genetic loci with arbitrary allelic frequencies [28]. GDM is suitable for the fitting analysis of loci with pairwise distance matrices (i.e., geographical and environmental distance matrices), which can be implemented by the R package *gdm* [71]. The advantages of the GF and GDM are to handle large-scale genomic data and provide insights into genomic regions ostensibly under local adaptation. In addition, compared with the other methods listed above, an especially powerful feature of GDM and GF is that they can be employed to evaluate the potential impacts of climate change on biodiversity at the genetic level and how these impacts vary spatially [19].

2 Analysis of the risk of non-adaptedness (RONA)

In the context of climate change, it is crucial to determine the “fitness” (i.e., adaptive capacity) of species under future climate scenarios. Traditional ecological niche models (ENMs) integrate species’ occurrence data and bioclimatic variables to predict their potential

distributions under future climate changes. ENMs can delineate areas or populations that necessitate priority protection based on the prediction results, informing the protection of species and decision-making on conservation interventions [72,73]. However, ENMs assume the species ‘as a whole’ and ignore the local adaptation of species to specific biotic or abiotic factors, which often varies across populations (i.e., intraspecific variation) [74-76].

In order to overcome the deficiencies of ENMs, the risk of non-adaptedness (RONA), which relies on simple linear regression, was developed and represents the average change in allele frequency at climate-associated loci theoretically required to match future climate conditions [77] (Figure 1). This method utilizes the RONA values to assess the possibility of the population persistence under future climate change. The higher the RONA value means the lower the potential of the population to adapt to future climate conditions and *vice versa*. Empirical studies have shown that if the expected allele frequency < 0.1 per decade might be able to match the projected climate change, whilst change greater than 0.1-0.2 per decade might result in high climate adaptation lags [78]. Compared with the traditional ENMs, the advantage of the RONA method is that it takes the species distributions and allelic frequency of species into account simultaneously.

When performing RONA analysis, researchers can first download the climate data under different representative concentration pathways (RCP). The allele frequency of individuals and the assumed value (*P*-value) matrix of environmental factors can be retrieved from LFMM. RONA can be executed by PYRONA which ranks environmental factors by the number of associations between genetic loci (e.g., SNPs) and environmental factors [29]. Furthermore, the average RONA value calculated by PYRONA is weighted by the R^2 value of each involved correlation, which can be interpreted as the adaptive potential of the population under future climate change. The recent study precisely defined the method of predicting the adaptive potential of species under the future environmental gradient as a future risk of non-adaptedness (f-RONA) [79]. Furthermore, this approach argued that the current risks were likely to be particularly important for species that are already declining due to climate change. Based on this principle, the researchers explicitly defined the current risk of non-adaptedness (c-RONA) by further extending the initial RONA [79], which can evaluate the average change in allele frequency at climate-associated loci required for a given species to match the estimate of optimum under the current environmental gradients.

3 Application of ecological adaptation research in forest ecology

With long generation times and large effective population sizes, trees often occupy highly heterogeneous environments, which are good model systems for ecological adaptation researches [40]. Early studies on tree ecological adaptation mainly focused on the method of population genetics [14,16]. Since 2010, landscape genomics has been widely used in the study of tree ecological adaptation, which provides a new avenue for revealing the mechanism of tree ecological adaptation [13,20]. For example, Rellstab et al. (2016) studied the local adaptation of three species of oak trees (*Quercus petraea*, *Q. pubescens*, and *Q. robur*) in Switzerland using pooled amplicon sequencing [77]. Environmental association analysis (EAA, also called GEAs) found that there were gene loci significantly associated with precipitation and clay content in the soil in the three species. Martins et al. (2018) studied the ecological adaptation of *Q. rugosa* at high altitudes in Mexico using genotyping by sequencing (GBS) [80]. They identified 5,354 SNPs from 103 individuals in 17 populations. EAA showed that 97 SNPs were significantly correlated with climate variables, especially seasonal precipitation, which provided preliminary details for future conservation and management strategies of this species.

Species in nature can usually respond to multiple external environmental variables, so multivariate statistical analysis is widely used in the study of species' genetic variation in response to the environments. Sork et al. (2016) identified 195 SNPs in 13 populations of *Quercus lobata* in California from 40 candidate genes related to bud/flowering, growth, and osmotic and temperature stress [26]. The association analysis between allele frequency and climate gradient further identified five SNPs involving genes related to budburst/flowering and temperature stress. RDA showed that climate variables explained 67% of the adaptive genetic variation, but only 33% of the neutral genetic variation. In addition, more studies have shown that there may not only be a simple linear correlation between the genetic variation of species and the environmental gradients but also a nonlinear correlation. Therefore, nonlinear models play important roles in analyzing the genetic variation of species. For example, Gugger et al. (2018) applied a nonlinear, multivariate environmental association method to examine the spatial genetic structure and its association with environmental variation in an ecologically and economically important tree species endemic to Hawaii, *Acacia koa* [70]. They identified more than 11,000 SNPs from 311 trees, GF and GDM both revealed strong correlations between spatial genetic structure and mean annual rainfall. Utilizing models for projected future climate on Hawaii Islands, they showed that the predicted changes in rainfall patterns may result in genetic offsets.

Researchers in China mostly use neutral markers (loci) to study the demographic history of trees, but few studies focus on the adaptive genetic variation of trees. In recent years, researchers have gradually begun to pay attention to the influences of environmental factors on the adaptive genetic variation of species. Xia et al. (2018) determined the influences of environment and geography on the population structure of Chinese pine (*Pinus tabulaeformis*) using GBS and mitochondrial markers [81]. The outlier detection identified by SNPs was significantly associated with temperature, precipitation, soil type, etc., and the RAD showed that these differentiation-specific loci had significant local adaptation signals, which provided insights into further studying the origin of adaptive variation in *P. tabulaeformis*. Gao et al. (2021) used restriction site-associated DNA sequencing (RAD-seq) to analyze the ecological adaptation of the forest tree species *Quercus acutissima* [82]. GEAs showed that the genetic differentiation-specific loci of *Q. acutissima* were significantly correlated with precipitation. In addition, IBE results showed that the gene flow among *Q. acutissima* populations is constrained by natural selection and ecological adaptation, which resulted in adaptive differentiation of some genetic loci and phenotypes. This study is of great significance for understanding the generation and maintenance mechanisms of genetic variation in forest tree species. Zhao et al. (2020) used exome capture sequencing to analyze the population structure and ecological adaptation of the key species *Pinus densata* inhabiting the Qinghai-Tibet Plateau [83]. After controlling for demographic processes, only *c.* 4% of loci were affected by selection and allele surfing, and GEAs revealed that the adaptive potential of alpine pine to cope with future climate change is limited.

The interaction between environmental heterogeneity and local adaptation is critical to understanding the evolutionary history of species. However, how precisely environmental heterogeneity drives population differentiation and how genomic variation contributes to adaptive evolution are still poorly understood for most species. Jia et al. (2020) used GBS to generate large-high-quality variants from *Platycladus orientalis* range-wide collection [84]. They clarified the pattern of the large-scale genetic variation of *P. orientalis* and established the genomic response model of this species under climate scenarios. Using GF, the authors identified the response patterns of different geographical populations *P. orientalis* and established utilization strategies of germplasm resources under future climate scenarios. Du and her colleagues [85,86] studied the tree species *Quercus aquifolioides* and carried out ecological adaptation research. By pool sequencing 65 candidate genes related to drought and oxygen stress, they identified 381 SNPs. GEAs showed that the genetic variation of the

Hengduan Mountains–Western Sichuan Plateau (HDM–WSP) and Tibet lineages are associated with the environment (precipitation) and geography (latitude and longitude), respectively. RONA showed that the HDM–WSP lineage had better adaptability to future precipitation variables. This study analyzed the effects of geographical/climatic factors on the ecological adaptation of *Q. aquifolioides* under climate change. The latest research on tree ecological adaptation showed that environmental variables play vital roles in adaptive genetic variation. Therefore, it is very important to determine the influences of environmental and landscape factors on the adaptive genetic variation of species. Landscape genomics opens the door to a better understanding of species' ecological adaptation to the environments.

4 Challenges of landscape genomics

However, there are still huge challenges in using landscape genomics methods to study ecological adaptation. First, researchers commonly use landscape genomics to reveal ecological adaptation or adaptive genetic variation of species, but they often focus on the species or subspecies level. Studies at the community level are largely ignored. In the future, landscape genomics should be used to carry out ecological adaptation of multiple species located in the same area under the framework of landscape community genomics. This method is beneficial to obtain the response of these species to the eco-evolutionary dynamics of the shared terrestrial landscape or environments, and to deepen our understanding of ecological adaptation or adaptive genetic variation [87]. In the study of ecological adaptation at the community level, the main challenge is how to simulate their local adaptation patterns and the adaptation potentials to future climate change. The landscape community genomics using GDM, GF model, and RONA might help to solve the above problems possibly.

Secondly, for the analysis of ecological adaptation using landscape genomics, there will never be a single analytical approach that is optimal for addressing all questions belonging to landscape genomics, and interdisciplinary collaboration will continue to be a cornerstone for progress in landscape genomics. The various array of approaches, with their different assumptions, advantages and limitations, also makes it challenging to synthesize results obtained from landscape genomics studies [13,39,88,89], thus a comprehensive explanation should be given in the specific analysis.

Finally, in the existing ecological adaptation studies using landscape genomics, researchers often underestimate the impacts of sampling strategies on the final results. Too often, genetic and genomic data are gathered for other research purposes, and landscape

genomic questions are only considered after sampling is finished [90]. In future studies on ecological adaptation using landscape genomics, sampling strategies must consider the eco-evolutionary dynamics of species and the representativeness of the environmental variability that species experience [13,90-92]. Recent studies have highlighted the importance of selecting an appropriate sample size in landscape genomics. To retrieve obvious adaptive signals, for species with limited dispersal ability, sample sizes above 200 units (sample size per location) are generally sufficient to detect most adaptive signals, while in random mating populations, this threshold should be increased to 400 units [93]. In addition, researchers necessitate considering the genomics resources available for the study of organisms in the future (such as whether there is a reference genome). Although there are already numerous genome-level sequencing methods that can be used in the study of ecological adaptation (such as RAD, transcriptome sequencing, etc.). The number of SNPs and whether adaptive loci with strong selection signals can be detected are strongly related to the selection of sequencing strategies [94].

In closing, when using landscape genomics methods in the future, population genetics methods should first be used to study species' genetic diversity, population structure, and gene flow within and between populations, to understand the demographic history of species. Based on the in-depth understanding of the evolutionary history of species, appropriate sampling strategies and analysis methods should be employed to evaluate the associations between genomic data and climate variables, we can explore the adaptability of species or communities to the environment and understand their eco-evolutionary dynamics. Finally, RONA can be used to predict future fate and reveal its adaptation mechanism of species, providing guidelines for the scientific protection of species and decision-making on precise management interventions. The landscape genomics research of protected species with adaptive genetic variation can be used for providing theoretical bases of corresponding conservation measures (such as adopting *in situ* conservation or *ex situ* conservation, using assisted gene flow or assisted migration [10,95]).

5 Summary

Understanding the genetic basis of ecological adaptation is the key issue of molecular ecology and evolutionary biology. By integrating the theories and methods of ecological adaptation research, this review proposes that it is necessary to use population genetics methods to understand species distribution patterns and the evolutionary history of species. Meanwhile,

combined with the methods of landscape genomics, this review recommends researchers integrate genetic variation, environmental factors, and landscape characteristics to illuminate the interactions between species' adaptive genetic variation and landscape characteristics. Projecting the adaptation potential under alternative climate scenarios via a newly developed RONA method can clarify the response mechanism of species' adaptive genetic variation under different spatial-temporal scales, which may provide new avenues for dissecting the ecological adaptation of species' ongoing rapid climate change.

Acknowledgments: We would like to thank Professor Dayong Zhang of Beijing Normal University, Xinzeng Wei, a researcher at Wuhan Botanical Garden of the Chinese Academy of Sciences, and Professor Jianfeng Mao of Beijing Forestry University for their comments on this article.

References

- [1] Bellard C, Bertelsmeier C, Leadley P, et al. Impacts of climate change on the future of biodiversity. *Ecol Lett*, 2012, 15: 365–377.
- [2] de Lafontaine G, Napier J D, Petit R J, et al. Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology*, 2018, 99: 1530–1546.
- [3] Nicotra A B, Atkin O K, Bonser S P, et al. Plant phenotypic plasticity in a changing climate. *Trends Plant Sci*, 2010, 15: 684–692.
- [4] Thompson P L, Fronhofer E A. The conflict between adaptation and dispersal for maintaining biodiversity in changing environments. *Proc Natl Acad Sci USA*, 2019, 116: 21061–21067.
- [5] Aitken S N, Yeaman S, Holliday J A, et al. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol Appl*, 2008, 1: 95–111.
- [6] Alberto F J, Aitken S N, Alía R, et al. Potential for evolutionary responses to climate change—Evidence from tree populations. *Glob Chang Biol*, 2013, 19: 1645–1661.
- [7] Corlett R T, Westcott D A. Will plant movements keep up with climate change? *Trends Ecol Evol*, 2013, 28: 482–488.
- [8] Hampe A, Petit R J. Conserving biodiversity under climate change: the rear edge matters. *Ecol Lett*, 2010, 8: 461–467.
- [9] Zhu K, Woodall C W, Ghosh S, et al. Dual impacts of climate change: forest migration and turnover through life history. *Glob Chang Biol*, 2012, 20: 251–264.
- [10] Aitken S N, Whitlock M C. Assisted gene flow to facilitate local adaptation to climate change. *Annu Rev Ecol Evol Syst*, 2013, 44: 367–388.

- [11] Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. *Nat Rev Genet*, 2013, 14: 807–820.
- [12] Isabel N, Holliday J A, Aitken S N. Forest genomics: advancing climate adaptation, forest health, productivity, and conservation. *Evol Appl*, 2020, 13: 3–10.
- [13] Rellstab C, Gugerli F, Eckert A J, et al. A practical guide to environmental association analysis in landscape genomics. *Mol Ecol*, 2015, 24: 4348–4370.
- [14] Wright S I, Gaut B S. Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol*, 2005, 22: 506–519.
- [15] Excoffier L, Hofer T, Foll M. Detecting loci under selection in a hierarchically structured population. *Heredity*, 2009, 103: 285–298.
- [16] Hohenlohe P A, Phillips P C, Cresko W A. Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *Int J Plant Sci*, 2010, 171: 1059–1071.
- [17] Eveno E, Collada C, Guevara M A, et al. Contrasting patterns of selection at *Pinus pinaster* ait. Drought stress candidate genes as revealed by genetic differentiation analyses. *Mol Biol Evol*, 2008, 25: 417–437.
- [18] Hoban S, Kelley J L, Lotterhos K E, et al. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am Nat*, 2016, 188: 379–397.
- [19] Fitzpatrick M C, Keller S R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett*, 2015, 18: 1–16.
- [20] Sork V L. Genomic studies of local adaptation in natural plant populations. *J Hered*, 2017, 109: 3–15.
- [21] Holderegger R, Kamm U, Gugerli F. Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecol*, 2006, 21: 797–807.
- [22] Sork V L, Davis F W, Westfall R, et al. Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Mol Ecol*, 2010, 19: 3806–3823.
- [23] Coop G, Witonsky D, Di Rienzo A, et al. Using environmental correlations to identify loci underlying local adaptation. *Genetics*, 2010, 185: 1411–1423.
- [24] Frichot E, Schoville S D, Bouchard G, et al. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol*, 2013, 30: 1687–1699.
- [25] Lasky J R, Des Marais D L, McKAY J K, et al. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol Ecol*, 2012, 21: 5512–5529.
- [26] Sork V L, Squire K, Gugger P F, et al. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *Am J Bot*, 2016, 103: 33–46.

- [27] Ellis N, Smith S J, Pitcher C R. Gradient forests: calculating importance gradients on physical predictors. *Ecology*, 2012, 93: 156–168.
- [28] Ferrier S, Manion G, Elith J, et al. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers Distrib*, 2007, 13: 252–264.
- [29] Pina-Martins F, Baptista J, Pappas Jr G, et al. New insights into adaptation and population structure of cork oak using genotyping by sequencing. *Glob Chang Biol*, 2018, 25: 337–350.
- [30] Beaumont M A, Balding D J. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol*, 2004, 13: 969–980.
- [31] Beaumont M A, Nichols R A. Evaluating loci for use in the genetic analysis of population structure. *P Roy Soc Lond B Bio*, 1996, 263: 1619–1626.
- [32] Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 2008, 180: 977–993.
- [33] Vitalis R, Dawson K, Boursot P. Interpretation of variation across marker loci as evidence of selection. *Genetics*, 2001, 158: 1811–1823.
- [34] Csilléry K, Lalagüe H, Vendramin G G, et al. Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. *Mol Ecol*, 2014, 23: 4696–4708.
- [35] Roschanski A M, Csilléry K, Liepelt S, et al. Evidence of divergent selection for drought and cold tolerance at landscape and local scales in *Abies alba* Mill. in the French Mediterranean Alps. *Mol Ecol*, 2016, 25: 776–794.
- [36] Forester B R, Lasky J R, Wagner H H, et al. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol Ecol*, 2018, 27: 2215–2233.
- [37] Joost S, Bonin A, Bruford M W, et al. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol Ecol*, 2007, 16: 3955–3969.
- [38] Joost S, Kalbermatten M, Bonin A. Spatial analysis method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Mol Ecol Resour*, 2008, 8: 957–960.
- [39] Fenderson L E, Kovach A I, Llamas B. Spatiotemporal landscape genetics: Investigating ecology and evolution through space and time. *Mol Ecol*, 2020, 29: 218–246.
- [40] Sork V L, Aitken S N, Dyer R J, et al. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genet Genomes*, 2013, 9: 901–911.
- [41] van Strien M J, Holderegger R, Van Heck H J. Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity*, 2015, 114: 27–37.
- [42] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for

Statistical Computing, Vienna, 2018.

- [43] Fricot E, François O. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol*, 2015, 6: 925–929.
- [44] MacCullagh P, Nelder J A. *Generalized Linear Models*. London: Chapman & Hall/CRC, 1989.
- [45] De Mita S, Thuillet A C, Gay L, et al. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol*, 2013, 22: 1383–1399.
- [46] Stucki S, Orozco-terWengel P, Forester B R, et al. High performance computation of landscape genomic models including local indicators of spatial association. *Mol Ecol Resour*, 2017, 17: 1072–1089.
- [47] Manthey J D, Moyle R G. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. *Mol Ecol*, 2015, 24: 3628–3638.
- [48] Legendre P, Legendre L F. *Numerical Ecology*. Amsterdam: Elsevier Ltd, 2012.
- [49] Diniz-Filho J A F, Soares T N, Lima J S, et al. Mantel test in population genetics. *Genet Mol Biol*, 2013, 36: 475–485.
- [50] Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*, 1967, 27: 209–220.
- [51] Harmon L J, Glor R E. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution*, 2010, 64: 2173–2178.
- [52] Rousset F. Genetic differentiation and estimation of gene flow from F -statistics under isolation by distance. *Genetics*, 1997, 145: 1219–1228.
- [53] Slatkin M. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 1993, 47: 264–279.
- [54] Hijmans R J. *Geosphere: Spherical Trigonometry*. R package geosphere version 1.5-7, 2014.
- [55] Goslee S C, Urban D L. The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Soft*, 2007, 22: 1–9.
- [56] Wang I J. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, 2013, 67: 3403–3411.
- [57] Wang I J, Glor R E, Losos J B. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol Lett*, 2013, 16: 175–182.
- [58] Wang I J, Bradburd G S. Isolation by environment. *Mol Ecol*, 2014, 23: 5649–5662
- [59] Smouse P E, Long J C, Sokal R R. Multiple regression and correlation extensions of the Mantel test

- of matrix correspondence. *Syst Zool*, 1986, 35: 627.
- [60] Lichstein J W. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol*, 2007, 188: 117–131.
- [61] Sork V L, Waits L. Contributions of landscape genetics—Approaches, insights, and future potential. *Mol Ecol*, 2010, 19: 3489–3495.
- [62] Oksanen J, Blanchet F G, Kindt R, et al. *vegan: Community Ecology Package*. R package version, 2017.
- [63] Forester B R, Jones M R, Joost S, et al. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol Ecol*, 2016, 25: 104–120.
- [64] Legendre P, Anderson M J. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr*, 1999, 69: 1–24.
- [65] ter Braak C J F. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 1986, 67: 1167–1179.
- [66] Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Stat Sci*, 2001, 16: 199–231.
- [67] van den Wollenberg A L. Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, 1977, 42: 207–219.
- [68] Liaw A, Wiener M. Classification and regression by randomForest. *R news*, 2002, 2: 18–22.
- [69] Smith S J, Ellis N, Pitcher C R. Conditional variable importance in R package extendedForest. R vignette, 2011.
- [70] Gugger P F, Liang C T, Sork V L, et al. Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evol Appl*, 2018, 11: 231–242.
- [71] Manion G, Lisk M, Ferrier S. Package 'gdm', A toolkit with functions to fit, plot, and summarize Generalized Dissimilarity Models: CRAN Repository, R. 2018.
- [72] Elith J, Leathwick J R. Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst*, 2009, 40: 677–697.
- [73] Maguire K C, Nieto-Lugilde D, Fitzpatrick M C, et al. Modeling species and community responses to past, present, and future episodes of climatic and ecological change. *Annu Rev Ecol Evol Syst*, 2015, 46: 343–368.
- [74] Ikeda D H, Max T L, Allan G J, et al. Genetically informed ecological niche models improve climate change predictions. *Glob Chang Biol*, 2017, 23: 164–176.
- [75] Peterson M L, Doak D F, Morris W F. Incorporating local adaptation into forecasts of species' distribution and abundance under climate change. *Glob Chang Biol*, 2019, 25: 775–793.
- [76] Smith A B, Beever E A, Kessler A E, et al. Alternatives to genetic affinity as a context for within-species response to climate. *Nat Clim Chang*, 2019, 9: 787–794.

- [77] Rellstab C, Zoller S, Walthert L, et al. Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Mol Ecol*, 2016, 25: 5907–5924.
- [78] Jordan R, Hoffmann A A, Dillon S K, et al. Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: Implications for adaptive potential to projected climate change. *Mol Ecol*, 2017, 26: 6002–6020.
- [79] Borrell J S, Zohren J, Nichols R A, et al. Genomic assessment of local adaptation in dwarf birch to inform assisted gene flow. *Evol Appl*, 2020, 13: 161–175.
- [80] Martins K, Gugger P F, Llanderal-Mendoza J, et al. Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evol Appl*, 2018, 11: 1842–1858.
- [81] Xia H, Wang B, Zhao W, et al. Combining mitochondrial and nuclear genome analyses to dissect the effects of colonization, environment, and geography on population structure in *Pinus tabuliformis*. *Evol Appl*, 2018, 11: 1931–1945.
- [82] Gao J, Liu Z L, Zhao W, et al. Combined genotype and phenotype analyses reveal patterns of genomic adaptation to local environments in the subtropical oak *Quercus acutissima*. *J Syst Evol*, 2021, 59: 541–556.
- [83] Zhao W, Sun Y Q, Pan J, et al. Effects of landscapes and range expansion on population structure and local adaptation. *New Phytol*, 2020, 228: 330–343.
- [84] Jia K H, Zhao W, Maier P A, et al. Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). *Evol Appl*, 2020, 13: 665–676.
- [85] Wang Y Y, Zhang Y, Fang J H, et al. A research method for ecological adaptation of tree species based on candidate genes: with *Quercus aquifolioides* as an example (in Chinese). *J Lanzhou Univ Nat Sci*, 2018, 54: 811–816.
- [86] Du F K, Wang T, Wang Y, et al. Contrasted patterns of local adaptation to climate change across the range of an evergreen oak, *Quercus aquifolioides*. *Evol Appl*, 2020, 13: 2377–2391.
- [87] Hand B K, Lowe W H, Kovach R P, et al. Landscape community genomics: understanding eco-evolutionary processes in complex environments. *Trends Ecol Evol*, 2015, 30: 161–168.
- [88] Balkenhol N, Waits L P, Dezzani R J. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, 2009, 32: 818–830.
- [89] Jaquière J, Stoeckel S, Nouhaud P, et al. Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex. *Mol Ecol*, 2012, 21: 5251–5264.
- [90] Balkenhol N, Dudaniec R Y, Krutovsky K V, et al. Landscape Genomics: understanding relationships between environmental heterogeneity and genomic characteristics of populations. In: Rajora O, ed. *Population Genomics*. Population Genomics. Heidelberg: Springer, 2017. 1–62.
- [91] Leempoel K, Duruz S, Rochat E, et al. Simple rules for an efficient use of geographic information systems in molecular ecology. *Front Ecol Evol*, 2017, 5: 33.

- [92] Manel S, Joost S, Epperson B K, et al. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Mol Ecol*, 2010, 19: 3760–3772.
- [93] Selmoni O, Vajana E, Guillaume A, et al. Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach. *Mol Ecol Resour*, 2020, 20: 154–169.
- [94] Lowry D B, Hoban S, Kelley J L, et al. Responsible RAD: striving for best practices in population genomic studies of adaptation. *Mol Ecol Resour*, 2017, 17: 366–369.
- [95] Aitken S N, Bemmels J B. Time to get moving: assisted gene flow of forest trees. *Evol Appl*, 2016, 9: 271–290.