

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/354709037>

# Redundancy Analysis (RDA): a Swiss Army knife for landscape genomics

Article in Methods in Ecology and Evolution · September 2021

DOI: 10.1111/2041-210x.13722

---

CITATION

1

READS

114

2 authors:



Thibaut Capblancq

University of Vermont

21 PUBLICATIONS 217 CITATIONS

[SEE PROFILE](#)



Brenna R Forester

U.S. Fish and Wildlife Service

40 PUBLICATIONS 1,036 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Adaptive potential of alpine tree species to climate change [View project](#)



Hybrid speciation in alpine Coenonympha butterflies [View project](#)

# Redundancy analysis: A Swiss Army Knife for landscape genomics

Thibaut Capblancq<sup>1</sup>  | Brenna R. Forester<sup>2</sup> 

<sup>1</sup>Department of Plant Biology, University of Vermont, Burlington, VT, USA

<sup>2</sup>Department of Biology, Colorado State University, Fort Collins, CO, USA

**Correspondence**

Thibaut Capblancq  
 Email: thibaut.capblancq@gmail.com  
 Brenna R. Forester  
 Email: brenna.forester@colostate.edu

**Handling Editor:** Javier Palarea-Albaladejo

## Abstract

1. Landscape genomics identifies how spatial and environmental factors structure the amount and distribution of genetic variation among populations. Landscape genomic analyses have been applied across diverse taxonomic groups and ecological settings, and are increasingly used to analyse datasets composed of large numbers of genomic markers and multiple environmental predictors.
2. It is in this context that multivariate methods show their strengths. Redundancy analysis (RDA) is a constrained ordination that, in a landscape genomics framework, models linear relationships among environment predictors and genomic variation, effectively identifying covarying allele frequencies associated with the multivariate environment. RDA can be used at both individual and population levels, can include covariates to account for confounding factors and can be used to directly infer genotype–environment associations on the landscape. The modelling of both multivariate response and explanatory variables allows RDA to accommodate the genomic and environmental complexity found in nature, producing a powerful and efficient tool for landscape genomics.
3. In this review, we outline the diverse uses of RDA in landscape genomics, including variable selection, variance partitioning, genotype–environment associations, and the calculation of adaptive indices and genomic offset. To illustrate these applications, we use a published dataset for lodgepole pine that includes genomic, phenotypic and environmental data. We provide an introduction to the statistical basis of RDA, a tutorial on its use and interpretation in landscape genomics applications, discuss limitations and provide guidelines to avoid misuse.
4. This review and associated tutorial provide a comprehensive resource to the landscape genomics community to improve understanding of RDA as a modelling framework, and encourage the appropriate use of RDA across diverse landscape genomics applications. RDA is truly a Swiss Army Knife for landscape genomics:

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society

a multipurpose, adaptable and versatile approach to identifying, evaluating and forecasting relationships between genetic and environmental variation.

#### KEY WORDS

adaptive genetic variation, genomic offset, genotype–environment associations, landscape genomics, multivariate statistics, redundancy analysis, variable selection, variance partitioning

## 1 | INTRODUCTION

Landscape genomics is a diverse and rapidly evolving field that integrates techniques from population genetics, landscape ecology, spatial statistics and community ecology. Broadly, landscape genomics investigates the processes underlying patterns of neutral and adaptive genetic variation in relation to environmental and spatial heterogeneity. Early work in neutral landscape genomics (sometimes referred to as ‘landscape genetics’) focused on identifying spatial patterns of neutral genetic variation (Manel & Holderegger, 2013; Manel et al., 2003), while later definitions emphasized explicit assessment of the effects of landscape configuration, composition, and matrix quality on gene flow and the spatial distribution of neutral variation (Storfer et al., 2007). The increased availability of dense, genomic-scale data has driven the emergence of adaptive landscape genomics (Balkenhol et al., 2019; Joost et al., 2007; Manel et al., 2010; Storfer et al., 2018), which not only allows for the identification of candidate adaptive loci associated with environmental variation (i.e. genotype–environment associations, GEA), but also facilitates downstream analyses such as the calculation of adaptive indices and genomic offsets, which are used to infer spatial adaptive gradients and their temporal shifts in response to environmental change. In the last two decades, these neutral and adaptive methods have been extended to investigate the complexities of terrestrial, freshwater and marine ecosystems. Recent applications range from the characterization of constrained movement corridors and asymmetrical gene flow across seascapes and riverscapes (Brauer et al., 2018; Xuereb et al., 2018), to the identification of adaptive genetic variation in birds, mammals and plants (Lasky et al., 2015; Razgour et al., 2019; Ruegg et al., 2018), and the study of local adaptation across current and future species ranges (Capblancq, Morin, et al., 2020; Steane et al., 2014).

Most methods used in adaptive landscape genomics are univariate, meaning that they test one genetic marker at a time, and may also test only one environmental predictor at a time (Rellstab et al., 2015). For example, one of the first GEAs used a chi-squared test to identify a signature of local adaptation to drought in ponderosa pines at a single locus with three alleles (Mitton et al., 1977). By contrast, modern landscape genomic analyses may include hundreds of thousands or even millions of genetic markers and multiple environmental predictors, requiring millions of univariate tests, which do not account for covariation among environmental variables and/or genetic markers. To address this issue, multivariate statistical methods are increasingly used to analyse all markers and predictors simultaneously, allowing researchers to identify covarying sets of genotypes and their relationship with the multivariate

environment, a major advantage in contemporary, high-dimensional landscape genomic research.

One of the most commonly used multivariate landscape genomic methods is redundancy analysis (RDA), a form of constrained ordination. Ordination methods have been used in community ecology for almost 70 years to evaluate how environmental features shape patterns of community composition (Bray & Curtis, 1957; Whittaker, 1956). These models are an intuitive fit for modern genomic datasets, where we are interested in understanding how the multivariate environment shapes patterns of genomic composition across landscapes. RDA is based on multivariate regression, and models linear combinations of the explanatory variables (e.g. environmental predictors) that explain linear combinations of the response variables (e.g. single nucleotide polymorphisms, SNPs), effectively identifying covarying loci associated with the multivariate predictors (Legendre & Legendre, 2012, Box 1). RDA is a highly flexible framework, which allows us to address four of the major questions in landscape genomics:

1. What environmental/spatial processes drive patterns of genetic variation?
2. What is the genetic basis of local adaptation to the environment?
3. How is adaptive genetic variation distributed across landscapes?
4. What are the impacts of climate and/or landscape change on the distribution of adaptive genetic variation?

In this review, we will explore the many uses of RDA in the field of landscape genomics, including variable selection, variance partitioning, GEA, and the calculation of adaptive indices and genomic offset. We also provide a tutorial on the use and interpretation of RDA for these analyses (available at <https://github.com/Capblancq/RDA-landscape-genomics>), discuss limitations and provide guidelines to avoid misuse. Our goals are to provide resources to the landscape genomics community to improve understanding of RDA as a modelling framework, and encourage the appropriate use of RDA across diverse landscape genomics applications.

## 2 | RDA-BASED LANDSCAPE GENOMICS APPLICATIONS

### 2.1 | Empirical dataset

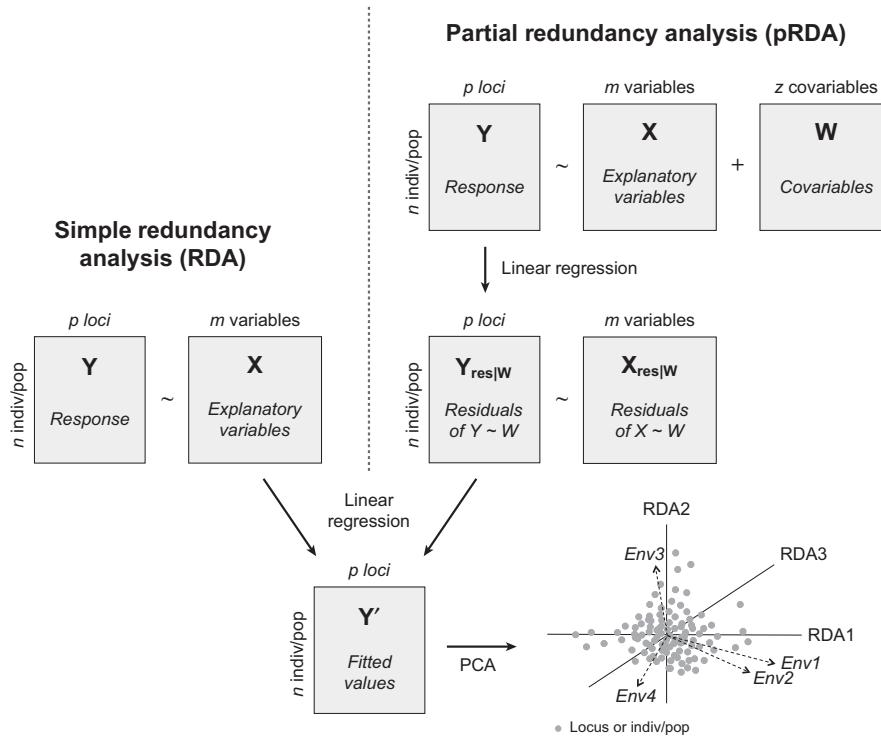
To illustrate different applications of RDA in landscape genomics, we used a large empirical dataset for lodgepole pine *Pinus contorta*,

### BOX 1 Understanding RDA

Redundancy analysis is a type of *asymmetric canonical analysis*, that is, a method that analyses two or more data frames simultaneously by combining *ordination* and *regression*. *Simple ordination* via principal components analysis (PCA) is likely familiar to most readers as a method for characterizing the main axes of variance in a matrix ( $\mathbf{Y}$ ) into uncorrelated (orthogonal) synthetic composite variables (the PC axes). For example, PCA of a genetic data matrix (i.e.  $n$  individuals or populations genotyped at  $p$  loci) can be used to infer similarities and differences among sampled individuals/populations based on their multilocus genotypes. By comparison, *multiple regression* is used to model a single response variable ( $\mathbf{y}$ ) using a set of multiple predictor variables ( $\mathbf{X}$ ). If we extend multiple regression to the multivariate response matrix used in PCA ( $\mathbf{Y}$ ), we are conducting an RDA. RDA therefore identifies linear combinations of the explanatory variables ( $\mathbf{X}$ ) that maximize the variance explained in linear combinations of the response ( $\mathbf{Y}$ ).

There are two steps to fitting a *simple RDA*: first, regress each variable in  $\mathbf{Y}$  on the  $\mathbf{X}$  explanatory variables and compute the fitted values ( $\mathbf{Y}'$ ); then conduct a PCA of the  $\mathbf{Y}'$  matrix to produce orthogonal constrained ('canonical') ordination axes, which are linear combinations of the predictors (Figure B1). Conditioning variables ( $\mathbf{W}$ ) can also be included to compute a *partial RDA* (pRDA). In this case, partial linear regression is used to adjust the linear effects of  $\mathbf{X}$  on  $\mathbf{Y}$  accounting for the covariables  $\mathbf{W}$  (Figure B1). In landscape genomics, covariables may include variables to account for population structure, such as ancestry coefficients, PC axes or spatial eigenvectors (e.g. Moran's eigenvector maps; Dray et al., 2006; Forester et al., 2018; Gutaker et al., 2020; Lasky et al., 2012).

When RDA is conducted with genomic response and environmental predictor data, the constrained ordination axes are comprised of *covarying sets of genotypes that are correlated with the multivariate environment*. This ability to model both multivariate response and multivariate explanatory variables, accommodating the genomic and environmental complexity found in nature, is what makes RDA such a unique and powerful tool in landscape genomics. We remind users that RDA identifies linear relationships among the response and predictor matrices; if nonlinear relationships are expected, other statistical frameworks may be more suitable, such as logistic regression (e.g. Joost et al., 2007), gradient forest (e.g. Fitzpatrick et al., 2021) or modified versions of RDA, such as polynomial RDA (Makarenkov & Legendre, 2002) and piecewise RDA (Vieira et al., 2019).



**FIGURE B1** A comparison of procedures for fitting a simple redundancy analysis (left) and a partial redundancy analysis (right). Both share a common last step of computing a principal components analysis (PCA) of the fitted values,  $\mathbf{Y}'$ . An example RDA biplot is shown, where RDA1, RDA2 and RDA3 are the orthogonal canonical axes, the dashed vectors represent the environmental predictors and the grey points represent either the loci or individuals/populations (depending on which the user chooses to represent). The arrangement of these items in the ordination space reflects their relationship with the canonical axes. Figure inspired by Legendre and Legendre (2012)

which includes both genomic (Mahony et al., 2020) and phenotypic (MacLachlan et al., 2017) data for 1,594 pine seedlings. The seedlings, grown in a common garden in Vancouver, British Columbia, came from 281 provenances (hereafter populations) distributed across the range of the species in British Columbia and Alberta (Canada). The genomic data consisted of 36,384 SNPs genotyped using a 50K SNP array (Mahony et al., 2020), including 3,934 intergenic SNPs used to account for neutral population structure and 32,407 SNPs used for downstream analyses. On this latter set of loci, we filtered out SNPs with a minor allele frequency <5% and estimated allele frequencies for populations at each of the remaining 29,941 loci. Population allele frequencies were then used as response variables in all RDA-based procedures described below. Note that RDA can efficiently analyse datasets containing hundreds of individuals and over 1 million SNPs (e.g. Zhang et al., 2019), making it particularly useful for contemporary, high-dimensional landscape genomic research. RDA can also accommodate individual-based datasets using genotypes as the response (coded as the count of one allele, i.e. 0/1/2). The lodgepole pine phenotypic data (used here to validate genomic offset calculations) included seedling height and average growth rate measured after three growing seasons (MacLachlan et al., 2017). Finally, we extracted the values of 27 bioclimate variables for all 281 source populations and for the common garden in Vancouver from ClimateNA (Wang et al., 2016). Using this combination of genomic, environmental and phenotypic data, we were able to test applications of RDA and compare our findings with published results for lodgepole pine. For all analyses, we use R v.3.6.3 (R Core Team, 2020) and the VEGAN package v.2.5-7 (Oksanen et al., 2015).

## 2.2 | Variable selection: Forward model building procedure

Variable selection is a critical preliminary step in ecological modelling, including landscape genomics. Depending on the analytical goals and our mechanistic understanding of the system, we may use either a predictive or explanatory approach to variable selection (Mac Nally, 2000). Here we present a predictive approach

using RDA with forward selection, where the goal is to maximize the genetic variance explained by a set of environmental predictors. This approach is well-suited when prediction is the primary objective and variable reduction is needed to avoid overfitting and collinearity (Dormann et al., 2013). The procedure begins with a test of significance of the global RDA model (all variables included). If this global model is significant (Blanchet et al., 2008), we begin forward selection with an empty model where multivariate genetic variation is explained by a fixed intercept. The model is then sequentially complexified by adding explanatory variables one by one, using two stopping criteria to prevent overestimation of the explained variance: a permutation-based significance test and the adjusted  $R^2$  of the global model. If the addition of a variable meets either stopping criteria (i.e. permuted  $p$ -value is greater than the threshold or the adjusted  $R^2$  decreases), the variable is rejected and forward selection ends. For example, Gibson and Moyle (2020) used forward selection with RDA to identify 14 abiotic variables associated with genetic variation in a wild tomato species. They found that variable selection differed when accounting for geographical distance among sampled populations using partial RDA, likely because of collinearity between geographical distance and environmental variation. Furthermore, other variable selection approaches based on RDA may better handle the issue of collinearity among variables (e.g. sparse RDA proposed in Csala et al., 2017) and could be promising additions for landscape genomic studies.

We applied forward selection to the lodgepole pine data using the *ordiR2step vegan* function with the following stopping criteria: variable significance of  $p < 0.01$  using 1,000 permutations, and the adjusted  $R^2$  of the global model (Supporting Information 1). In this and all subsequent analyses, we standardized the predictors (i.e. subtracted the mean and divided by the standard deviation) to ensure the variable units were comparable (Legendre & Legendre, 2012). Note that RDA analyses can include qualitative predictors such as soil type using binary ('dummy variable') coding (Legendre & Legendre, 2012). In total, 9 of the 27 bioclimate variables were selected (Table 1), showing a strong influence of temperature extremes, solar radiation and moisture deficit. Mahony et al. (2020) identified winter temperature and mean annual precipitation as important drivers of local adaptation in lodgepole pine. Our expanded

Variables	$R^2_{adj}$	Cum $R^2_{adj}$	F-value	p-value
MAR - Mean Annual solar Radiation	0.0220	0.022	7.31	0.002**
EMT - Extreme Minimum Temp.	0.0223	0.044	7.50	0.002**
MWMT - Mean Warmest Month Temp.	0.0099	0.054	3.90	0.002**
CMD - Climatic Moisture Deficit	0.0086	0.063	3.55	0.002**
Tave wt - Winter Mean Temp.	0.0041	0.067	2.22	0.002**
DD18 - Degree Days below 18°C	0.0030	0.070	1.87	0.002**
MAP - Mean Annual Prec.	0.0017	0.072	1.51	0.002**
Eref - Potential Evaporation (Hargreave)	0.0016	0.073	1.47	0.002**
PAS - Prec. As Snow	0.0018	0.075	1.53	0.002**

**TABLE 1** Climatic variables identified as significantly associated with genetic variation using forward variable selection with RDA (redundancy analysis)

\*\* $p \leq 0.01$ .

set of nine predictors may lead to differences in interpretation when searching for drivers of selection (Section 4).

An important issue with any predictive approach to variable selection is that statistical selection of variables optimizes variance explained but cannot, by itself, identify the ecological or mechanistic drivers of genetic variation (Mac Nally, 2000). To our knowledge, the impact of variable selection has not been specifically evaluated in a landscape genomics context, but we can learn from the species distribution modelling community, which has published an abundant literature on this topic (Araújo & Guisan, 2006; Dormann et al., 2013; Fournier et al., 2017). A misuse of variable selection can potentially bias both biological conclusions and downstream analyses. Especially when using large bioclimatic variable sets, pairwise predictor correlations can be very high, for example, among seasonal calculations of temperature or precipitation. While one variable may maximize variance explained, it may be another, correlated variable, potentially even unmeasured, that is the mechanistic driver of variation. The ubiquitous nature of environmental correlation means that it is critical to carefully investigate selected variables but also avoid overinterpretation of variable importance in downstream analyses unless mechanistic data support observed relationships.

### 2.3 | Variance partitioning: Disentangling the drivers of genetic variation

Another common goal of landscape genetics is to investigate the influence of environment [Isolation By Environment (IBE)], geographical features [Isolation By Distance (IBD) or Isolation By Resistance (IBR)] and demographic history (i.e. neutral genetic structure) on the distribution of genetic variation across species ranges (Orsini et al., 2013). Variance partitioning with pRDA can identify the contribution of these factors to reducing gene flow and triggering genetic divergence among populations. Variance partitioning estimates the proportion of variance explained by one set of explanatory variables (e.g. climate) once the influence of other variables (e.g. geography and/or neutral population structure) has been removed (Legendre & Legendre, 2012). When comparing the amount of variance explained

by each pRDA and the variance of a model including all explanatory variables (full model), it is possible to estimate the independent contribution of each set of variables together with the confounded effect induced by collinearity (Peres-Neto et al., 2006). Partial RDAs have been used to estimate the different contribution of climate and space in explaining genetic variation for *Arabidopsis thaliana* (Lasky et al., 2012), beech (Capblancq, Morin, et al., 2020), sorghum (Lasky et al., 2015) and rice (Gutaker et al., 2021). Recent studies have extended the use of pRDA to investigate the relative contributions of current circulation, connectivity and physical distance in shaping neutral or adaptive genetic variation in marine landscapes (Benestan et al., 2016; Xuereb et al., 2018).

We apply pRDA-based variance partitioning to the lodgepole pine data to decompose the contribution of climate, neutral population structure and geography in explaining genetic variation. We used three sets of variables: (a) nine bioclimate variables ('clim', Table 1); (b) three proxies of neutral genetic structure (population scores along the first three axes of a genetic PCA conducted on the 3,934 neutral loci; 'struct'); and (c) population coordinates (longitude and latitude) to characterize geographical variation ('geog'). We used population allele frequencies as the response variable in four different models (Table 2, Supporting Information 1).

Together, climate, neutral genetic structure and geography significantly explained 15% of the total genetic variance across lodgepole pine populations (Table 2). The effect of climate was highly significant even when controlling for population structure and geography, and explained 4% of total genetic variation (27% of the variation explained by the full model), suggesting association between genetic variation and environmental gradients (IBE). The pure effect of genetic structure accounted for 3% of total genetic variance (21% of explained variation) while geographical coordinates accounted for only 1% (5% of explained variation). Demography and geography are not always differentiated when conducting variance partitioning, since isolation by distance (IBD) is often the main driver (or a good proxy) of intraspecific genetic structure (Lasky et al., 2015). Nonetheless, it can be useful to assess their independent contributions when the studied species has experienced a complex demographic history in combination with IBD. Finally, we found that

**TABLE 2** The influence of climate, geography and neutral genetic structure on genetic variation decomposed with pRDA (partial redundancy analysis). Inertia is analogous to variance. The proportion of explainable variance represents the total constrained variation explained by the full model

Partial RDA models	Inertia	R <sup>2</sup>	p (>F)	Proportion of explainable Variance	Proportion of total Variance
Full model: F ~ clim. + geog. + struct.	85.3	0.146	0.001***	1	0.15
Pure climate: F ~ clim.   (geog. + struct.)	22.7	0.039	0.001***	0.27	0.04
Pure structure: F ~ struct.   (clim. + geog.)	17.7	0.030	0.001***	0.21	0.03
Pure geography: F ~ geog.   (clim. + struct.)	4.2	0.007	0.004***	0.05	0.01
Confounded climate/structure/geography	40.6			0.48	0.07
Total unexplained	498.4				0.85
Total inertia	583.6				1.00

\*\*\*p ≤ 0.001.

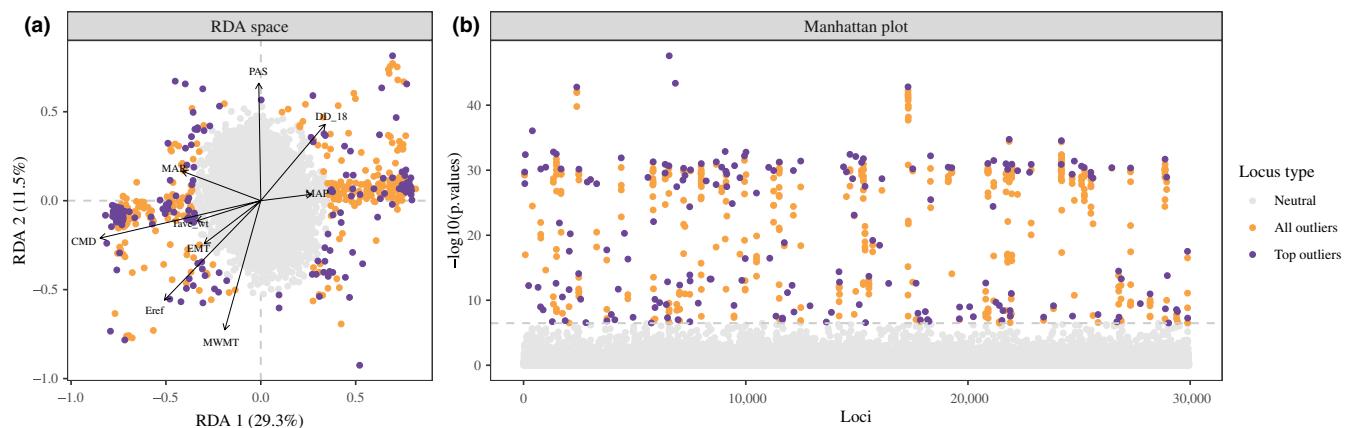
the largest proportion of genetic variance could not be uniquely attributed to any of the three sets of predictors (48% of explained variation). This confounded effect reflects a high degree of collinearity between explanatory variables and especially among the three groups of predictors (climate, structure and geography; Supporting Information 2). In other words, while we know that climate, neutral genetic structure and geography explain a large amount of genetic variation in lodgepole pine, almost half of this explainable variation is statistically impossible to associate with any specific predictor. This is critical information given that most landscape genomic studies look for correlation between climatic and genetic variation (i.e. GEA) and either assume no collinearity between environment and other drivers of genetic divergence or, on the contrary, totally remove this commonly explained variation by removing the effect of population structure or geographical variation (Frichot et al., 2015). In the first case, GEA detections could potentially be subject to high false-positive rates, while in the latter case detections might show high false-negative rates (Forester et al., 2018). Selecting an appropriate approach to account for demographic history and geographical distance is of major importance when searching for selection in the genome (Hoban et al., 2016). Variance partitioning can be a useful step to explore the (statistical) association among available descriptors, to better understand the covariation of environmental and genetic gradients, and to determine how much overall genetic variation is shaped by environmental, geographical and demographic factors before conducting further landscape genomics study.

## 2.4 | Genotype–Environment Associations: Identifying loci under selection

Genotype–environment associations regress genetic variation against environmental variables to identify loci showing statistical associations with the predictors. Identifying genetic adaptation to complex and/or multiple environmental gradients has triggered the

development of multivariate GEA methods (Fitzpatrick et al., 2021; Forester et al., 2018; Lasky et al., 2012). RDA appears to be, to date, one of the best-performing multivariate GEA solution, with higher true-positive and lower false-positive rates compared to univariate and other multivariate tests (Capblancq et al., 2018; Forester et al., 2018). RDA is also comparatively robust to sampling design, with higher power than other approaches even when sampling does not maximize environmental differentiation (Forester et al., 2018). Slightly different procedures have been described in the literature but all of them start by using an RDA model with multilocus genetic variation as response (either individual genotypes or population allele frequencies) and a set of environmental predictors as explanatory variables, sometime adding geographical coordinates or proxies for neutral genetic structure as conditioning variables in a pRDA to account for their confounded effects. By definition, the RDA will model linear relationships among the genetic markers and explanatory variables, and ordinate this variation along orthogonal axes that condense patterns of covariation between genetic markers and environmental predictors (Box 1). Each genetic marker included in the response will have a score or ‘loading’ along each of the newly built RDA axes. The loci putatively under selection are the markers showing extreme loadings along one or multiple axes. RDA has been used to find the genetic basis of adaptation in organisms with varied life-history traits including mammals (Razgour et al., 2019), echinoderms (Xuereb et al., 2018), fish (Brauer et al., 2018), birds (Friis et al., 2018; Stuart et al., 2020), crops (Lasky et al., 2015), annual plants (Lasky et al., 2012; Massatti & Knowles, 2020) and long-lived trees (Capblancq, Morin, et al., 2020; Steane et al., 2014). It has also been adapted to the analysis of gene expression data (Rougeux et al., 2019), phenotypes (Blanco-Pastor et al., 2020; Carvalho et al., 2020), copy number variants (Dorant et al., 2020) and haplotype variation (Morin et al., 2021).

Here, we used pRDA to identify candidate adaptive markers in the lodgepole pine data using population allele frequencies for



**FIGURE 1** Results of the genotype–environment association using partial RDA (redundancy analysis). (a) The projection of loci and environmental variables along the first two RDA axes. The locus scores are rescaled to an unshown axis for better visibility. (b) A Manhattan plot with the distribution of  $-\log_{10}(p\text{-values})$ . Loci identified as outliers ( $p\text{-value} < 3.4 \times 10^{-7}$  and the lowest  $p\text{-value}$  per contig) are coloured for both panels

the 29,941 loci as response and the 9 selected bioclimate variables (Table 1) as predictors. We followed the procedure described in Capblancq et al. (2018) where outliers are identified based on their extremeness along a distribution of Mahalanobis distances estimated between each locus and the centre of the RDA space using a certain number of axes ( $K$ ). We used the first two axes ( $K = 2$ ), which explained most of the genetic variance associated with the predictors (Figure 1) and partialled out the first three axes of a PCA of the neutral genetic markers (i.e. using only the 3,934 intergenic SNPs) to account for population structure. The Mahalanobis distances were corrected for the inflation factor (François et al., 2016) and transformed into  $p$ -values using a chi-squared distribution with  $K = 2$  degrees of freedom (Luu et al., 2017). The loci returning a  $p$ -value inferior to  $0.01/\text{number of tests}$  ( $0.01/29,941 = 3.4 \times 10^{-7}$ ), and the lowest  $p$ -value per contig were identified as candidate adaptive outliers (Supporting Information 1). Note that an alternative post-processing approach directly uses the loadings on each RDA axis independently, without a transformation to  $p$ -values (Forester et al., 2018; Lasky et al., 2012).

The RDA biplot illustrates how the genetic variation correlates with environmental predictors and how the selected genetic component varies along adaptive gradients (Figure 1a). The first axis, accounting for 29% of the variation, was primarily associated with moisture indices and summer precipitation variables, while the second axis (11.5%) was driven by summer temperatures and snow accumulation. We identified 583 loci showing extreme association with these two major axes of variation (i.e.  $p$ -values  $<3.4 \times 10^{-7}$ ; Figure 1b), reduced to a set of 183 unlinked outliers when retaining the best hit for each genomic contig. 126 of these 183 markers (69%) were also identified by Mahony et al. (2020) using Bayenv2 (Günther & Coop, 2013), and a third (326) of the 865 loci identified with Bayenv2 were identified in the larger set of 583 RDA candidates (Supporting Information 3).

In this example, we accounted for population structure, as is commonly recommended when conducting genome scans; however, there are cases where this approach can be overly conservative (Forester et al., 2018). Variance partitioning (Section 3) indicated a high proportion of confounded variance among climatic variation and neutral population structure, with no possibility to statistically disentangle the two factors (Table 2). In this case, correcting for population structure will remove statistical signal associated with both factors, reducing true positive rates. By contrast, not accounting for population structure can lead to an increased number of false-positive detections if signatures of neutral structure follow selective gradients (Excoffier et al., 2009). To investigate this issue in the lodgepole pine data, we compared the candidates identified with the partial RDA to those identified using a simple RDA. Our results show some degree of overlap between the two approaches, with 42% of the adaptive loci (424/1012) and 24% of the top hits per contig (95/399) being identified by both partial and simple RDA-based genome scans (Supporting Information 4). We emphasize that there is no correct answer to this common dilemma of confounded variation

in GEA, differentiation-based and genome-wide association analyses. Instead, researchers must balance their tolerance for false-negative and false-positive rates, and use an analytical approach that best matches the objectives of the study as well as the potential ability to validate candidate loci. Awareness of the trade-offs associated with decisions such as population structure correction will allow users to appropriately interpret their results, both statistically and biologically.

## 2.5 | Adaptive landscape: Projecting adaptive gradient(s) across space

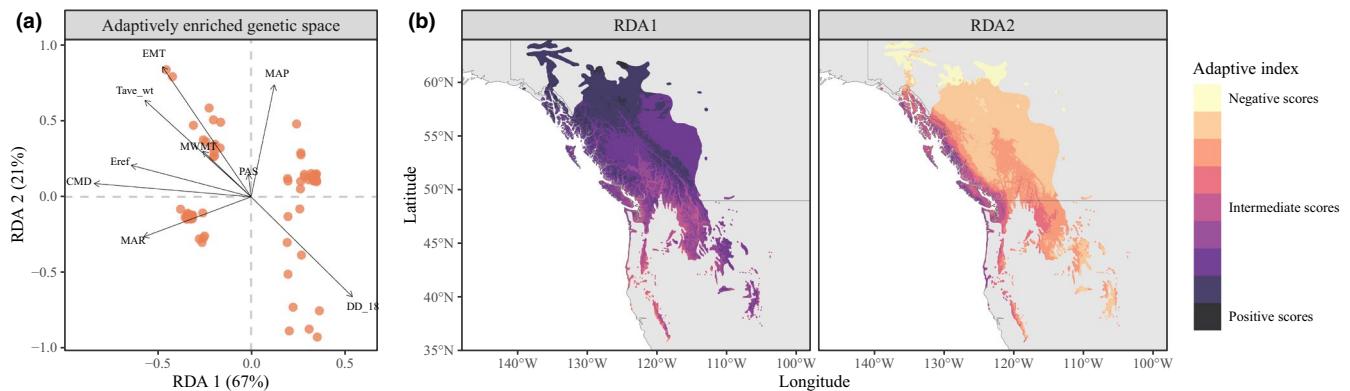
Once candidate adaptive markers and their potential environmental drivers have been identified with GEA, the genotype–environment relationship can be used to predict an index of genetic composition for any environment across the landscape (Capblancq et al., 2020). Once again, RDA provides a flexible framework for this analysis and has been used to map adaptive variation in common beech across the French Alps (Capblancq, Morin, et al., 2020), and to extrapolate eucalyptus adaptation to aridity in southeastern Australia (Steane et al., 2014).

We investigated the lodgepole pine adaptive landscape in western North America using the 95 loci identified as outliers (top hit per contig) by both the simple and partial RDAs conducted in Section 4. These loci were used as the multivariate response in an ‘adaptively enriched’ RDA using the nine selected bioclimate variables (Table 1) as explanatory variables. We followed the procedure described in Steane et al. (2014) that uses the scores of the environmental variables along the RDA axes to calculate a genetic-based index of adaptation for each environmental pixel of the landscape. This index is estimated independently for each RDA axis of interest using the formula:

$$\text{Adaptive Index} = \sum_{i=1}^n a_i b_i,$$

where  $a$  is the climatic variable score (loading) along the RDA axis,  $b$  is the standardized value for this particular variable at the focal pixel and  $i$  refers to one of the  $n$  different variables used in the RDA model (Supporting Information 1). The climatic variable scores used here are estimated based on how each predictor affects (magnitude and direction) adaptive genetic variation along the RDA axes. The adaptive index thus provides an estimate of adaptive genetic similarity or difference of all pixels on the landscape as a function of the values of the environmental predictors at that location.

The ‘adaptively enriched genetic space’ shows how the genetic markers correlate with the different bioclimate variables as well as how the bioclimate variables correlate with each other (Figure 2a). De facto, most of the variation is aggregated on RDA1 (67% of the variance) along which allele frequencies were either associated with high summer precipitation and a high number of degree days below 18°C (positive RDA1 scores) or with low moisture deficit, high



**FIGURE 2** Lodgepole pine adaptive landscape with (a) the adaptively enriched genetic space showing association between adaptive loci and climatic drivers of adaptation and (b) spatial projection of adaptive genetic turnover across the species range. The locus loadings in panel (a) are rescaled to an unshown axis for better visibility

potential evapotranspiration and high solar radiation (negative RDA1 scores). The extrapolation of this adaptive gradient on the landscape (Figure 2b) showed how RDA1 index contrasts southern and low elevation areas (negative RDA1 scores) from high elevation and northern areas (positive RDA1 scores). The adaptive gradients associated with RDA2 (21% of the variance) differentiated the coastal and southern areas characterized by warmer temperatures and more precipitation during winter (positive RDA2 scores) to inland and mountainous areas characterized by colder extreme temperatures and later start of the frost-free period and more degree days below 18°C (negative RDA2 scores). Mahony et al. (2020), using a totally different approach, found a similar gradient of genetic adaptation for lodgepole pine, with groups of adaptive loci that varied along the elevation gradient (Cluster 4 in their study) or distinguished coastal from continental populations (Clusters 4 and 6 in their study).

Using RDA to visualize adaptive landscapes is straightforward, flexible and useful when studying the distribution of adaptive alleles across species ranges or when looking for adaptive units within species for conservation or management purposes. Nonetheless, any prediction procedure such as this relies on the extrapolation of correlative relationships and carries potential pitfalls (Yates et al., 2018). Predicting any genetic-based adaptive index in unsampled areas could be strongly biased if the value or combination of the environmental predictors in those projected areas exceed those used to train the model. Similarly, any projection could be inaccurate if the genetic-environment relationship is not well characterized, either because the scale of the analyses (e.g. size of the pixel) does not capture the complexity of the association or because the distribution of the identified genetic variation is not optimal across the current and/or future environmental landscape (e.g. due to recent demographic history, barriers to gene flow, genetic drift in small populations, fluctuating trait heritability, varying selection pressures or genetic architecture; Kardos & Luikart, 2021).

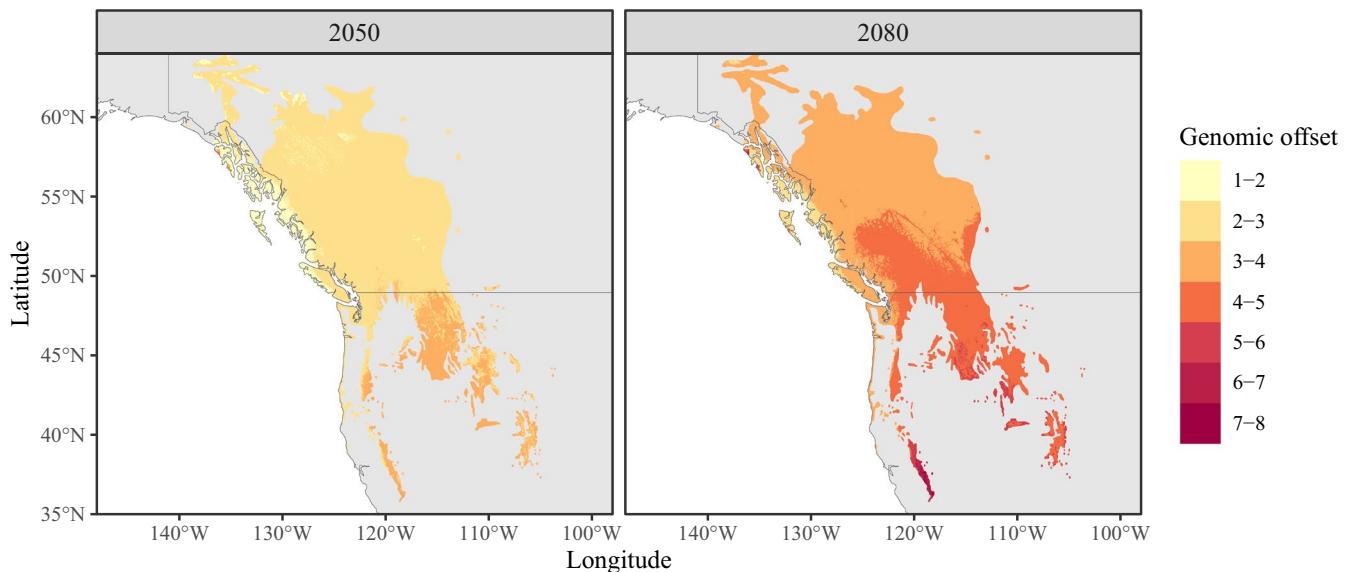
Finally, beyond predicting adaptive indices, RDA can be used to estimate individual genotypes or population allele frequencies as a function of environmental variation. For example, Carvalho et al. (2020) used RDA to identify suitable genetic provenances for

revegetation efforts at highly disturbed mining sites, an application that could also be used to inform climate-adjusted provenancing and assisted gene flow efforts (Aitken & Bemmels, 2016). There are potential statistical problems associated with these predictions, however, such as negative estimates of allele counts or frequencies. One attempt to address these issues combines calibration and prediction of genotype-environment models using RDA with the use of generalized additive models to constrain predictions within a realistic range (Kindt, 2021). Efforts such as these will continue to expand and refine the RDA toolkit for landscape genomics applications.

## 2.6 | Predicting local maladaptation: Genomic offset

The genetic-environment relationship used above can be extrapolated to future environments to predict a potential shift in adaptive optimum induced by climate change (Capblancq, Fitzpatrick, et al., 2020). If the change of local environments is significant enough that populations are not able to track it, *in situ*, by modifying their genetic composition, it could translate into maladaptation and eventually reduce population fitness (Bay et al., 2018). Investigating how best to predict future maladaptation has become a topic of recent interest (Capblancq, Fitzpatrick, et al., 2020; Fitzpatrick & Keller, 2015; Rellstab et al., 2016), including through the use of RDA models (Capblancq, Morin, et al., 2020). The RDA-based method to predict future maladaptation is relatively simple. As described above, RDA can be used to predict the optimal adaptive index for each environmental pixel under consideration (Section 5), using both current and future environmental conditions. The difference between the two predictions provides an estimate of the shift in adaptive index that would be required to track climate change. This proxy has been successively named genetic/genomic offset (Fitzpatrick & Keller, 2015), risk of non-adaptedness (Rellstab et al., 2016) or genomic vulnerability (Bay et al., 2018).

We estimated genomic offset for lodgepole pine by applying the same procedure used for current climate (Section 5) to future climate projections (Wang et al., 2016), predicting adaptive indices based on RDA1 and RDA2. We then calculated the Euclidean



**FIGURE 3** Mean predicted genomic offset for lodgepole pine in 2050 and 2080 using the RCP8.5 climate scenario and the ensemble projection of 15 CMIP 5 AOGCMs (see <http://climatena.ca>)

distance between each pixel score under current and future climates for each of the first two RDA axes to produce a genomic offset value (Supplementary material 1). Genomic offsets estimated for 2050 and 2080 were mapped across the species range to show the areas where maladaptation may pose a risk under future climate change.

Our model predicted a contrasted pattern for 2050 (Figure 3), with low values of genomic offset in the Canadian part of the species range and a higher risk of maladaptation in the southern edge (genomic offset >1). The prediction for 2080 shows a clear increase in genomic offset, especially in the Californian Sierra Nevada and southern Rocky Mountains, while the coastal and northern populations will be less impacted (Figure 3).

A challenge associated with any predictive method is validating the predictions, which are usually based on statistical association only (Capblancq, Fitzpatrick, et al., 2020). Here, we use an approach that takes advantage of the common garden experiment conducted in lodgepole pine. We estimated genomic offset between the climate of each source population and the climate of the garden where the seedlings were grown (in Vancouver, BC, Canada) and were then able to test the influence of the predicted genomic offset on the seedling fitness traits (Fitzpatrick et al., 2021). We found a strongly negative relationship between genomic offset and seedling fitness in the common garden (Height:  $p$ -value =  $1.2 \times 10^{-9}$  and  $R^2 = 0.12$ ; Growth rate:  $p$ -value =  $7.9 \times 10^{-15}$  and  $R^2 = 0.19$ ; Supporting Information 5), indicating that fitness decreases as genomic offset increases. This validation approach assumes that a transfer across the range of the species (space) could be equivalent to a future change in local environmental conditions (time; Blois et al., 2013). Common garden data could also be used to assess the capacity of the model to predict beyond its training range if the common garden experienced environmental conditions outside the range of the sampled populations, including extreme values or new combinations of environmental

variables (Fitzpatrick et al., 2018). This validation procedure confirms that genomic offset can be a powerful tool to design management or conservation strategies in the context of climate change.

### 3 | CONCLUSIONS

Redundancy analysis is a Swiss Army Knife for landscape genomic analyses: a multipurpose, adaptable and versatile statistical framework for identifying, evaluating and projecting relationships between genetic and environmental variation. Because RDA models both multivariate responses and predictors, it effectively identifies covarying sets of loci that are correlated with the multivariate environment, in contrast to univariate methods. This allows RDA to better accommodate and represent the genomic and environmental complexity found in nature. Additionally, RDA can be used to model genetic data at the individual or population levels, can account for conditioning variables, such as data reflecting population structure, and is computationally efficient when applied to modern, large genomic datasets. These characteristics make RDA an intuitive fit for the analysis of covariation between high-dimensional genomic data and complex landscape and environmental features.

Redundancy analysis has one important limitation: it models linear relationships among response and predictors, meaning that nonlinear relationships will not be detected. In landscape genomics and GEA studies, linearity is often assumed even if it does not always capture the exact relationship between environmental gradients and the induced adaptive genetic variation. Fortunately, there are nonlinear statistical methods that have been adapted to landscape genomic analyses, such as logistic regression and gradient forest. Similarly, variations on the classic RDA method, such as polynomial and piecewise RDA, have been developed to accommodate

nonlinear relationships, though these have yet to see development or use in landscape genomics. Ongoing testing of the strengths, limitations and complementary applications of the varied statistical methods, including RDA, in landscape genomics will continue to be important in improving the insights and value of landscape genomic applications.

## ACKNOWLEDGEMENTS

We would like to thank Colin Mahony and Brandon Lind for producing and sharing the data used in this review, including previously unpublished lists of outlier loci. We would also like to thank Jesse Lasky for his helpful feedback on the manuscript, as well as Sean Schoville and an anonymous reviewer for their comments which greatly improved the paper. T.C. was supported by a postdoctoral associate programme awarded by the Office of the Vice President for Research of the University of Vermont. B.R.F. was supported by a David H. Smith Conservation Research Fellowship.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHORS' CONTRIBUTIONS

All the work conducted for this review study, from data analysis to manuscript writing, result from a team effort shared by T.C. and B.R.F.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13722>.

## DATA AVAILABILITY STATEMENT

Data available from the Dryad Digital Repository <https://doi.org/10.5061/dryad.56j8vq8>. (Mahony et al., 2019) and ClimateNA (Wang et al., 2016). Scripts and a Tutorial are available on Github at the following address <https://github.com/Capblancq/RDA-landscape-genomics> and Zenodo (Capblancq & Forester, 2021).

## ORCID

Thibaut Capblancq  <https://orcid.org/0000-0001-5024-1302>

Brenna R. Forester  <https://orcid.org/0000-0002-1608-1904>

## REFERENCES

- Aitken, S. N., & Bemmels, J. B. (2016). Time to get moving: Assisted gene flow of forest trees. *Evolutionary Applications*, 9, 271–290. <https://doi.org/10.1111/eva.12293>
- Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677–1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
- Balkenhol, N., Dudaniec, R. Y., Krutovsky, K. V., Johnson, J. S., Cairns, D. M., Segelbacher, G., Selkoe, K. A., von der Heyden, S., Wang, I. J., Selmoni, O., & Joost, S. (2019). Landscape genomics: Understanding relationships between environmental heterogeneity and genomic characteristics of populations. In O. P. Rajora (Ed.), *Population genomics: Concepts, approaches and applications* (pp. 261–322). Springer International Publishing. [https://doi.org/10.1007/13836\\_2017\\_2](https://doi.org/10.1007/13836_2017_2)
- Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, 359, 83–86. <https://doi.org/10.1126/science.aan4380>
- Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., Rochette, R., & Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, 25(20), 5073–5092. <https://doi.org/10.1111/mec.13811>
- Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology*, 89, 2623–2632. <https://doi.org/10.1890/07-0986.1>
- Blanco-Pastor, J. L., Barre, P., Keep, T., Ledauphin, T., Escobar-Gutiérrez, A., Roschanski, A. M., Willner, E., Dehmer, K. J., Hegarty, M., Muylle, H., Veerckman, E., Vandepoele, K., Ruttink, T., Roldán-Ruiz, I., Manel, S., & Sampoux, J.-P. (2020). Canonical correlations reveal adaptive loci and phenotypic responses to climate in perennial ryegrass. *Molecular Ecology Resources*, 21(3), 849–870. <https://doi.org/10.1111/1755-0998.13289>
- Blois, J. L., Zarnetske, P. L., Fitzpatrick, M. C., & Finnegan, S. (2013). Climate change and the past, present, and future of biotic interactions. *Science*, 341, 499–504. <https://doi.org/10.1126/science.1237184>
- Brauer, C. J., Unmack, P. J., Smith, S., Bernatchez, L., & Beheregaray, L. B. (2018). On the roles of landscape heterogeneity and environmental variation in determining population genomic structure in a dendritic system. *Molecular Ecology*, 27, 3484–3497. <https://doi.org/10.1111/mec.14808>
- Bray, R. J., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27, 325–347. <https://doi.org/10.2307/1942268>
- Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic prediction of (Mal)adaptation across current and future climatic landscapes. *Annual Review of Ecology Evolution and Systematics*, 51, 245–271. <https://doi.org/10.1146/annurev-ecolsys-020720-042553>
- Capblancq, T., & Forester, B. (2021). Data from: Capblancq/RDA-landscape-genomics: First release RDA-landscape-genomics (v1.0.0). Zenodo, <https://doi.org/10.5281/zenodo.5494709>
- Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, 18, 1223–1233. <https://doi.org/10.1111/1755-0998.12906>
- Capblancq, T., Morin, X., Gueguen, M., Renaud, J., Lobreaux, S., & Bazin, E. (2020). Climate-associated genetic variation in *Fagus sylvatica* and potential responses to climate change in the French Alps. *Journal of Evolutionary Biology*, 33, 783–796.
- Carvalho, C. S., Forester, B. R., Mitre, S. K., Alves, R., Imperatriz-Fonseca, V. L., Ramos, S. J., Resende-Moreira, L. C., Siqueira, J. O., Trevelin, L. C., Caldeira, C. F., Gastauer, M., & Jaffé, R. (2020). Combining genotype, phenotype, and environmental data to delineate site-adjusted provenance strategies for ecological restoration. *Molecular Ecology Resources*, 44–58. <https://doi.org/10.1111/1755-0998.13191>
- Csala, A., Voorbraak, F. P. J. M., Zwinderman, A. H., & Hof, M. H. (2017). Sparse redundancy analysis of high-dimensional genetic and genomic data. *Bioinformatics*, 33, 3228–3234. <https://doi.org/10.1093/bioinformatics/btx374>
- Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., Normandieu, E., Rochette, R., & Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype-temperature association in a marine species. *Molecular Ecology*, 29, 4765–4782. <https://doi.org/10.1111/mec.15565>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J.,

- Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196, 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, 103, 285–298. <https://doi.org/10.1038/hdy.2009.74>
- Fitzpatrick, M. C., Blois, J. L., Williams, J. W., Nieto-Lugilde, D., Maguire, K. C., & Lorenz, D. J. (2018). How will climate novelty influence ecological forecasts? Using the Quaternary to assess future reliability. *Global Change Biology*, 24, 3575–3586. <https://doi.org/10.1111/gcb.14138>
- Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y., & Keller, S. R. (2021). Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources*, <https://doi.org/10.1111/1755-0998.13374>
- Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, 18, 1–16. <https://doi.org/10.1111/ele.12376>
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, 27, 2215–2233. <https://doi.org/10.1111/mec.14584>
- Fournier, A., Barbet-Massin, M., Rome, Q., & Courchamp, F. (2017). Predicting species distribution combining multi-scale drivers. *Global Ecology and Conservation*, 12, 215–226. <https://doi.org/10.1016/j.gecco.2017.11.002>
- François, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, 25, 454–469. <https://doi.org/10.1111/mec.13513>
- Frichot, E., Schoville, S. D., De Villemereuil, P., Gaggiotti, O. E., & François, O. (2015). Detecting adaptive evolution based on association with ecological gradients: Orientation matters! *Heredity*, 115, 22–28. <https://doi.org/10.1038/hdy.2015.7>
- Friis, G., Fandos, G., Zellmer, A. J., McCormack, J. E., Faircloth, B. C., & Milá, B. (2018). Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. *Molecular Ecology*, 27, 5137–5153. <https://doi.org/10.1111/mec.14946>
- Gibson, M. J. S., & Moyle, L. C. (2020). Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species. *Molecular Ecology*, 29, 2204–2217. <https://doi.org/10.1111/mec.15477>
- Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, 195, 205–220. <https://doi.org/10.1534/genetics.113.152462>
- Guataker, R. M., Groen, S. C., Bellis, E. S., Choi, J. Y., Pires, I. S., Bocinsky, R. K., Slatton, E. R., Wilkins, O., Castillo, C. C., Negrão, S., Oliveira, M. M., Fuller, D. Q., Guedes, J. A. D' A., Lasky, J. R., & Purugganan, M. D. (2020). Genomic history and ecology of the geographic spread of rice. *Nature Plants*, 6, 492–502. <https://doi.org/10.1038/s41477-020-0659-6>
- Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist*, 188, 379–397. <https://doi.org/10.1086/688018>
- Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., & Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*, 16, 3955–3969. <https://doi.org/10.1111/j.1365-294X.2007.03442.x>
- Kardos, M., & Luikart, G. (2021). The genetic architecture of fitness drives population viability during rapid environmental change. *The American Naturalist*, 197, 511–525. <https://doi.org/10.1086/713469>
- Kindt, R. (2021). AlleleShift: An R package to predict and visualize population-level changes in allele frequencies in response to climate change. *PeerJ*, 9, e11534. <https://doi.org/10.7717/peerj.11534>
- Lasky, J. R., Des Marais, D. L., McKay, J. K., Richards, J. H., Juenger, T. E., & Keitt, T. H. (2012). Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. *Molecular Ecology*, 21, 5512–5529.
- Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., Juenger, T. E., Hyma, K., Acharya, C., Mitchell, S. E., Buckler, E. S., Brenton, Z., Kresovich, S., & Morris, G. P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, 1, 1–14. <https://doi.org/10.1126/sciadv.1400218>
- Legendre, P., & Legendre, L. (2012). *Numerical ecology* (Vol. 24, 3rd ed.). Elsevier.
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17, 67–77.
- Mac Nally, R. (2000). Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – ‘predictive’ and ‘explanatory’ models. *Biodiversity and Conservation*, 9, 655–671.
- MacLachlan, I. R., Wang, T., Hamann, A., Smets, P., & Aitken, S. N. (2017). Selective breeding of lodgepole pine increases growth and maintains climatic adaptation. *Forest Ecology and Management*, 391, 404–416. <https://doi.org/10.1016/j.foreco.2017.02.008>
- Mahony, C., MacLachlan, I., Lind, B., Yoder, J., Wang, T., & Aitken, S. (2019). Data from: Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. *Dryad Digital Repository*, <https://doi.org/10.5061/dryad.56j8vq8>
- Mahony, C. R., MacLachlan, I. R., Lind, B. M., Yoder, J. B., Wang, T., & Aitken, S. N. (2020). Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. *Evolutionary Applications*, 13, 116–131. <https://doi.org/10.1111/eva.12871>
- Makarenkov, V., & Legendre, P. (2002). Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. *Ecology*, 83, 1146–1161. [https://doi.org/10.1890/0012-9658\(2002\)083\[1146:NRAACC\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[1146:NRAACC]2.0.CO;2)
- Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution*, 28, 614–621. <https://doi.org/10.1016/j.tree.2013.05.012>
- Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S., Scribner, K. T., Bonin, A., & Fortin, M.-J. (2010). Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, 19, 3760–3772. <https://doi.org/10.1111/j.1365-294X.2010.04717.x>
- Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18, 189–197. [https://doi.org/10.1016/S0169-5347\(03\)00008-9](https://doi.org/10.1016/S0169-5347(03)00008-9)
- Massatti, R., & Knowles, L. L. (2020). The historical context of contemporary climatic adaptation: A case study in the climatically dynamic and environmentally complex southwestern United States. *Ecography*, 43, 735–746. <https://doi.org/10.1111/ecog.04840>
- Mitton, J. B., Linhart, Y. B., Hamrick, J. L., & Beckman, J. S. (1977). Observations on the genetic structure and mating system of ponderosa pine in the Colorado front range. *Theoretical and Applied Genetics*, 51, 5–13. <https://doi.org/10.1007/BF00306055>

- Morin, P. A., Forester, B. R., Forney, K. A., Crossman, C. A., Hancock-Hanser, B. L., Robertson, K. M., Barrett-Lennard, L. G., Baird, R. W., Calambokidis, J., Gearin, P., Hanson, M. B., Schumacher, C., Harkins, T., Fontaine, M. C., Taylor, B. L., & Parsons, K. M. (2021). Population structure in a continuously distributed coastal marine species, the harbor porpoise, based on microhaplotypes derived from poor-quality samples. *Molecular Ecology*, 30, 1457–1476. <https://doi.org/10.1111/mec.15827>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., & Wagner, H. (2015). *vegan: Community ecology package*. R Package Version 2.3-2. Retrieved from <https://CRAN.R-project.org/package=vegan/>; <https://github.com/vegandevels/vegan/>; <https://doi.org/10.4236/issn.17477067>
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, 22, 5983–5999. <https://doi.org/10.1111/mec.12561>
- Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology*, 87, 2614–2625. [https://doi.org/10.1890/0012-9658\(2006\)87\[2614:VPOSDM\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2)
- R Core Team. (2020). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org/>
- Razgour, O., Forester, B., Taggart, J. B., Bekaert, M., Juste, J., Ibáñez, C., Puechmaille, S. J., Novella-Fernandez, R., Alberdi, A., & Manel, S. (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 10418–10423. <https://doi.org/10.1073/pnas.1820663116>
- Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24, 4348–4370. <https://doi.org/10.1111/mec.13322>
- Rellstab, C., Zoller, S., Walther, L., Lesur, I., Pluess, A. R., Graf, R., Bodénès, C., Sperisen, C., Kremer, A., & Gugerli, F. (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Molecular Ecology*, 25(23), 5907–5924. <https://doi.org/10.1111/mec.13889>
- Rougeux, C., Gagnaire, P. A., Praebel, K., Seehausen, O., & Bernatchez, L. (2019). Polygenic selection drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic sister species complex. *Molecular Ecology*, 28, 4388–4403. <https://doi.org/10.1111/mec.15226>
- Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., Paxton, E. H., & Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters*, 21, 1085–1096. <https://doi.org/10.1111/ele.12977>
- Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E., & Byrne, M. (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology*, 23, 2500–2513. <https://doi.org/10.1111/mec.12751>
- Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Dezzani, R., Delmelle, E., Vierling, L., & Waits, L. P. (2007). Putting the 'landscape' in landscape genetics. *Heredity*, 98, 128–142. <https://doi.org/10.1038/sj.hdy.6800917>
- Storfer, A., Patton, A., & Fraik, A. K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Frontiers in Genetics*, 9. <https://doi.org/10.3389/fgene.2018.00068>
- Stuart, K. C., Cardilini, A. P. A., Cassey, P., Richardson, M. F., Sherwin, W. B., Rollins, L. A., & Sherman, C. D. H. (2020). Signatures of selection in a recent invasion reveal adaptive divergence in a highly vagile invasive species. *Molecular Ecology*, 1–16. <https://doi.org/10.1111/mec.15601>
- Vieira, D. C., Brustolin, M. C., Ferreira, F. C., & Fonseca, G. (2019). *segRDA: An R package for performing piecewise redundancy analysis*. *Methods in Ecology and Evolution*, 10, 2189–2194.
- Wang, T., Hamann, A., Spittlehouse, D., & Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS ONE*, 11, 1–17. <https://doi.org/10.1371/journal.pone.0156720>
- Whittaker, R. H. (1956). Vegetation of the Great Smoky Mountains. *Ecological Monographs*, 26, 1–80. <https://doi.org/10.2307/1943577>
- Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M. J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology*, 27(10), 2347–2364. <https://doi.org/10.1111/mec.14589>
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. M. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33, 790–802. <https://doi.org/10.1016/j.tree.2018.08.001>
- Zhang, M., Suren, H., Holliday, J. A., & Gaut, B. (2019). Phenotypic and genomic local adaptation across latitude and altitude in *Populus trichocarpa*. *Genome Biology and Evolution*, 11, 2256–2272. <https://doi.org/10.1093/gbe/evz151>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

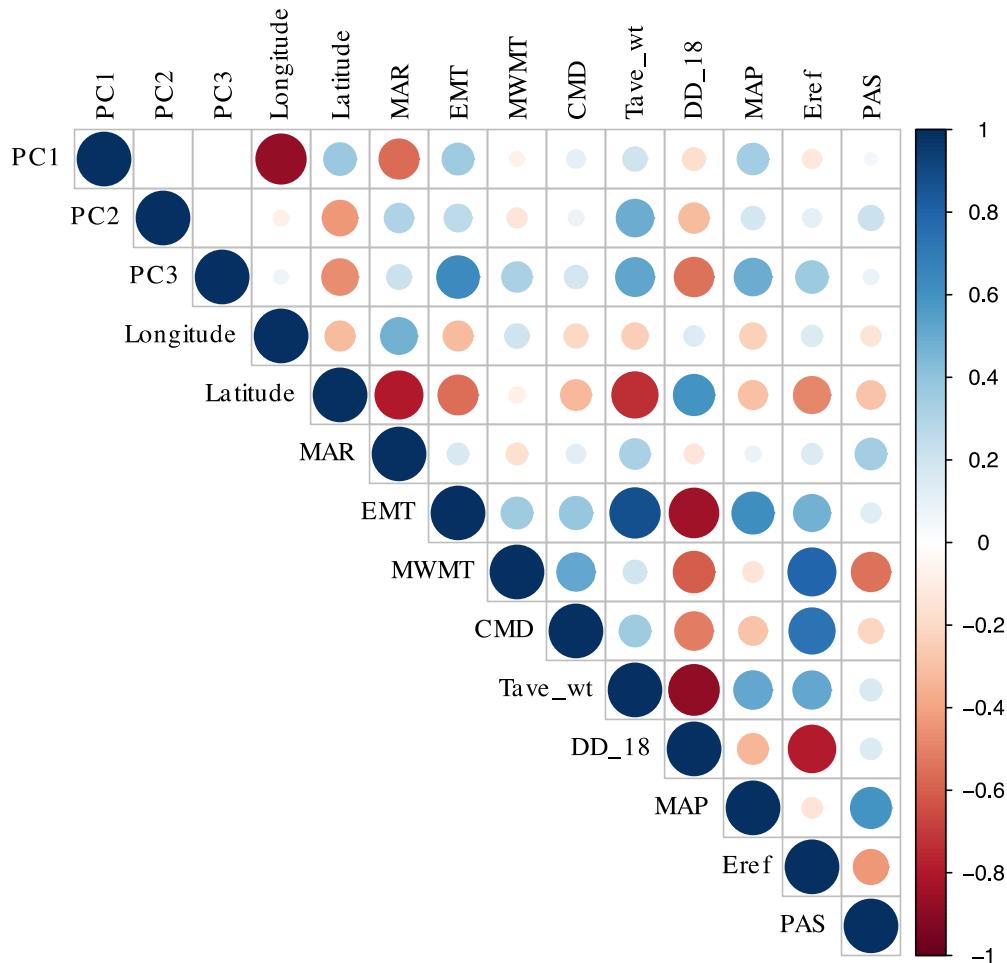
**How to cite this article:** Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, 00, 1–12. <https://doi.org/10.1111/2041-210X.13722>

# Redundancy Analysis (RDA): a Swiss Army knife for landscape genomics

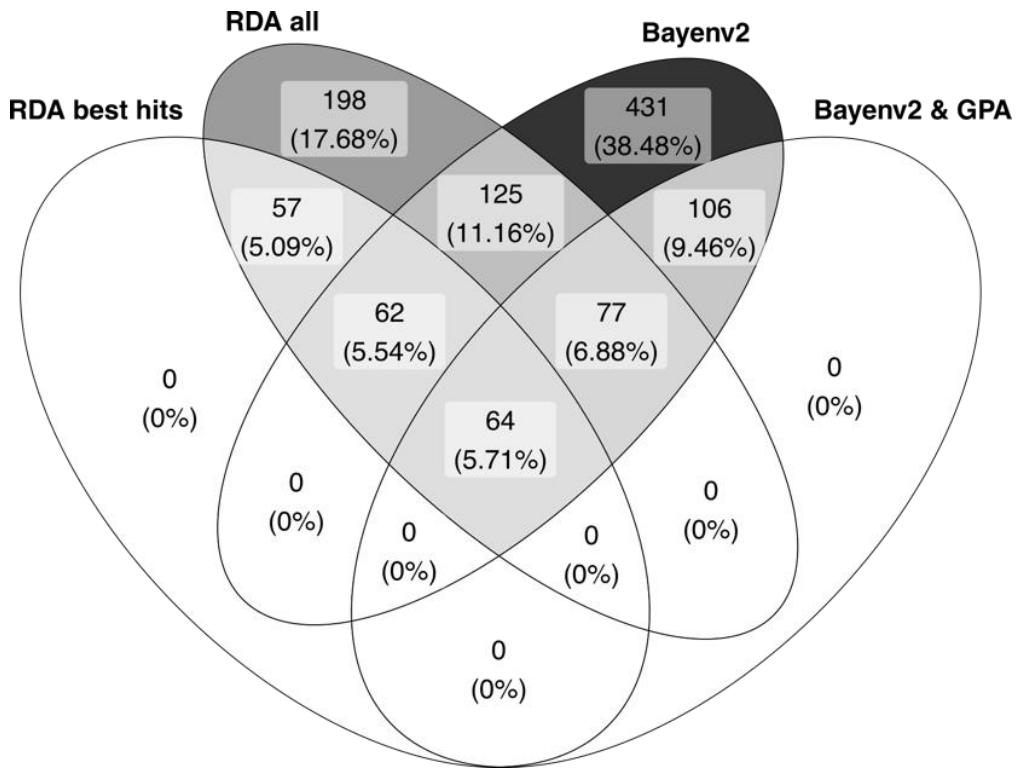
Thibaut CAPBLANCQ and Brenna R. FORESTER

## SUPPORTING INFORMATION

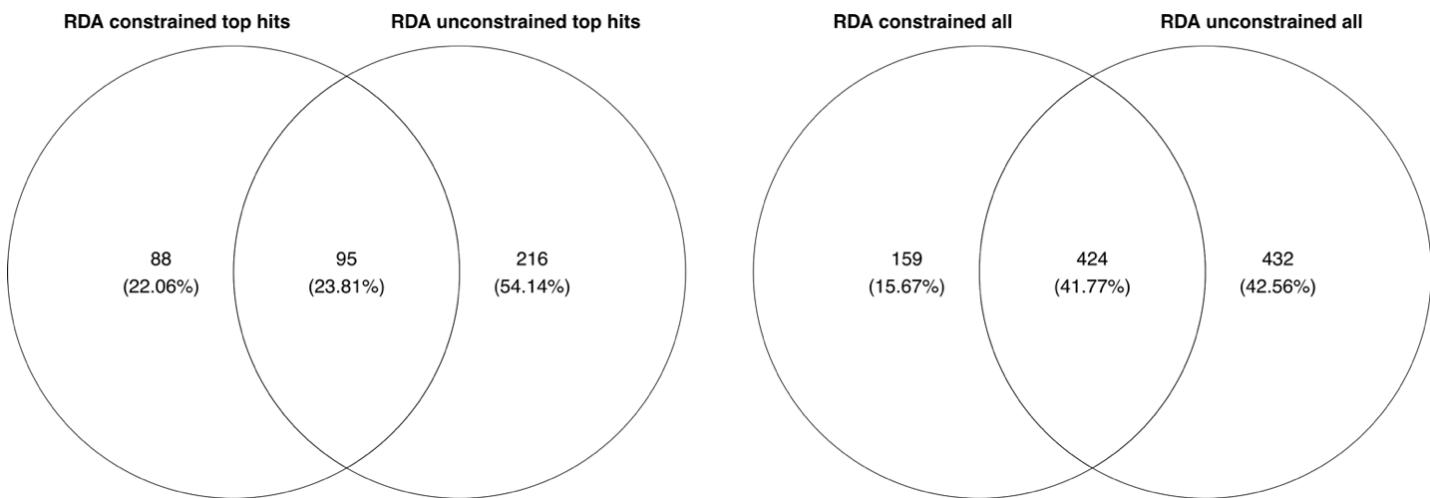
**Supporting information 1:** A detailed R tutorial to conduct the different RDA-based procedures described in this review is available at <https://github.com/Capblancq/RDA-landscape-genomics>



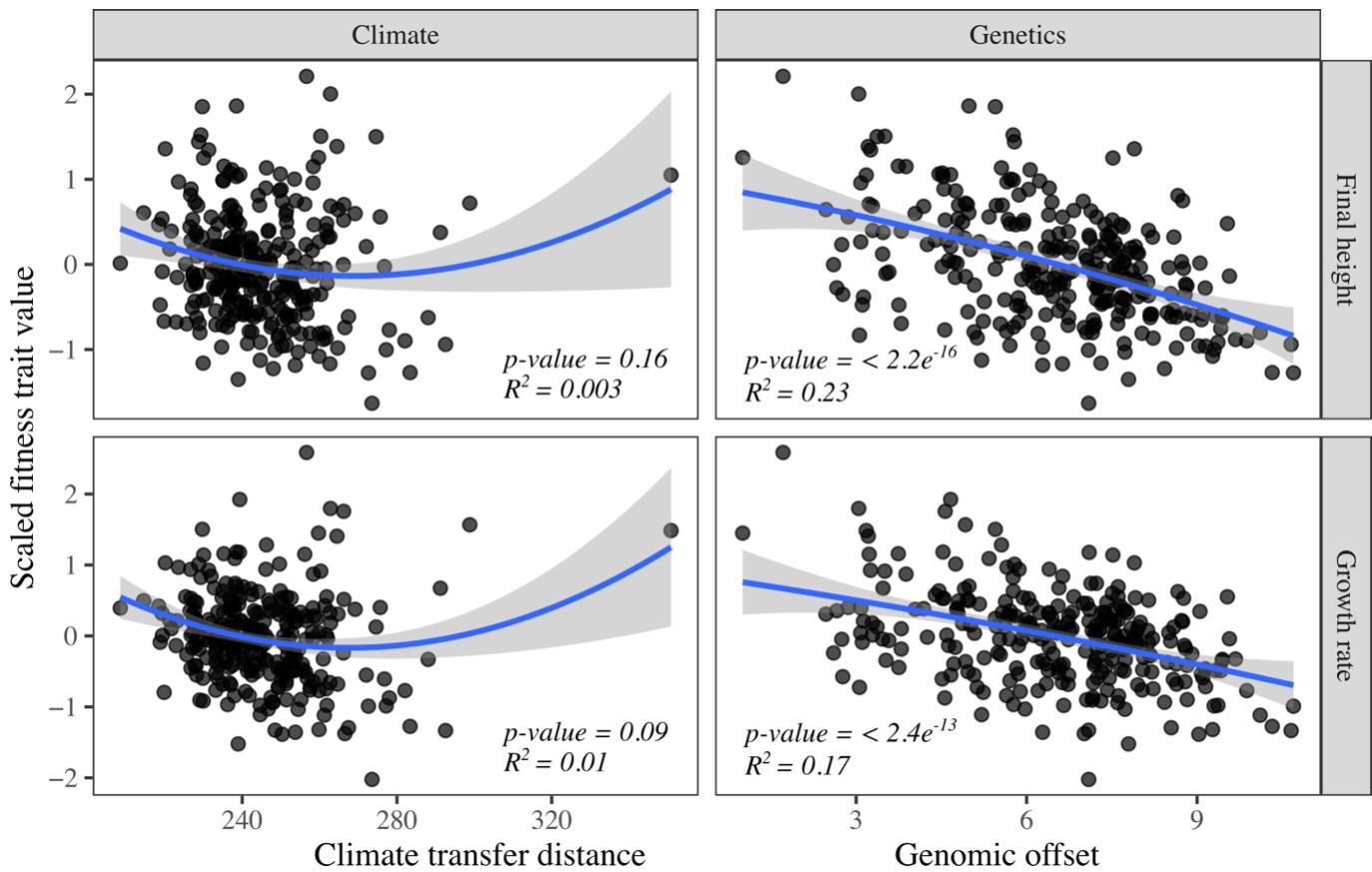
**Supporting information 2:** Correlogram of the different variables estimated at each sampling site, including the nine selected climatic variables, the three first genetic PCs, longitude and latitude.



**Supporting information 3:** Venn diagram showing the degree of overlap between the genetic outliers identified with the unconstrained RDA-based genome scan and the outliers identified by Mahony and colleagues (2020) using Bayenv2 and a genotype-phenotype association (GPA) method.



**Supporting information 4:** Venn diagram showing the degree of overlap between the genetic outliers identified with the constrained or unconstrained RDA-based genome scans.



**Supporting information 5:** Association between climate transfer distance (left column) and genomic offset (right column) and two fitness traits measured on the seedlings of the common garden in Vancouver (rows). Climate transfer distances are Mahalanobis distances estimated between the garden and provenances climates using the 9 selected climatic variables. Genomic offset was predicted using the RDA-based method described in Section 5 and using the two first RDA axes. P-value and  $R^2$  resulting from quadratic regressions together with the fitted regression lines and associated standard deviation show the strength of the link between genomic offset and the two traits.