

# Spatial areas of genotype probability: Predicting the spatial distribution of adaptive genetic variants under future climatic conditions

Estelle Rochat<sup>1</sup>  | Oliver Selmoni<sup>1</sup>  | Stéphane Joost<sup>1,2</sup> 

<sup>1</sup>Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

<sup>2</sup>The NEXTGEN Consortium

## Correspondence

Stéphane Joost, Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

Email: stephane.joost@epfl.ch

## Abstract

In a context of rapid global change, one of the key components for the survival of species is their genetic evolutionary potential for adaptation. Many methods have been developed to identify genetic variants underpinning adaptation to climate, but few tools were made available to integrate this knowledge into conservation management. We present here the SPatial Areas of Genotype probability (SPAG), a method to transpose the results of genotype–environment association studies into an evolutionary potential spatial prediction framework. We define a univariate model predicting the spatial distribution of a single-locus adaptive genotype and three multivariate models allowing the integration of several adaptive loci in a composite genotype. Unlike existing methods, SPAGs provide (a) a flexible approach to combine loci under different types of intergenic relationships and (b) a cross-validation framework to assess the pertinence of evolutionary potential predictions. SPAGs can be integrated with climate change projections to forecast the future spatial distribution of genotypes. The analysis of the mismatch between current and future SPAGs (“genomic offset”) makes it possible to identify vulnerable populations potentially lacking the adaptive genotypes necessary for future survival. We tested the SPAG approach on a simulated population and applied it to characterize the evolutionary potential of 161 Moroccan goats to bioclimatic conditions. We identified seven regions of the Moroccan goat genome strongly associated with the precipitation seasonality and used the SPAG approach to predict the evolutionary potential. We then forecasted the shift in SPAGs under a strong climate change scenario and uncovered the goat populations likely to be threatened in future conditions. The SPAG methodology is an efficient and flexible tool to characterize the evolutionary potential across a landscape and to transpose evolutionary information into conservation frameworks.

## KEYWORDS

adaptation, biodiversity, *Capra hircus*, climate change, conservation, evolutionary potential, genotype probabilities, goats, landscape genomics

Estelle Rochat and Oliver Selmoni contributed equally to this work.

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. *Diversity and Distributions* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Climate change is causing a shift away from the favourable conditions necessary for the survival of many animal and plant species (IPBES, 2019). In order to avoid extinction under these conditions, threatened species can either move to more favourable areas or adapt to their new environment (Waldvogel et al., 2020). Due to limitations in dispersal capacity, loss of favourable habitats and increased landscape fragmentation, the possibilities for dispersal to new areas are often limited (McGuire et al., 2016; Opdam & Wascher, 2004). Tolerance to stressful environments relies on phenotypic changes, which can be induced by genetic evolution (Fox et al., 2019; Merilä & Hendry, 2014).

The evolutionary potential for adaptation (hereafter referred to as evolutionary potential) of a given population can be defined as the presence of the genetic variants conferring resistance against an environmental constraint (Harrisson et al., 2014). Information on evolutionary potential can have major implications in the design of efficient conservation strategies (Hoffmann & Sgrò, 2011; Nicotra et al., 2015; Sgrò et al., 2011; Shafer et al., 2015). For instance, it could complement models on species distribution and provide a more pertinent view of biodiversity richness (Waldvogel et al., 2020). In addition, the current distribution of adaptive genetic variants could be compared with climate change projections to identify populations lacking the genetic characteristics necessary for future survival (notion defined as "genomic offset," or "genomic vulnerability"; Bay et al., 2018; Capblancq et al., 2020; Fitzpatrick & Keller, 2015). The characterization of the evolutionary potential of a population follows a two-step framework: first, the genetic variants involved in local adaptation are uncovered, and then, their distribution across an area of interest is predicted (Peterson et al., 2019).

The fields of population and landscape genomics have produced a wide spectrum of approaches to uncover genetic variants involved in local adaptation of natural populations (Luikart et al., 2003; Rellstab et al., 2015). In outlier tests (or genome scans), loci deviating from the neutral distribution of genetic variation (outliers) are considered as potentially adaptive (Luu et al., 2017). Such analyses are often complemented by genotype–environment association (GEA) methods, which investigate how the genetic variation of a population matches the environmental variability of the landscape where the population is living (Rellstab et al., 2015). For instance, redundancy analysis (RDA) can be used to assess how the distribution of a group of loci can be explained by a set of environmental descriptors (i.e. multi-loci adaptive genotypes; Legendre & Legendre, 2012). Other GEA methods investigate the association between environmental variables and single loci (i.e. single-locus adaptive genotypes). These methods are implemented in software like Bayenv (Günther & Coop, 2013), LFMM (Frichot et al., 2013), gINLAnd (Guillot et al., 2014) or Samþada (Stucki et al., 2017). The latter is designed for high-performance computation and is therefore attractive for whole genome-scale genotyping. In addition, Samþada has recently been complemented with an R-package facilitating the uptake of the

method (R. Samþada, Duruz et al., 2019, with dedicated guidelines to boost statistical power via sampling strategy as proposed in Selmoni et al., 2020).

Over the last decade, different methods have been proposed to transpose the results of GEA analyses into evolutionary potential spatial prediction frameworks (Waldvogel et al., 2020). One of the most straightforward approaches defines the evolutionary potential as the sum of the adaptive genotypes observed in an individual (polygenic scores; Gagnaire & Gaggiotti, 2016). For instance, Babin et al. (2017) applied this method to adaptive genotypes (uncovered using RDA and Bayenv) in American eel. Linear and generalized linear models were then employed to evaluate how the polygenic scores of sampled individuals matched the environmental conditions at the sampling locations. A similar approach was applied to loci adaptive for aridity (identified using LFMM and gINLAnd) in sugar beet (Manel et al., 2018). Importantly, the study is one of the few that cross-validated the predictions on evolutionary potential on a separated dataset. There are also more sophisticated methods to predict evolutionary potential at population-level and across multivariate environments. For instance, Razgour and colleagues (2019) recently assessed the frequency of adaptive genotypes in distinct populations using LFMM and RDA, and then employed the RDA ordination axes to project the expected genotype frequencies across the environmental space. Another framework based on genotype frequencies at population-level is the one proposed by Fitzpatrick and Keller (2015). This framework first employed outlier tests and Bayenv to identify adaptive loci in balsam poplar populations and then alternatively used a nonlinear extension of matrix regression method (generalized dissimilarity modelling) or a nonparametric machine learning algorithm (Gradient Forest) to predict evolutionary potential.

Despite the usefulness of the existing approaches for predicting the evolutionary potential, there are still important gaps that need to be filled. One of the main issues is that evolutionary potential is often described as a sum of the effects of single adaptive genotypes (Gagnaire & Gaggiotti, 2016). Such a view of additive genetic effects is well established, but neglects the fact that epistatic genetic structures can also underpin adaptation (Hansen, 2013). Under an epistatic view of adaptation, the selective advantage provided by an adaptive genotype depends on the presence or absence of other adaptive genotypes. For instance, two adaptive genotypes can be considered as equivalent when the presence of one genotype, regardless of the other, is sufficient to confer the selective advantage (Thompson & Jiggins, 2014). This situation can also concern loci implicated in adaptation to different environmental constraints, since the molecular basis of stress responses can converge towards common physiological mechanisms (Pandey et al., 2015). Understanding how the combined effects of different loci contribute to the evolutionary potential requires (a) a flexible framework to combine adaptive loci in different additive/epistatic relationships when computing the evolutionary potential and (b) a rigorous cross-validation approach to assess how pertinent evolutionary potential predictions are under the different additive/epistatic relationships.

We propose here a novel approach to predict the frequency of adaptive genotypes called SPAtial Areas of Genotypes Probabilities (SPAG). The SPAG approach is based on logistic genotype–environment associations (Joost et al., 2007) and takes as input the results of the GEA software Samβada (Stucki et al., 2017). By relying on conditional probability theory, the SPAG approach allows for the combination of adaptive genotypes in different types of intergenic relationships. The resulting SPAGs can be used to portray the evolutionary potential across an area of interest or to identify vulnerable populations that may be threatened by climate change. Importantly, the SPAG approach features a cross-validation step to assess the accuracy of the predictions on evolutionary potential. We introduce the theoretical bases of SPAGs and validate the approach with a simulated dataset. We then apply our approach to a case study on Moroccan goats to characterize their evolutionary potential to current climatic conditions and to predict genomic offset to climate change.

## 2 | MATERIAL AND METHODS

### 2.1 | SPAG approach

The SPAG approach is based on the spatial analysis mMethod (SAM; Joost et al., 2007) implemented in the GEA software Samβada to uncover candidate adaptive single-locus genotypes (Stucki et al., 2017). We first describe the SAM method and then show how the results of SAM can be used to predict spatial areas of genotype probabilities. Four different types of SPAGs are described (univariate SPAG, Intersection-SPAG, Union-SPAG and K-Percentage-SPAG), mirroring different possibilities of intergenic relationships between adaptive loci.

#### 2.1.1 | Logistic regressions (SAM)

The SAM method assumes a linear response of the genotype to the environmental variable and uses logistic regressions (Formula 1) to assess the probability of presence of a (single-locus) genotype  $G_1$  as a function of an environmental variable ( $x_1$ ),

$$p(G_1) = p(G_1 = 1 | x_1) = \frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}}, \quad (1)$$

where  $\beta_0$  and  $\beta_1$  are the parameters of the regression to be fitted. Independent univariate logistic regressions can be computed between each genotype and each environmental predictor, and significant associations can be identified using statistical tests. Joost et al. (2007) suggested the combined use of a likelihood ratio (G score) and a Wald test. The likelihood ratio (G) compares the likelihood of a model with the likelihood of a null model without the environmental variable of interest. The null hypothesis is that the model considered does not explain more variance than the null model, that

is the environmental variable considered does not help in predicting the probability of presence of the genotype. The Wald test is a common statistic used to estimate whether a parameter is equal to a given value. In our case, it is used to reject the null hypothesis that the  $\beta$  parameter associated with an environmental variable is equal to 0, which would also indicate that it has no effect on the probability of finding the genotype.

#### 2.1.2 | Univariate SPAG

Once a single-locus genotype involved into a significant association with an environmental variable has been identified, Formula (1) enables the estimation of the probability of presence of the genotype for any value of the environmental variable ( $x_1$ ). We consequently used it to estimate and delimit on a map the probability of presence of a genotype over the whole region of interest (Joost, 2006; Rochat et al., 2016). We named such a delimited surface *univariate Spatial Area of Genotype Probability* (SPAG).

#### 2.1.3 | Intersection-SPAG (I-SPAG)

The *Intersection* model (I-SPAG) is used to compute the probability that a set of single-locus genotypes are all simultaneously present. This corresponds to the situation where the evolutionary potential is defined by the additive effect of complementary loci (Gagnaire & Gaggiotti, 2016). Following the theory of conditional probability (Kolmogorov, 1956), the probability of simultaneous presence of  $n$  adaptive single-locus genotypes  $G_i$ ,  $i = 1:n$  can be computed using Formula 2:

$$p\left(\bigcap_{i=1}^n G_i\right) = p\left(\bigcap_{i=1}^{n-1} G_i\right) p\left(G_n | \bigcap_{i=1}^{n-1} G_i\right) \quad (2)$$

where  $p(G_n | \bigcap_{i=1}^{n-1} G_i)$  is a conditional probability that can be estimated using a logistic regression where  $\bigcap_{i=1}^{n-1} G_i$  is integrated as a covariate (Formula 3).

$$p\left(G_n | \bigcap_{i=1}^{n-1} G_i\right) = \frac{e^{\beta_0 + \beta_1 x_n + \beta_2 \bigcap_{i=1}^{n-1} G_i}}{1 + e^{\beta_0 + \beta_1 x_n + \beta_2 \bigcap_{i=1}^{n-1} G_i}} \quad (3)$$

However, as we would like to use this model to predict the probability of presence of the genotypes for any point of the region of interest, that is also where  $G_i$  values are unknown, we suggested to estimate  $\bigcap_{i=1}^{n-1} G_i$  by  $p(\bigcap_{i=1}^{n-1} G_i)$ .

Using the associative property of the intersection operator, the intersection of  $n$  genotypes can be computed by starting with the univariate model  $p(G_1)$ , which is used as a covariate to compute  $p(G_1 \cap G_2)$ , itself used to compute  $p(G_3 \cap (G_1 \cap G_2))$ , etc. Formula (3) can thus be implemented with a recursive model based on the univariate formula in which covariates are added (see File S1 for more details). We implemented this model as an R function available

following the link given in the section “code availability” at the end of the manuscript.

### 2.1.4 | Union-SPAG (U-SPAG)

The *union* model (U-SPAG) is used to compute the probability of finding at least one of the adaptive single-locus genotypes of interest. This view of evolutionary potential corresponds to the epistatic situation where the presence of an adaptive single-locus genotype is sufficient for securing the selective advantage, while additional adaptive genotypes do not confer additional advantages (Hansen, 2013). We implemented it with the inclusion-exclusion principle (e.g. for two genotypes:  $p(G1 \cup G2) = p(G1) + p(G2) - p(G1 \cap G2)$ ). We implemented the generalized formula for  $n$  adaptive genotypes (Formula 4), as an R function based on the intersection model previously described (see code availability and File S1).

$$p\left(\bigcup_{i=1}^n G_i\right) = \sum_{i=1}^n p(G_i) - \sum_{i < j} p(G_i \cap G_j) + \sum_{i < j < k} p(G_i \cap G_j \cap G_k) + \dots + (-1)^{n-1} p\left(\bigcap_{i=1}^n G_i\right) \quad (4)$$

### 2.1.5 | K-percentage SPAG (K-SPAG)

Finally, we developed a *K-percentage* model (K-SPAG) to estimate the probability that an individual carries K% of  $n$  adaptive single-locus genotypes. This view of evolutionary potential is at the crossroads between the additive situation, described in the I-SPAG, and the epistatic situation of U-SPAG. Here we assume that a higher number of adaptive single-locus genotypes increases the selective advantage (additive effect), but also that when a sufficient number of adaptive genotypes is present, there is no additional selective advantage associated with additional genotypes (epistatic effect). This probability can be computed by combining formulas from the union and intersection models (Formula 5, explained in more detail in File S1). Again, this formula was implemented as an R function (see code availability).

$$p(K\% G_{i=1,\dots,n}) = p\left(\bigcup_{i=1}^n \bigcap_{1 \leq i_1 < i_2 < \dots < i_{\lfloor K\%n \rfloor + 1}} (G_{i_1} \cap G_{i_2} \cap \dots \cap G_{i_k})\right) \quad (5)$$

Note that all multivariate models allow the integration of adaptive genotypes associated with various environmental variables since the environmental variable  $x_i$  used to compute  $p(G_i)$  can be different for each  $i$ .

## 2.2 | Simulation study

In order to test the SPAG approach, we first computed a simulated dataset using the individual-based population genetics model software CDPOP 1.3 (Landguth & Cushman, 2010; Landguth et al., 2020). We simulated individual genetic exchanges and natural selection across 300 non-overlapping generations among 200

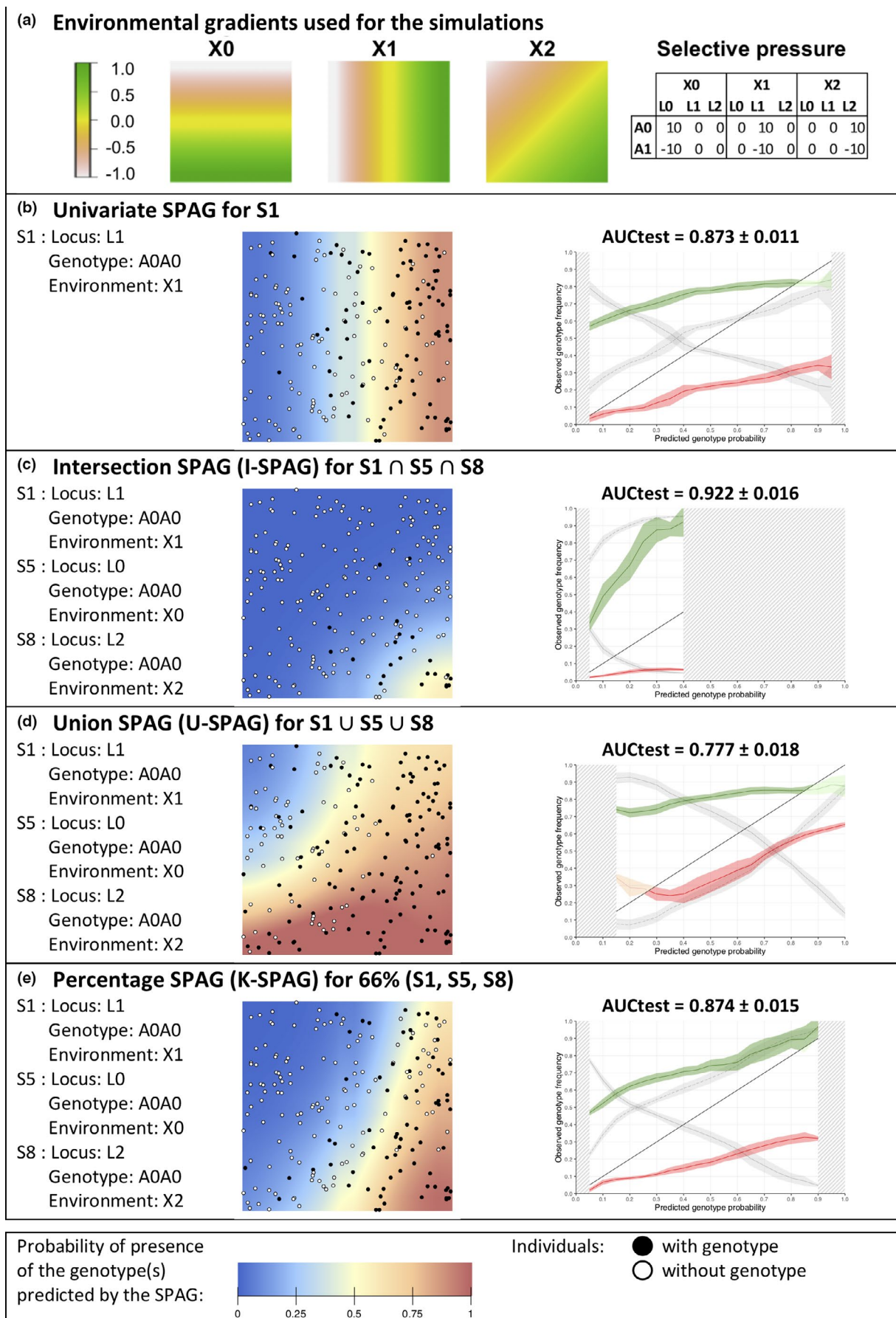
individuals randomly located in a  $500 \times 500$  gridded landscape. For the breeding parameters, we considered a sexual reproduction, with random mating, both male and female with replacement, no selfing, no philopatry, no multiple paternity, equal sex ratio and each mated pair producing three offspring. The movement of the individuals was linearly restricted as a function of the Euclidean distance, with a maximum dispersal corresponding to 25% of the entire landscape. We simulated 50 diallelic loci, with three loci under selection (L0, L1 and L2). The selection was implemented using three  $500 \times 500$  raster gradients, the first from north to south (X0), the second from east to west (X1) and the third from northwest to southeast (X2; see Figure 1a). We set the average effects  $b_{L0A0A0} = 10$  and  $b_{L0A1A1} = -10$  for the locus L0 with the environmental variable X0, which indicates that the genotype A0A0 from locus L0 will be favoured in the south (where  $X0 = 1$ ), whereas A1A1 will be favoured in the north (where  $X0 = -1$ ). We set similar effects for the locus L1 with the environmental variable X1 and L2 with X2. All other beta effects were set to 0, indicating no influence of the environmental variable to the distribution of genotypes. All genotypes were randomly initialized at the beginning of the simulations. The exact list of simulation parameters used is provided in File S2.

Univariate logistic regressions were then applied to the genetic data of individuals at the 300th generation in order to identify the most significant associations, which should highlight loci under selection. Univariate and multivariate SPAGs were then applied to estimate the probability to find these genotypes across the simulated landscape. The results were validated using a cross-validation procedure presented below.

## 2.3 | Moroccan goat

### 2.3.1 | Genetic data

The real case study focussed on goats (*Capra hircus*) from Morocco. The genetic dataset was produced in the context of the NEXTGEN project (Alberto et al., 2018). The NEXTGEN project produced whole genome sequences data for 161 Moroccan goats from 6 different local breeds. Since goat production system in Morocco is mainly free range, these goats are living from 8 to 12 months outdoors (Boujenane, 2005) and are confronted to contrasting environmental conditions, from the Sahara desert to the Atlas Mountains (see File S3). The goats were sampled in 161 farms chosen such to be representative of the range of environmental conditions observed in Morocco (Stucki, 2014). The sequencing method is described by Benjelloun et al. (2015) and allows to genotype 31.8 M of SNPs mapped to the goat's reference genome CHIR v1.0 (Dong et al., 2013). The genetic dataset was filtered such as to keep only autosomal, bi-allelic SNPs, with a maximum missingness per individuals and per site of 0.05 and a maximum major genotype frequency of 0.9. The final dataset contains 8,497,971 SNPs.



**FIGURE 1** SPAG—Simulated dataset. Univariate and multivariate spatial areas of genotypes probability (SPAG) for the simulated dataset. The identifiers of the presented models (S1, S5, S8) refer to Table 1. Please refer to Box 1 to interpret the validation graphs shown on the right of each map



### 2.3.2 | Environmental data

The climatic conditions of the sampling locations were characterized using the 19 bioclimatic variables (File S4) from the WorldClim database (<https://www.worldclim.org/>), representative of the period 1960–1990 (Hijmans et al., 2005). Each variable was retrieved as a raster layer with a spatial resolution of 30 arc-seconds (approx. 1 km<sup>2</sup>), and values were extracted for all sampling locations using the *extract* function from the R-package *raster* (Hijmans & van Etten, 2012). In order to get a comparable scale of values for all bioclimatic variables, we applied a standardization by subtracting the mean and dividing by the standard deviation. Some of the bioclimatic variables are highly correlated. However, we choose to keep all of them to be able to identify a posteriori which variable had the strongest effect. Since no models computed involved more than one environmental variable simultaneously, this collinearity will not impact the results.

### 2.3.3 | Population structure

We ran a preliminary analysis of population structure to evaluate the possible confounding role of neutral genetic variation on the detection of adaptive genotypes (Li et al., 2012). The genetic population structure was estimated with a principal component analysis (Price et al., 2006; Reich et al., 2008) computed with the function *snpgdsPCA* from the *SNPRelate* R-package (Zheng et al., 2012). In order to avoid a strong influence of SNP clusters on this analysis, we used here a pruned set of SNPs that are in approximate linkage equilibrium with each other. The pruning was performed with the function *snpgdsLDpruning* from the *SNPRelate* package, with a threshold  $D' = 0.2$ . The resulting datasets contain 59,224 SNPs.

The cumulated variance explained by the 10 first PCA components on the SNP markers represents only 8.1% of the total variance, and the increase in variance explained is almost proportional to the number of components, which highlights that there is no clear substructure (File S5). We thus performed the GEA analysis without any covariates for population structure.

### 2.3.4 | Logistic regressions and SPAGs

Logistic models were computed for all the combinations of single-locus genotypes with the 19 bioclimatic variables. The statistical significance of the model was assessed using Wald test and log-likelihood ratio (G), both corrected for the false discovery rate due to multiple comparisons using the procedure proposed by Benjamini and Hochberg (1995), under an expected false discovery rate (FDR) of 0.05 (i.e. 5% of the results expected to be false positives). An association was considered as significant if both the models without covariates and with population covariates were significant.

In order to identify potential functions of the SNPs involved into the significant associations, we used the NCBI Genome Data

Viewer ([https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF\\_000317765.1](https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_000317765.1)) to search for the presence of annotated genes in the genomic region of 10kbp surrounding the SNPs of interest. All analyses were computed using a combination of the *SamBada* software (Stucki et al., 2017) and a custom R-script based on the *glm* function.

Finally, the significant logistic regressions uncovered by the genotype–environment association analysis were used to compute the SPAGs. We computed (a) the univariate SPAG for every significant single-locus genotype; (b) the bivariate and trivariate I-SPAGs and U-SPAGs for every combination of significant single-locus genotypes; and (c) the trivariate K-SPAGs (percentage = 0.66) for every combination of significant single-locus genotypes. Note that all SPAGs were computed using the absolute value of the coefficient of regression ( $\beta_1$ ) describing the relationship between genotypes and environmental variable.

### 2.4 | Validation procedure

SPAGs were validated with a cross-validation procedure, using 25% of the individuals to compute the SPAG (i.e. 50 individuals for the simulated datasets, 41 individuals for the Moroccan goats) and the remaining 75% to test it. Training individuals were selected such to represent the entire range of values of the environmental variable under study (see File S1 for the exact procedure). The model was validated using the area under the receiver operating curve computed with the testing dataset (AUCtest, Fielding & Bell, 1997) and a custom validation graph presented in Box 1. The cross-validation procedure was repeated 10 times. In order to avoid bias due to rare genotypes occurrences, we applied this cross-validation procedure only when at least 10 individuals carried the genotype of interest.

### 2.5 | Genomic offset under climate change

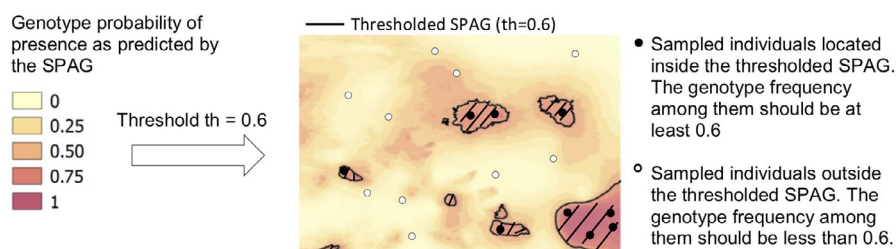
We used the Moroccan goat dataset to present an application of the SPAG for predicting the distribution of adaptive genotype(s) under climate change. In order to predict the genotype frequency optimal for future conditions, we retrieved WorldClim data for the year 2070, corresponding to a strong climate change scenario from the Max Planck Institute Earth System Model (MPI-ESM-LR; Giorgetta et al., 2013) with a Representative Concentration Pathway equals to 8.5 (RCP 8.5). We then assume that the optimal genotype frequency for future conditions should be close to the genotype frequency currently observed in areas with climatic conditions resembling the future ones. We thus applied the current parameters of the logistic regressions on the future environmental variables in order to derive the future SPAGs for the genotypes of interest. We then studied the genomic offset between the current and future SPAGs to identify vulnerable populations for which specific genotype frequencies should be higher for individuals to adapt to the future conditions.

### BOX 1 Validation graph

SPAGs indicate the probability of finding one or more genotypes of interest in a territory (panel below). For a given probability threshold value (e.g.  $th = 0.6$ ), we can thus use the SPAG to delimit the area where the probability of finding the genotype(s) of interest is predicted to be greater or equal to this threshold (e.g.  $p \geq 0.6$ ). If the SPAG is valid, the frequency of the genotype(s) observed among the testing individuals located within the thresholded SPAG should effectively be greater or equal to the threshold value, whereas it should be less outside.

The goal of the validation graph presented in this box is to evaluate the predictive power of SPAGs for different levels of genotype probability. We thus calculated the observed genotype frequencies inside and outside the thresholded SPAGs for each threshold value between 0 and 1 (with a step of 0.1), and we presented the results on a graph (panel below). The green line indicates the genotype frequency observed inside the thresholded SPAG, whereas the red line shows the genotype frequency observed outside it. A black line indicates the limit case where the observed genotype frequency is equal to the threshold value. For a given level of genotype probability, the SPAG is thus validated if the green line remains above the black line and the red line remains below it. The green and red areas around the lines indicate the 95% confidence intervals for each line, computed on the basis of the 10 cross-validation runs. We also presented on the graph the percentage of individuals located within the thresholded SPAG (grey line) and outside it (dotted grey line). The hatched grey areas indicate ranges of testing values where there was less than 5 individuals remaining inside or outside the thresholded SPAG, which was therefore considered not to be usable for the validation.

For a threshold value  $th = 0.4$ , the SPAG is validated since 50% of the testing individuals located within the area  $SPAG \geq 0.4$  carry the genotype of interest, whereas only 26% carry it outside ( $SPAG < 0.4$ ). Inversely, the model is not validated for  $th = 0.6$ , since only 54% of individuals carry the genotype of interest within the area where the SPAG predicted a probability of at least 0.6 ( $SPAG \geq 0.6$ ). The model is also not valid for a value  $th = 0.2$  since 23% of individuals carry the genotype of interest in the area  $SPAG < 0.2$ .



## 3 | RESULTS

### 3.1 | Simulated dataset—logistic regressions

The 10 most significant genotype–environment associations obtained with the simulated dataset ranked on the basis of the likelihood ratio (G) are presented in Table 1. We observed that the three loci simulated as under selection (L0, L1, L2) were coherently identified as the most significantly associated with the environmental variables under study.

### 3.2 | Simulated dataset—SPAGs

We focussed on the three single-locus genotypes under selection (corresponding to models S1, S5 and S8), and we evaluated how using different types of SPAGs modulated the predictive power of the approach. Figure 1 displays an example for every type of SPAG.

Univariate SPAGs resulted in an average AUC of  $0.84 \pm 0.05$  (Table S1). Figure 1b presents the univariate SPAG for the model S1

(Locus L1, genotype A0A0 associated with the environmental variable X1). Since this locus was simulated as under selection with the east–west gradient X1, the resulting SPAG coherently shows a similar gradient. The results are validated by the validation graph for almost the entire range of probability values, except close to 0.8, where the SPAG slightly overestimated the probability of presence of the genotype.

Intersection-SPAGs (I-SPAGs) generally showed higher AUCs, compared to univariate SPAGs (bivariate I-SPAGs:  $AUC = 0.91 \pm 0.03$ ; trivariate I-SPAG:  $AUC = 0.92 \pm 0.02$ ; Table S1). Figure 1c shows the I-SPAG for the three single-locus genotypes under selection. It indicates a very low probability of finding them all simultaneously, except in the southeast of the simulated area (where all environmental gradients had values close to 1).

Union SPAGs (U-SPAGs) resulted in AUCs systematically lower than in univariate and I-SPAGs (bivariate U-SPAGs:  $AUC = 0.80 \pm 0.03$ ; trivariate U-SPAG:  $AUC = 0.78 \pm 0.02$ ; Table S1). Figure 1d presents a U-SPAG of three single-locus genotypes under selection. It shows that the probability of finding at least one of them is higher than 0.5 in most of the area. The validation graph indicates that the model failed to estimate the

**TABLE 1** Most significant models—Simulated datasets. 10 most significant models obtained for the genotype–environment association (GEA) analysis of the simulated datasets. The simulated dataset sets an association between locus L0 and environmental variable X0, locus L1 and environmental variable X1 and locus L2 and environmental variable X2. GEA models are ranked based on the likelihood ratio. Env, environmental predictor; Geno, genotype; *p*G, *p*-value associated with the G score; *p*W, *p*-value associated with the Wald score;  $\beta$ 0 and  $\beta$ 1 are the parameters of the logistic regression

ID	Marker	Locus	Geno.	Env.	G score	<i>p</i> G	<i>p</i> W	$\beta$ 0	$\beta$ 1
S1	L1A0A0	L1	A0A0	X1	92.30	7.44E-22	9.41E-14	−0.46	2.65
S2	L1A1A1	L1	A1A1	X1	90.49	1.86E-21	1.51E-12	−1.24	−2.94
S3	L2A1A1	L2	A1A1	X2	68.63	1.19E-16	9.02E-11	−1.59	−4.37
S4	L0A1A1	L0	A1A1	X0	65.69	5.29E-16	4.04E-11	−0.66	−2.35
S5	L0A0A0	L0	A0A0	X0	64.95	7.69E-16	6.92E-11	−1.58	2.61
S6	L2A0A0	L2	A0A0	X1	63.86	1.34E-15	1.37E-11	−0.55	2.04
S7	L2A1A1	L2	A1A1	X1	61.04	5.59E-15	4.12E-10	−1.56	−2.40
S8	L2A0A0	L2	A0A0	X2	60.94	5.88E-15	2.52E-10	−0.54	3.45
S9	L0A0A0	L0	A0A0	X2	43.43	4.39E-11	2.81E-08	−1.41	3.11
S10	L42A0A0	L42	A0A0	X0	42.44	7.30E-11	5.93E-09	0.06	1.67

probability of finding the genotypes for threshold values below 0.2 and above 0.8.

The AUC of the K-SPAG was higher than the univariate and U-SPAGs ( $AUC = 0.87 \pm 0.02$ ). Figure 1e depicts the probability of finding at least two of the three single-locus genotypes under selection, which is high in the southeast of the map. The validation graph indicates that the predictive power of the K-SPAG was usable for any level of genotype probability.

### 3.3 | Moroccan goats—logistic regressions

More than 483 million logistic association models were computed. After correction for false discovery rate with a significant

threshold of 5%, no model was significant according to the Wald score, but seven models were significant according to the G score (Table S2a). Among them, three models were strongly associated with the precipitation seasonality (bio15, Figure 3a), which is the

coefficient of variation ( $CV = \frac{SD}{mean}$ ) of the variation of monthly

precipitation over the year. Following this initial result, we investigated in more details the adaptation to this bioclimatic variable. When considering only the associations involving bio15 (25,447,348 models), 78 models were significant after FDR correction of G score, with a significant threshold of 5% (Table S2b). The SNPs involved in these models are located on seven different genomic regions (Table 2), corresponding to four annotated genes

**TABLE 2** Significant models for precipitation seasonality—Moroccan datasets. Significant models obtained for the analysis of Moroccan datasets with precipitation seasonality (bio15) after FDR correction. Chr, chromosome; End, end in base pairs of the region; G, G score (Log-Likelihood ratio); Genes, annotated genes on the genomic region; Geno, corresponding SNP Genotype; GF, corresponding genotype frequency; ID, model identifier; Peak SNP, SNP of the most significant model in that region; qG, corresponding *p*-value corrected for FDR; Start, start in base pairs of the region identified as under selection;  $\beta$ 0 and  $\beta$ 1, parameters of the logistic regression

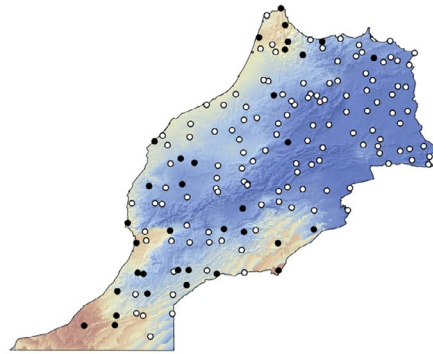
ID	Chr	Start (BP)	End (BP)	Peak (BP)	Geno	GF	G	qG	$\beta$ 0	$\beta$ 1	Genes
M1	6	12'174'332	12'298'321	12'276'168	AA	21.74	38.74	0.004	−1.80	1.51	-
M2	13	43'436'394	43'438'732	43'436'394	GG	10.56	29.02	0.042	−3.14	1.80	-
M3	24	19'436'980	19'436'980	19'436'980	CC	76.40	34.75	0.008	1.55	1.29	-
M4	24	25'852'900	25'860'754	25'860'754	AG	38.51	34.75	0.008	−0.29	−1.07	DSG4
M5	24	28'799'029	28'833'762	28'833'253	TT	12.42	27.87	0.046	−2.72	1.58	CDH2
M6	24	30'566'869	30'584'692	30'566'869	TT	2.48	27.99	0.046	−25.69	−15.44	KCTD1
M7	27	25'930'079	25'933'133	25'930'079	GG	78.88	32.88	0.012	1.76	−1.35	WRN

**FIGURE 2** SPAG—Moroccan dataset. Univariate and Multivariate spatial areas of genotypes probability (SPAG) for the Moroccan dataset. The identifiers of the presented models (M1, M3, M4) refers to Table 2. The maps show the average genotype(s) frequency(ies) based on the 10 runs computed with different random selection of training sets containing 25% of the total number of individuals. AUCtest indicates the mean value of the AUC computed with the testing dataset over the 10 runs. Please refer to Box 1 to interpret the validation graphs shown on the right of each map

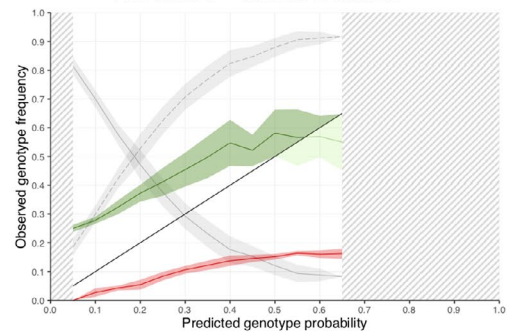


**(a) Univariate SPAG for M1**

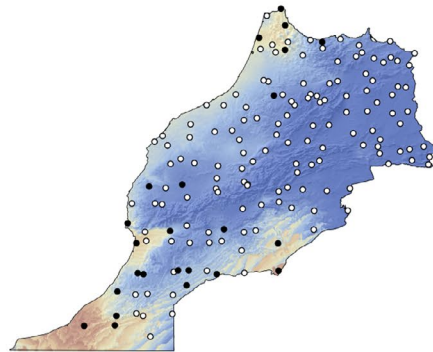
M1 : Chr: 6  
Pos : 12276168  
Geno: AA



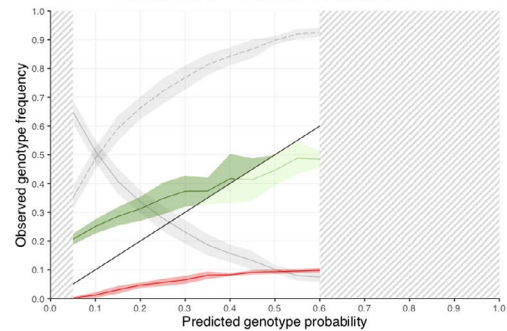
AUCtest =  $0.815 \pm 0.023$

**(b) Intersection SPAG (I-SPAG) for  $M1 \cap M4$** 

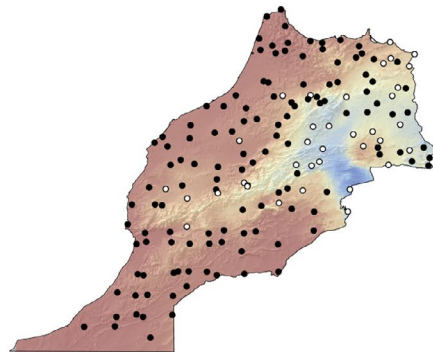
M1 : Chr: 6  
Pos : 12276168  
Geno: AA  
M4 : Chr: 24  
Pos : 25860754  
Geno: AG



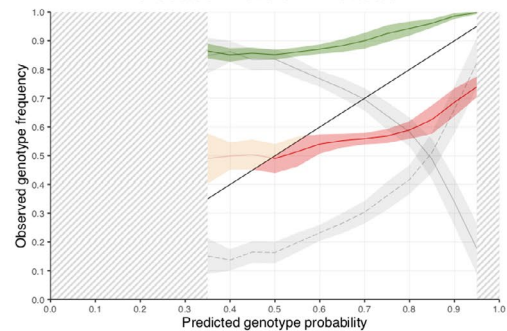
AUCtest =  $0.847 \pm 0.022$

**(c) Union SPAG (U-SPAG) for  $M1 \cup M3$** 

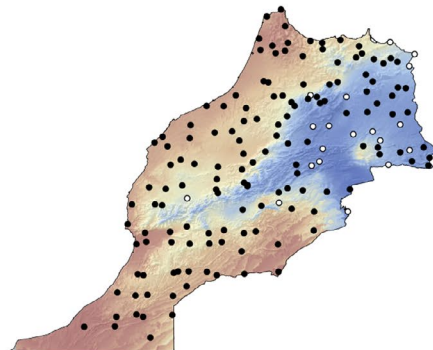
M1 : Chr: 6  
Pos : 12276168  
Geno: AA  
M3 : Chr: 24  
Pos : 19436980  
Geno: CC



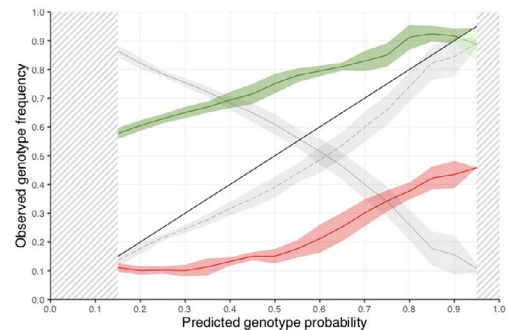
AUCtest =  $0.837 \pm 0.009$

**(d) Percentage SPAG (K-SPAG) for 66% ( $M1, M3, M4$ )**

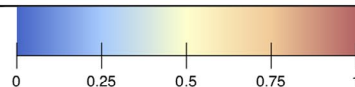
M1 : Chr: 6  
Pos : 12276168  
Geno: AA  
M3 : Chr: 24  
Pos : 19436980  
Geno: CC  
M4 : Chr: 24  
Pos : 25860754  
Geno: AG



AUCtest =  $0.863 \pm 0.009$



Probability of presence  
of the genotype(s)  
predicted by the SPAG:



Individuals:

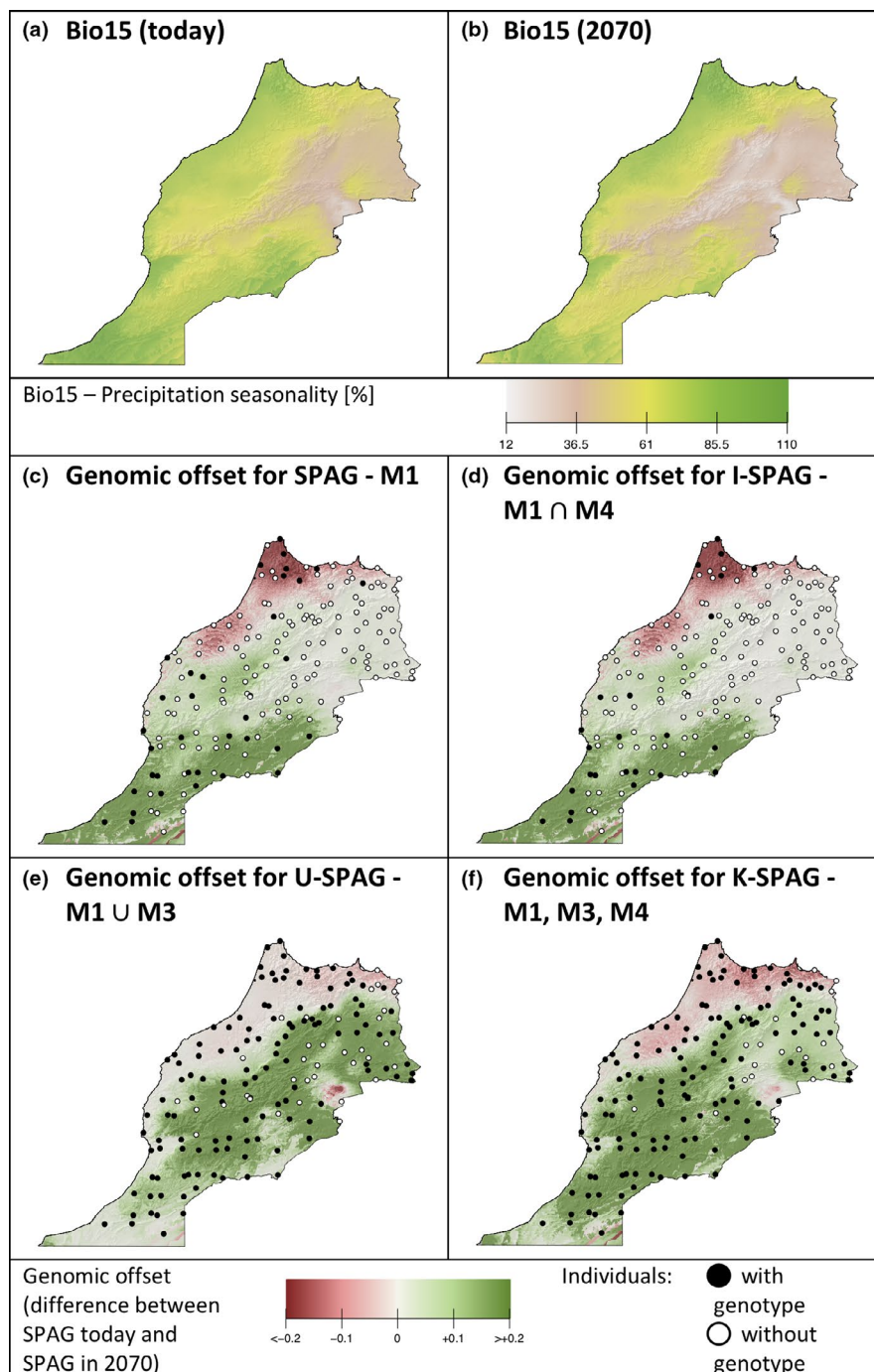
● with genotype  
○ without genotype

(DSG4, CDH2, KCTD1 and WRN) on the reference genome CHIR 1.0.

### 3.4 | Moroccan goats—SPAGs

The seven significant single-locus genotypes uncovered in the GEA were employed to compute univariate SPAGs, I-SPAGs, U-SPAGs (both bi- and trivariate) and K-SPAGs (trivariate only). We assessed the predictive power of the approach under different types of SPAGs using AUC, and we show some examples of SPAGs in Figure 2.

Univariate SPAGs resulted in an average of AUC of  $0.79 \pm 0.03$  (Table S3; Figure 2a shows the univariate SPAG for the genotype of model M1 presented in Table 2 (see File S6 for the other univariate SPAGs). The predicted probability of presence of the genotype is the highest in the extreme southwest of the country. In this region, the precipitation seasonality is the greatest ( $>100\%$ , Figure 3a) and all goats carry the genotype of interest. In northern and western coastal areas, the predicted probability of finding the genotype is close to 0.5, since the precipitation seasonality is relatively high ( $>70\%$ ). Finally, in the eastern regions of the country, the probability of finding the genotype of model M1 is much lower ( $<0.2$  in most areas). In these regions, variations of precipitation



**FIGURE 3** Moroccan goats—Genomic offset in 2070. Maps A and B display the current and forecasted (2070) precipitation seasonality (Bio15), respectively. Maps C–F show the genomic offset for different types of spatial areas of genotype probability (SPAGs), that is the difference between the probability of adaptation today and the probability forecasted on future Bio15 values. Red areas indicate regions where a higher probability of presence of adaptive genotype is expected in 2070, compared with today. Green areas indicate regions where a lower probability of presence of adaptive genotype is expected in 2070

are less important (<40%) and most of the goats sampled do not carry the genotype.

Intersection-SPAGs (I-SPAGs) displayed an average AUC of  $0.81 \pm 0.07$  in bivariate SPAGs and of  $0.83 \pm 0.05$  in trivariate SPAGs. Figure 3b shows the I-SPAG for the genotypes involved in GEA models M1 and M4 (Table 2). This I-SPAG indicates that the simultaneous presence of the two single-locus genotypes is very unlikely ( $p < 0.2$ ) for most of the territory, with the exception for the southwest of the country.

Union SPAGs (U-SPAGs) generally resulted in lower AUC, compared to I-SPAGs (bivariate U-SPAGs: AUC =  $0.81 \pm 0.03$ ; trivariate U-SPAGs:  $0.82 \pm 0.03$ ). Indeed, we found only one combination of single-locus genotypes with higher AUC under U-SPAG ( $0.84 \pm 0.01$ ) than under I-SPAG ( $0.81 \pm 0.02$ ). This U-SPAG employed genotypes from M1 and M3 and is shown in Figure 3c. The probability of finding at least one of these two genotypes is close to 1 for most of the country except in the north-eastern region.

Percentage SPAGs (K-SPAGs) showed AUC values generally lower than I-SPAGs (AUC =  $0.82 \pm 0.05$ ). One exception was the K-SPAG shown in Figure 3d, employing at least two of the genotypes from M1, M3 and M4. For these three loci, the K-SPAG showed an higher AUC ( $0.86 \pm 0.01$ ) than the I-SPAG ( $0.83 \pm 0.02$ ) and the U-SPAG ( $0.84 \pm 0.03$ ). Unlike the I-SPAG and U-SPAG shown in Figure 3b,c, the validation plot shows that this K-SPAG is validated for any level of probability of presence of the adaptive genotype.

### 3.5 | Moroccan goats—genomic offset

Figure 3 shows the genomic offset, which represents the differences between the current SPAGs (presented in Figure 2) and the SPAGs corresponding to their projections for 2070. In Morocco, precipitation seasonality (bio15) is predicted to increase in the north of the country, with a maximum increase of 5%–10% in the extreme north-western region, and to decrease (from –10% to –20%) in other areas, especially in the Atlas Mountains (centre of the country, in the eastern region) and near the Sahara (South region; Figure 3a,b).

The genomic offsets of the univariate SPAG for model M1 (Figure 3c) and of the I-SPAG for models M1–M4 (Figure 3d) display comparable spatial patterns. In both SPAGs, the northern regions of Morocco show the highest values of genetic offset (up to +20%), indicating that in these regions, the probability of finding the adaptive genotypes would have to be much higher for individuals to be adapted to 2070 conditions. Today, such genotypes are observed only in the extreme northern region, while they are absent in the northwest and northeast.

The genomic offsets of the U-SPAG for models M1–M3 (Figure 3e) and of the K-SPAG for models M1–M3–M4 (Figure 3f) show a similar spatial pattern to the one described above, even though with a different level of magnitude. Indeed, in the northern regions of Morocco the probability of presence of adaptive genotypes only have to be slightly higher (up to +10%) for individuals to be adapted to 2070 conditions. However, the multi-locus genotypes described

by the U-SPAG and the K-SPAG were already found in most of the sampled goats, with the exception of the north-eastern regions.

## 4 | DISCUSSION

### 4.1 | Mapping genotype probabilities

The different SPAGs computed in this work generally showed an AUC above 0.8 on both simulated and real data, stressing the predictive power of the approach. While previous works had already described methods to transpose GEA results into predictions of evolutionary potential (Babin et al., 2017; Fitzpatrick & Keller, 2015; Gagnaire & Gaggiotti, 2016; Razgour et al., 2019), the cross-validation of such predictions is rare (Manel et al., 2018). In this regard, it is noteworthy to mention that the SPAG approach showed a generally high predictive power despite using a limited number of training samples (i.e. 50 simulated individuals, 41 Moroccan goats).

As we compared different types of SPAGs, we noticed that the I-SPAGs generally showed a higher predictive power. Indeed, for a same combination of single-locus genotypes, the I-SPAGs resulted in higher AUCs than the corresponding U-SPAGs and univariate SPAGs. We expected to observe this in the simulated dataset, because the three adaptive genotypes were set as independent from each other (i.e. without any epistatic relationship). The same was observed in the Moroccan goat case study. Two mutually non-exclusive hypotheses might explain this observation: (a) additive genetic effects are predominant in local adaptation; (b) GEA methods are more sensitive to adaptive loci in additive relationships between each other.

Of note, the Moroccan goat study also showed one case where two single-locus genotypes (M1 and M3) showed a higher predictive power when considered as alternative to each other (U-SPAG), instead of complementary (I-SPAG). A redundancy in the functional annotation of the genes controlled by these loci might corroborate the view of an epistatic relationship (Phillips, 2008). Unfortunately, such interpretation is hampered by the fact that both loci are located in intergenic regions of the genome and therefore possibly act as regulatory regions on distant genes (Brodie et al., 2016).

Nevertheless, previous research expects that the evolutionary potential is shaped by the co-existence of adaptive loci underpinning both additive and epistatic relationships (Hansen, 2013; Thompson & Jiggins, 2014). The existing methods to predict evolutionary potential overlook the epistatic component of adaptation either because (a) they explicitly focus on additive effects (e.g. “polygenic scores”; Gagnaire & Gaggiotti, 2016); (b) they work at population-level (Fitzpatrick & Keller, 2015). Indeed, in population-level approaches the presences/absences of single-locus genotypes are averaged as genotype frequencies by population. Consequently, it is impossible to assess whether two loci co-occur (or not) in the same individual and to establish whether they are complementary (or alternative) to each other.

The K-SPAG overcomes these issues, by providing a flexible approach to use information at individual-level to combine adaptive loci



underlying different types of intergenic relationships. For instance, in the Moroccan case study we constructed a composite genotype combining alternative loci (M1 and M3) with additive loci (M1 and M4), and the resulting K-SPAG showed higher predictive power than the corresponding I-SPAGs and U-SPAGs alone.

## 4.2 | From SPAG to conservation: The Moroccan goat example

Information on the evolutionary potential can have major implications in predicting species response to climate change (Razgour et al., 2019). One of the main application in conservation is the prioritization of areas based on the information on evolutionary processes (Orr & Unckless, 2008). The use of the SPAGs to characterize the genomic offset under climate change conditions can help identify (a) well-adapted populations, where individuals currently show adaptive genotypes expected to be optimal under future conditions, (b) threatened populations where adaptive genetic variants are currently rare or missing and will be needed to ensure future survival.

The Moroccan goat case study offers a concrete example of the usefulness of SPAGs in determining such conservation priorities. Livestock populations in developing country are often bred in outdoor pastoralist systems, therefore exposed to climatic constraints (e.g. heat stress, limited water availability, presence of parasites; FAO, 2015). Increase in precipitation seasonality is a threat for such populations, as it implies a more heterogeneous distribution of rainfall across the year, exacerbating the risk of drought and floods (Nardone et al., 2010). Climate forecasts expect an increase in precipitation seasonality in the northern regions of Morocco (Tramblay et al., 2012). We estimated the genomic offset using different types of SPAGs, and all the resulting predictions agree that in the northern regions of Morocco an increase in the evolutionary potential for adaptation to precipitation seasonality will be necessary by 2070.

Importantly, such increase in the evolutionary potential depends on natural gene flow, which is not accounted for in the genomic offset calculations. Here, SPAGs can be used to evaluate whether adaptive genotypes are already present in the populations that will need them in the future. For instance, we can observe that if we consider univariate SPAGs or SPAGs accounting for additive genetic effects (I-SPAGs), only the extreme north of Morocco is expected to already host such genotypes, which are rarer in the north-western and absent in the north-eastern regions. However, if we assume that epistatic genetic effects contribute to the evolutionary potential (U-SPAGs, K-SPAGs), we can see that almost the entire northern region already hosts adapted individuals, with the only exception of the extreme northeast.

This mismatch in the results underlines the importance of assessing how complex relationships between adaptive loci contribute to the evolutionary potential (Hansen, 2013). In this regard, it is recommendable to complement the use of SPAGs in conservation plans

with other information on phenotypic or physiological characteristics associated with the different adaptive genotypes (Waldvogel et al., 2020).

## 4.3 | Limitations and perspectives

The SPAG approach presented appears to be powerful for mapping the probability of finding locally adapted genetic variants in a landscape. However, the adaptation process is complex and can involve large numbers of loci with small individual effects (Pritchard & Rienzo, 2010), for which the detection power of the single-locus GEA methods may be reduced (Harrisson et al., 2014; Villemereuil et al., 2014). In this case, it may be advisable to use multi-loci GEA methods to detect adaptive genotypes (Forester et al., 2017; Legendre & Legendre, 2012) and to use multi-loci prediction frameworks to predict evolutionary potential (Fitzpatrick & Keller, 2015).

In addition, since the genomic offset calculations may be highly dependent on the climate change scenario considered, computations should be performed with various scenarios and less weight should be given to the conclusions not consistent within scenarios (Reside et al., 2018). It is also possible that the forecasted environmental conditions exceed the climatic ranges observed today, leading to extrapolation (and the related uncertainty) in the calculation of future evolutionary potential (Elith et al., 2010). Such uncertainty should always be considered in the interpretation of the genomic offset calculated using the SPAGs.

Furthermore, in order to assess the real vulnerability of populations, an analysis of connectivity should be carried out to highlight the potential of natural gene flow to increase the probability of finding favourable genotypes in threatened populations. Such information on connectivity could be used to develop a statistical framework assessing the likelihood of the shift from a current SPAG to a future SPAG.

The SPAG approach can be extended to the output of other types of GEA methods, assuming that such output is convertible into a binary measure of genotype presence/absence. This might be relatively straightforward for GEA methods that are individual-based and focus on single-locus genotypes. With LFMM (Frichot et al., 2013) for instance, the adaptive genotype (or allele) can be easily identified by comparing the variation of the genetic variant of interest across the environmental gradient driving adaptation. For population-level GEA methods (e.g. RDA; Legendre & Legendre, 2012), the use of SPAGs is more complex and less pertinent. In fact, with this type of GEA methods the frequencies of single-locus genotypes are generally averaged by population and grouped via ordination techniques, and it is therefore difficult to assess in which individuals the adaptive genotype is present (or absent).

Finally, GEA and SPAGs could also be used to predict the presence of genotype(s) associated with other pressures showing a spatial distribution, such as the presence of a parasite (Vajana et al., 2018) or a predator (Cousyn et al., 2001) or the urbanization level (Harris & Munshi-South, 2016). Including such selective

pressure will portray a more realistic view of how different species are shaping their evolutionary potential in response to climate change.

## ACKNOWLEDGMENTS

We thank Paolo Ajmone Marsan and Elia Vajana for the fruitful discussions we had about the subject of this research. The authors declare that there is no conflict of interest regarding the publication of this paper.

## Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ddi.13256>.

## DATA AVAILABILITY STATEMENT

The Moroccan goat dataset is available at <https://projects.ensembl.org/nextgen/> (population MOCH). The R functions developed to compute the univariate and multivariate SPAGs are available at: <https://github.com/estellerochat/SPAG>.

## ORCID

Estelle Rochat  <https://orcid.org/0000-0002-7978-5239>

Oliver Selmoni  <https://orcid.org/0000-0003-0904-5486>

Stéphane Joost  <https://orcid.org/0000-0002-1184-7501>

## REFERENCES

- Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., Barbato, M., Zamani, W., Alberti, A., Engelen, S., Stella, A., Joost, S., Ajmone-Marsan, P., Negrini, R., Orlando, L., ... Pompanon, F. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*, 9(1), <http://dx.doi.org/10.1038/s41467-018-03206-y>.
- Babin, C., Gagnaire, P.-A., Pavey, S. A., & Bernatchez, L. (2017). RAD-Seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American Eel (*Anguilla rostrata*). *Genome Biology and Evolution*, 9, 2974–2986.
- Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Rugg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, 359, 83–86.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.
- Benjelloun, B., Alberto, F. J., Streeter, I., Boyer, F. A., Coissac, E., Stucki, S., BenBati, M., Ibbelbachyr, M., Chentouf, M., Bechchari, A., Leempoel, K., Alberti, A., Engelen, S., Chikhi, A., Clarke, L., Flicek, P., Joost, S. A., Taberlet, P., & Pompanon, F. A. (2015). Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. *Frontiers in Genetics*, 6. <https://doi.org/10.3389/fgene.2015.00107>
- Boujenane, I. (2005). *Small Ruminant Breeds of Morocco*.
- Brodie, A., Azaria, J. R., & Ofran, Y. (2016). How far from the SNP may the causative genes be? *Nucleic Acids Research*, 44, 6046–6054. <https://doi.org/10.1093/nar/gkw500>
- 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, 245–269.
- Cousyn, C., Meester, L. D., Colbourne, J. K., Brendonck, L., Verschuren, D., & Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 6256–6260.
- de Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, 23, 2006–2019. <https://doi.org/10.1111/mec.12705>
- Dong Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., ... Wang, W. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology*, 31(2), 135–141. <http://dx.doi.org/10.1038/nbt.2478>.
- Duruz S., Sevane, N., Selmoni, O., Vajana, E., Leempoel, K., Stucki, S., Orozco-terWengel, P., Rochat, E., Dunner, S., Bruford, M. W., Joost, S., (2019). Rapid identification and interpretation of gene-environment associations using the new R.SamBada landscape genomics pipeline. *Molecular Ecology Resources*, 19(5), 1355–1365. <http://dx.doi.org/10.1111/1755-0998.13044>.
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330–342.
- FAO. (2015). The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture.
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24, 38–49.
- 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.
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2017). *Using genotype-environment associations to identify multilocus local adaptation*.
- Fox, R. J., Donelson, J. M., Schunter, C., Ravasi, T., & Gaitán-Espitia, J. D. (2019). Beyond buying time: The role of plasticity in phenotypic adaptation to rapid environmental change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374, 20180174.
- Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, 30, 1687–1699.
- Gagnaire, P.-A., & Gaggiotti, O. E. (2016). Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*, 62, 603–616.
- Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H., Ilyina, T., Kinne, S., Kornbluh, L., Matei, D., Mauritsen, T., ... Stevens, B. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, 5(3), 572–597. <http://dx.doi.org/10.1002/jame.20038>.
- Guillot, G., Vitalis, R., le Rouzic, A., & Gautier, M. (2014). Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. *Spatial Statistics*, 8, 145–155.
- Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, 195, 205–220.
- Hansen, T. F. (2013). Why epistasis is important for selection and adaptation. *Evolution*, 67, 3501–3511.
- Harris, S. E., & Munshi-South, J. (2016). Scans for positive selection reveal candidate genes and local adaptation of *Peromyscus leucopus* populations to urbanization. *bioRxiv*, 038141.



- Harrison, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, 7, 1008–1025.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.
- Hijmans, R. J., & van Etten, J. (2012). raster: Geographic analysis and modeling with raster data. R package version 2.0-12. Retrieved from <http://CRAN.R-project.org/package=raster>
- Hoffmann, A. A., & Sgrò, C. M. (2011). Climate change and evolutionary adaptation. *Nature*, 470, 479–485.
- IPBES. (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. E. S. Brondizio, J. Settele, S. Díaz, & H. T. Ngo (Eds). Bonn, Germany: IPBES secretariat.
- Joost, S. (2006). The geographical dimension of genetic diversity: a GIScience contribution for the conservation of animal genetic resources. *EPFL*. Retrieved from <https://infoscience.epfl.ch/record/64345?ln=en>
- 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(18), 3955–3969. <http://dx.doi.org/10.1111/j.1365-294X.2007.03442.x>
- Kolmogorov, A. (1956). *Foundations of the theory of probability*. Chelsea Pub Co.
- Landguth, E. L., & Cushman, S. A. (2010). cdpop: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, 10, 156–161.
- Landguth, E. L., Forester, B. R., Eckert, A. J., Shirk, A. J., Menon, M., Whipple, A., Day, C. C., Cushman, S. A. (2020). Modelling multilocus selection in an individual-based, spatially-explicit landscape genetics framework. *Molecular Ecology Resources*, 20(2), 605–615. <http://dx.doi.org/10.1111/1755-0998.13121>
- Legendre, P., & Legendre, L. (2012). *Numerical ecology*. Elsevier.
- Li, J., Li, H., Jakobsson, M., Li, S., Sjödin, P., & Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: Where do we stand and where could we go? *Molecular Ecology*, 21, 28–44. <https://doi.org/10.1111/j.1365-294X.2011.05308.x>
- Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, 4, 981–994.
- 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.
- Manel, S., Andreello, M., Henry, K., Verdelet, D., Darracq, A., Guerin, P. E., Desprez, B., Devaux, P. (2018). Predicting genotype environmental range from genome-environment associations. *Molecular Ecology*, 27(13), 2823–2833. <http://dx.doi.org/10.1111/mec.14723>
- McGuire, J. L., Lawler, J. J., McRae, B. H., Nuñez, T. A., & Theobald, D. M. (2016). Achieving climate connectivity in a fragmented landscape. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 7195–7200.
- Merilä, J., & Hendry, A. P. (2014). Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evolutionary Applications*, 7, 1–14.
- Nardone, A., Ronchi, B., Lacetera, N., Ranieri, M. S., & Bernabucci, U. (2010). Effects of climate changes on animal production and sustainability of livestock systems. *Livestock Science*, 130, 57–69.
- Nicotra, A. B., Beever, E. A., Robertson, A. L., Hofmann, G. E., & O'Leary, J. (2015). Assessing the components of adaptive capacity to improve conservation and management efforts under global change. *Conservation Biology*, 29, 1268–1278.
- Opdam, P., & Wascher, D. (2004). Climate change meets habitat fragmentation: Linking landscape and biogeographical scale levels in research and conservation. *Biological Conservation*, 117, 285–297.
- Orr, H. A., & Unckless, R. L. (2008). Population extinction and the genetics of adaptation. *The American Naturalist*, 172, 160–169.
- Pandey, P., Ramegowda, V., & Senthil-Kumar, M. (2015). Shared and unique responses of plants to multiple individual stresses and stress combinations: Physiological and molecular mechanisms. *Frontiers in Plant Science*, 6. <https://doi.org/10.3389/fpls.2015.00723>
- Peterson, M. L., Doak, D. F., & Morris, W. F. (2019). Incorporating local adaptation into forecasts of species' distribution and abundance under climate change. *Global Change Biology*, 25, 775–793.
- Phillips, P. C. (2008). Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9, 855–867.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38, 904–909.
- Pritchard, J. K., & Rienzo, A. D. (2010). Adaptation – Not by sweeps alone. *Nature Reviews Genetics*, 11, 665–667.
- Razgour, O., Forester, B., Taggart, J. B., Bekaert, M., Juste, J., & Ibáñez, C. (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.
- Reich, D., Price, A. L., & Patterson, N. (2008). Principal component analysis of genetic data. *Nature Genetics*, 40, 491–492.
- Relstab, 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.
- Reside, A. E., Butt, N., & Adams, V. M. (2018). Adapting systematic conservation planning for climate change. *Biodiversity and Conservation*, 27, 1–29.
- Rochat, E., Leempoel, K., Vajana, E., Colli, L., Ajmone-Marsan, P., Joost, S., & the NEXTGEN Consortium. (2016). *Map of genotype frequency change in autochthonous Moroccan sheep breeds due to global warming, Conference on Conservation genomics*, Vairao, Portugal.
- Selmoni, O., Vajana, E., Guillaume, A., Rochat, E., & Joost, S. (2020). Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach. *Molecular Ecology Resources*, 20, 154–169.
- Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, 4, 326–337.
- Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill, J. A., Hoesel, A. R., Höglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., ... Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, 30(2), 78–87. <http://dx.doi.org/10.1016/j.tree.2014.11.009>
- Stucki, S. (2014). Développement d'outils de géo-calcul haute performance pour l'identification de régions du génome potentiellement soumises à la sélection naturelle. *Infoscience*.
- Stucki, S., Orozco-terWengel, P., Forester, B. R., Duruz, S., Colli, L., Masembe, C., Negrini, R., Landguth, E., Jones, M. R., Bruford, M. W., Taberlet, P., Joost, S. (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources*, 17(5), 1072–1089. <http://dx.doi.org/10.1111/1755-0998.12629>
- Thompson, M. J., & Jiggins, C. D. (2014). Supergenes and their role in evolution. *Heredity*, 113, 1–8.
- Tramblay, Y., Badi, W., Driouech, F., El Adlouni, S., Neppel, L., & Servat, E. (2012). Climate change impacts on extreme precipitation in Morocco. *Global and Planetary Change*, 82–83, 104–114.
- Vajana, E., Barbato, M., Colli, L., Milanese, M., Rochat, E., Fabrizi, E., Mukasa, C., Del Corvo, M., Masembe, C., Muwanika, V. B., Kabi, F., Sonstegard, T. S., Huson, H. J., Negrini, R., Joost, S., & Ajmone-Marsan,

- P. (2018). Combining landscape genomics and ecological modelling to investigate local adaptation of indigenous Ugandan cattle to east coast fever. *Frontiers in Genetics*, 9. <https://doi.org/10.3389/fgene.2018.00385>
- Waldvogel, A. M., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C., Kofler, R., Mock, T., Schmid, K., Schmitt, I., Bataillon, T., Savolainen, O., Bergland, A., Flatt, T., Guillaume, F., Pfenninger, M. (2020). Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters*, 4(1), 4–18. <http://dx.doi.org/10.1002/evl3.154>
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326–3328.

#### BIOSKETCH

The research team is the Laboratory of Geographic Information System (LASIG) at the Ecole Polytechnique Fédérale de Lausanne (EPFL). For about fifteen years, LASIG contributed to the development of statistical methods in landscape genomics and featured numerous research works related to adaptation of livestock and wild species.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Rochat E, Selmoni O, Joost S. Spatial areas of genotype probability: Predicting the spatial distribution of adaptive genetic variants under future climatic conditions. *Divers Distrib*. 2021;00:1–15. <https://doi.org/10.1111/ddi.13256>