Spatial Areas of Genotype Probability (SPAG): predicting the spatial distribution of adaptive genetic variants under future climatic conditions

Estelle Rochat[1, 2], Stéphane Joost[1, 2, 3]

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

[2] The ADAPTMAP Consortium

[3] The NEXTGEN Consortium

Abstract

In a context of rapid global change, one of the key components for the survival of species is their genetic adaptive potential. Many methods have been developed to identify adaptive genetic variants, but few tools were made available to integrate this knowledge into conservation management. We present here the SPatial Areas of Genotype probability (SPAG), using genotype-environment logistic associations to map the probability of finding beneficial variants in a study area. We define a univariate model predicting the spatial distribution of a single genotype, and three multivariate models allowing the integration of several genotypes, potentially associated with various environmental variables. We then integrate climate change projections to map the corresponding future distribution of genotypes. The analysis of the mismatch between current and future SPAGs makes it possible to identify a) populations that are better adapted to the future climate through the presence of genetic variants able to cope with future conditions, and b) vulnerable populations where genotype(s) of interest are not frequent enough for the individuals to adapt to the future climate. We validate the SPAG approach using simulations and we use it to study the potential adaptation of 161

Moroccan and 382 European goats to the bioclimatic conditions. In Morocco, using whole genome sequence data, we identify seven genomic regions strongly associated with the precipitation seasonality (WorldClim database). The predicted shift in SPAGs under a strong climate change scenario for 2070 highlights goat populations likely to be threatened by the expected increase in precipitation variation in the future. In Europe, we find genomic regions associated with low precipitation, the shift in SPAGs highlighting vulnerable populations not adapted to the very dry conditions expected in 2070. The SPAG methodology is successfully validated using cross-validations and provides an efficient tool to take the adaptive potential into account in general conservation frameworks.

Introduction

Climate change has altered the conditions of various organisms, causing an average warming of 0.2°C per decade over the past 30 years, as well as a sea level rise of more than 3 mm per year since the start of the century and an increase in the frequency of extreme weather events such as storms, droughts and floods (IPBES, 2019). These changes are likely to continue in the future (IPCC, 2014). When such important changes occur, many animal and plant species are confronted with a shift away from the favourable conditions necessary for their survival (IPBES, 2019). In order to avoid extinction under these conditions, they can either move to more favourable areas or adapt to their new environment (Hughes, 2000). Due to limitations in dispersal capacity, loss of favourable habitats and increased landscape fragmentation, the possibilities for dispersal to new areas are often limited (Opdam and Wascher, 2004; McGuire *et al.*, 2016). Population adaptation relies on phenotypic changes, which can be induced either by phenotypic plasticity or genetic evolution (Merilä and Hendry, 2014; Fox *et al.*, 2019). Phenotypic plasticity can allow species to rapidly evolve by changing their behaviour, physiology or morphology (Reed *et al.*, 2011; Fox *et al.*, 2019). However, it can also potentially lead to a fitness reduction (Duputié *et al.*, 2015) and, since it is not based on heritable genetic variations, it will not necessarily ensure the persistence of adaptation for the next generations. In order to preserve biodiversity, it is therefore crucial to promote the

conservation of the genetic adaptive potential of populations (Hoffmann and Sgrò, 2011; Sgrò *et al.*, 2011; Nicotra *et al.*, 2015; Shafer *et al.*, 2015).

Conservation of this adaptive potential is also of major importance for livestock management in order to ensure herd persistence (Hoffmann, 2010; FAO, 2015). Indeed, many livestock populations, especially in developing country, are breed in pastoralist systems and live most of the time outdoors, confronted to difficult production conditions (e.g. heat stress, poor food resources and the presence of parasites and diseases), and without significant food or water supplies (FAO, 2015). These populations show signatures of local adaptation to their climatic conditions (McManus *et al.*, 2009, 2011; FAO, 2015; Bertolini *et al.*, 2018), to limited food resources (Silanikove, 2000) or to the presence of parasites (Noyes *et al.*, 2011; FAO, 2015; Vajana *et al.*, 2018). However, due to the increasing demand for food production, these local breeds currently tend to be replaced by high-producing commercial breeds imported from developed countries (Rischkowsky and Pilling, 2007; Hoffmann, 2010). This has led to a loss of genetic diversity, which threatens the adaptive potential of livestock species to environmental changes (FAO, 2015). In addition, imported breeds lack in the locally adapted genetic variants, which may reduce their fitness (FAO, 2015). It is therefore essential to highlight the adaptive potential of livestock species in order to encourage farmers to conserve local traditional breeds, and to carefully design cross-breeding, translocation or artificial selection (Scherf *et al.*, 2008; Allendorf *et al.*, 2010).

One of the essential components of the adaptive capacity of populations is genetic diversity (Allendorf and Leary, 1986). Since mutation rates are generally low, adaptation to rapid environmental changes largely depends on the amount of genetic variants already present in populations, i.e. standing genetic diversity (Orr and Unckless, 2008). With the recent increase in the availability of genetic data and the development of conservation genomics, various tools have been developed to integrate genetic diversity into conservation frameworks (Bonin *et al.*, 2007; Vandergast *et al.*, 2011; Thomassen *et al.*, 2011). However, conserving neutral variation in populations may not be sufficient to allow rapid adaptation to increasingly stressful

conditions (Reed and Frankham, 2001), and it could be more valuable to specifically preserve adaptive variation, i.e. variation associated with a trait involved in fitness (Hoffmann and Willi, 2008; Sgrò *et al.*, 2011; Willoughby *et al.*, 2018) or to combine both approaches (Funk *et al.*, 2012; Pauls *et al.*, 2013). Increasing attention is currently being paid to this issue in conservation discussions (Funk *et al.*, 2019; Hoelzel *et al.*, 2019; Mable, 2019).

Several methods have been developed to identify signatures of local adaptation, based on various assumptions and with different limitations and advantages (Schoville et al., 2012; Joost et al., 2013; Vitti et al., 2013; Hoban et al., 2016). The results have notably been used to establish prediction of future habitat range of species facing climate change (Hällfors et al., 2016; Ikeda et al., 2017; Garzón et al., 2019; Razgour et al., 2019). However, there is currently a need to integrate this knowledge in order to predict the distribution of locally adapted genetic variants along environmental gradients, and to project the probability of finding them in un-sampled areas or under future climatic conditions (Bay et al., 2017). Nevertheless, very few studies have addressed this issue. Fournier *et al.* (2011) identified locally adapted SNP alleles and then used the Maxent species distribution model (Phillips et al., 2006) to estimate their spatial distribution. Fitzpatrick and Keller (2015) proposed another approach based on two community-level modelling methods (Generalised Dissimilarity Modelling and Gradient Forest). They used it to map the current and future spatial distribution of several adaptive variants and to assess the “genetic offset” of populations under climate change as a function of the mismatch between the current distributions and future predictions. However, practical applications of these methods remain limited and there is still a need to develop new tools to ease the integration of the adaptive potential into conservation practices, especially by considering several loci, potentially non-independent and adapted to different environmental conditions.

We propose here a novel approach to predict genotype frequencies and map SPatial Areas of Genotypes Probabilities (SPAG) based on logistic genotype-environment associations (Joost *et* al., 2007) and conditional probability theory. SPAGs can be used to a) predict the probability

of presence of one or many locally adapted genetic variants in non-sampled areas b) identify areas where there is a greater probability of finding individuals better adapted to future climatic conditions, c) identify vulnerable populations that may be threatened by climate change and d) integrate the results into conservation frameworks by means of an easy combination with other georeferenced layers. The concept of applying logistic regressions on an environmental layer to predict the probability of presence of a genotype had been sketched out several years ago (Joost 2006; page 138). Here, we formalise this concept and extend it to multivariate models. We introduce the theoretical bases of SPAGs and validate the approach with a simulated dataset. We then present an application of our approach to two case studies in order to analyse the local adaptive potential of Moroccan and European goat populations.

Material and Methods

SPAG's approach

Logistic regressions (SAM)

The Spatial Analysis Method (SAM, Joost *et al.*, 2007) can be used to detect genotypes that are strongly associated with an environmental variable and are therefore potential adaptive variants (strictly additive). This 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 genotype *G1* as a function of an environmental variable *(x1)*,

$$p(G1) = p(G1 = 1 \mid x_1) = \frac{e^{\beta_0+\beta_1 x_1}}{1 + e^{\beta_0+\beta_1 x_1}}$$

Formula 1

where β0 and β1 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

121 ratio (G) compares the likelihood of a model with the likelihood of a null model without the
122 environmental variable of interest. The null hypothesis is that the model considered is no better
123 than the null model, i.e. the environmental variable considered does not help in predicting the
124 probability of presence of the genotype. The Wald test is a common statistic used to estimate
125 whether a parameter is equal to a given value. In our case, it is used to reject the null
126 hypothesis that the β parameter associated with an environmental variable is equal to 0, which
127 would also indicate that it has no effect on the probability of finding the genotype. The SAM
128 approach has been implemented in the Samβada software and validated against other
129 methods for identifying signatures of natural selection (Stucki *et al.*, 2017).

130 Univariate SPAG

131 Once the genotype(s) involved into significant associations with the environmental variables
132 have been identified, Formula 1 enables the estimation of the probability of presence of a
133 genotype for any value of an environmental variable (*x1*). We consequently used it to estimate
134 and delimit on a map the probability of presence of a genotype over the whole region of interest
135 (Joost, 2006; Rochat *et al.*, 2016). We named such a delimited surface univariate Spatial Area
136 of Genotype Probability (SPAG).

137 As more than one adaptive locus are usually identified, we also developed multivariate models
138 to compute a single map showing the probability of presence of multiple genotypes. Three
139 different multivariate models were developed to date: the Intersection, Union and K-
140 Percentage.

141 Intersection SPAG (I-SPAG)

142 The **Intersection** model (I-SPAG) is used to compute the probability that the variants of
143 interest are all simultaneously present. Following the theory of conditional probability
144 (Kolmogorov, 1956), the probability of simultaneous presence of *n* adaptive genotypes *Gi,*
145 *i=1:n* can be computed using Formula 2:

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

Formula 2

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

$$p(G_n | \cap_{i=1}^{n-1} G_i) = \frac{e^{\beta_0+\beta_1 x_n+\beta_2 \cap_{i=1}^{n-1} G_i}}{1+e^{\beta_0+\beta_1 x_n+\beta_2 \cap_{i=1}^{n-1} G_i}}$$

Formula 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, i.e. also where *Gi* values are unknown, we suggested to estimate $\cap_{i=1}^{n-1} G_i$ by $p(\cap_{i=1}^{n-1} G_i)$.

Using the associative property of the intersection operator, the intersection of *n* genotypes can be computed by starting with the univariate model *p(G1)*, which is used as a covariate to compute *p(G1 ∩ G2)*, itself used to compute *p(G3 ∩ (G1 ∩ G2)),* etc.. Formula 3 can thus be implemented with a recursive model based on the univariate formula in which covariates are added (see Supp. File 1 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.

<u>Union SPAG (U-SPAG)</u>

The **union** model (U-SPAG) is used to compute the probability of finding at least one of the adaptive genotypes of interest. We implemented it with the inclusion-exclusion principle (e.g. for two genotypes: p(G1 ∪ G2) = p(G1) + p(G2) – p(G1 ∩ G2)). We implemented the generalised formula for *n* adaptive genotypes (Formula 4), as an *R* function based on the intersection model previously described (see code availability and Suppl. File 1).

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

Formula 4

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 genotypes. This probability can be computed by combining formulas from the union and intersection models (Formula 5, explained in more detail in Supp. File 1). Again, this formula was implemented as an *R* function (see code availability).

$$p(K\%\ G_{i=1\ldots n}) = p\left(\bigcup_{i=1}^{n} \bigcap_{1 \le i_1 < i_2 < \cdots < i_{(K\% * n+1)}} (G_{i_1} \cap G_{i_2} \cap \ldots \cap G_{i_k})\right)$$

Formula 5

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

Simulation study

In order to test the SPAG's approach, we first computed a simulated dataset using the individual-based population genetics model software CDPOP 1.3 (Landguth and 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 500x500 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

189 three loci under selection (L0, L1 and L2). The selection was implemented using three 500x500
190 raster gradients, the first from north to south (X0), the second from east to west (X1) and the
191 third from northwest to southeast (X2) (see Figure1A). We set the average effects
192 bL0A0A0=10 and bL0A1A1=-10 for the locus L0 with the environmental variable X0, which
193 indicates that the genotype A0A0 from locus L0 will be favoured in the South (where X0=1),
194 whereas A1A1 will be favoured in the North (where X0=-1). We set similar effects for the locus
195 L1 with the environmental variable X1 and L2 with X2. All other beta effects were set to 0,
196 indicating no influence of the environmental variable to the distribution of genotypes. All
197 genotypes were randomly initialised at the beginning of the simulations. The exact list of
198 simulation parameters used is provided in Supp. File 2.

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

204 **Moroccan and European goats**

205 Genetic data

206 Two genetic datasets characterising goats (Capra hircus) were used for the analyses
207 presented here. The first one was produced in the context of the NEXTGEN project (Alberto
208 *et al.*, 2018) and the second was collected by the ADAPTMAP consortium (Stella *et al.*, 2018;
209 http://www.goatadaptmap.org/).

210 The NEXTGEN project produced whole genome sequences data for 161 Moroccan goats from
211 6 different local breeds. Since goat production system in Morocco is mainly free-range, these
212 goats are living from 8 to 12 months outdoors (Boujenane, 2005), and are confronted to
213 contrasting environmental conditions, from the Sahara desert to the Atlas Mountains (see
214 Figure A in Suppl. File 3). 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 ADAPTMAP consortium gathered genetic data for 4'563 goats from 144 breeds, sequenced worldwide with the CaprineSNP50 BeadChip and mapped on the most recent goat reference genome ARS1 (Bickhart *et al.*, 2017). The goats were georeferenced to the place where they have been sampled. We used here a subset of these data, constituted of individuals from Switzerland, North of Italy and France. This represented 458 individuals distributed in 196 locations, with 1 to 39 individuals per site. In order to avoid overweighting of some locations, we selected a maximum of five individuals per sampling site. These five individuals were chosen as the subset showing the highest Nei's genetic distances, computed with the function *dist.genpop* from the package *adegenet* (Jombart, 2008) in the *R* environment (R Development Core Team, 2008). The resulting dataset contains 382 individuals from 196 locations and 11 different breeds (see Figure B in Supp. File 3).

Both genetic datasets were 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 datasets contain 8,497,971 SNPs for the Moroccan goats and 46,294 SNPs for the European ones.

<u>Environmental data</u>

The climatic conditions of the sampling locations were characterised using the 19 bioclimatic variables (Supp. File 4) 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. 1km2) and values were extracted for all sampling locations using the *extract* function from the R-package *raster* (Hijmans and van Etten, 2012). In order to get similar ranges of values for all bioclimatic variables, which makes it easier to compare the subsequently derived models, all variables

241 were standardised for each dataset, by subtracting the mean and dividing by the standard
242 deviation. Some of the bioclimatic variables are highly correlated. However, we choose to keep
243 all of them to be able to identify *a posteriori* which variable had the strongest effect. Since no
244 models computed involved more than one environmental variable simultaneously, this
245 collinearity will not impact the results.

246 Population Structure

247 The genetic population structure was estimated with a Principal Component Analysis (Price *et*
248 *al.*, 2006; Reich *et al.*, 2008) computed with the function *snpgdsPCA* from the *SNPRelate* R-
249 package (Zheng *et al.*, 2012). In order to avoid a strong influence of SNP clusters on this
250 analysis, we used here a pruned set of SNPs that are in approximate linkage equilibrium with
251 each other. The pruning was performed with the function *snpgdsLDpruning* from the
252 *SNPRelate* package, with a threshold D'=0.2. The resulting datasets contain 59,224 SNPs for
253 the Moroccan goats and 14,571 SNPs for the European ones.

254 Logistic regressions and SPAGs

255 For the two datasets, logistic models were computed for each genotype with the 19 bioclimatic
256 variables. The statistical significance of the model was assessed using Wald test and log-
257 likelihood ratio (G), both corrected for the false-discovery rate due to multiple comparisons
258 using the procedure proposed by Benjamini and Hochberg (1995), under an expected false
259 discovery rate (FDR) of 0.05 (i.e. 5% of the results expected to be false positives).

260 In order to lower the number of false positive resulting from demographic processes instead of
261 natural selection (Li *et al.*, 2012), logistic models were computed with the addition of covariates
262 corresponding to the coordinates of individuals on the significant components of the PCA. The
263 significance of the models with population covariates was assessed using a Wald test and a
264 log likelihood ratio which compares the model with environment and covariates to the model
265 with covariates only. An association was considered as significant if both the models without
266 covariates and with population covariates were significant.

267 Finally, to identify potential functions of the SNPs involved into the significant associations, we
268 used the NCBI Genome Data Viewer
269 (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_000317765.1) to
270 search for the presence of annotated genes in the genomic region of 10kbp surrounding the
271 SNPs of interest. All analyses were computed using a combination of the Samβada software
272 (Stucki *et al.*, 2017) and a custom R-script based on the *glm* function.

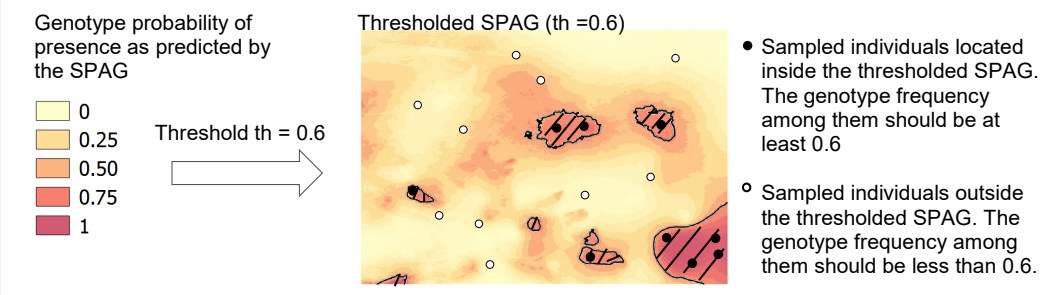
273 **Validation procedure**

274 For the simulated data as well as for the two goat datasets, SPAGs were validated with a
275 cross-validation procedure, using 25% of the individuals to compute the SPAG (i.e. 50
276 individuals for the simulated datasets, 41 individuals for the Moroccan goats and 96 for the
277 European) and the remaining 75% to test it. Training individuals were selected such to
278 represent the entire range of values of the environmental variable under study (see Supp. File
279 1 for the exact procedure). The model was validated using the Area Under the Receiver
280 Operating Curve computed with the testing dataset (AUCtest, (Fielding and Bell, 1997)) and a
281 custom validation graph presented in Box 1. The cross-validation procedure was repeated 10
282 times.

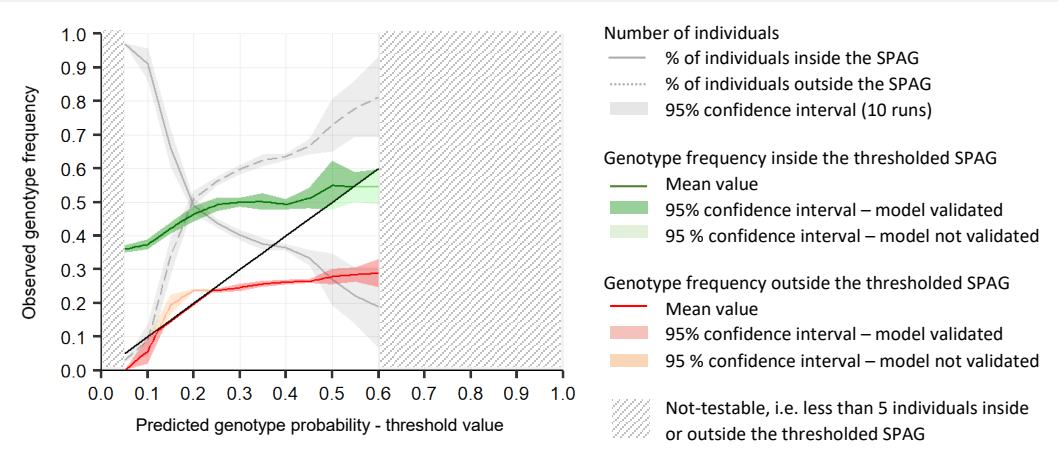
283

284 **Box 1: Validation Procedure**

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



To validate the SPAGs, 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. 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.



306 *For a threshold value th=0.4, the SPAG is validated since 50% of the testing individuals located within*
307 *the area SPAG≥0.4 carry the genotype of interest, whereas only 26% carry it ouside (SPAG<0.4).*
308 *Inversely, the model is not validated for th=0.6, since only 54% of individuals carry the genotype of*
309 *interest whithin the area where the SPAG predicted a probability of at least 0.6 (SPAG ≥0.6). The model*
310 *is also not valid for a value th=0.2 since 23% of individuals carry the genotype of interest in the area*
311 *SPAG<0.2.*
312

313 **Projections under climate change**

314 We use the two goat datasets to present an application of the SPAG for predicting the
315 distribution of adaptive genotype(s) under climate change. In order to predict the genotype
316 frequency optimal for future conditions, we retrieved Worldclim data for the year 2070,
317 corresponding to a strong climate change scenario from the Max Planck Institute Earth System
318 Model (MPI-ESM-LR) (Giorgetta *et al.*, 2013) with a Representative Concentration Pathway
319 equals to 8.5 (RCP 8.5). We then assume that the optimal genotype frequency for future
320 conditions should be close to the genotype frequency currently observed in areas with climatic
321 conditions resembling the future ones. We thus applied the current parameters of the logistic
322 regressions on the future environmental variables in order to derive the future SPAGs for the
323 genotypes of interest. We then study the shift between the current and future SPAGs to identify
324 vulnerable populations for which specific genotype frequencies should be much higher so that
325 individuals can adapt to the future conditions.

326 ***Results***

327 **Simulated results**

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

332 ***Table 1 – Most significant models – Simulated datasets***
333 10 most ignificant models obtained for the analysis of the simulated datasets, ranked based on the
334 likelihood ratio. Geno = Genotype, Env=environmental predictor, pG=p-value associated with the G
335 score, pW=p-value associated with the Wald score, β0 and β1 are the parameters of the logistic
336 regression.

ID	Marker	Locus	Geno.	Env.	G score	pG	pW	β0	β1
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

337 Figure 1B presents the univariate SPAG for the model S1 (Locus L1, genotype A0A0
338 associated with the environmental variable X1). Since this locus was simulated as under
339 selection with the east-west gradient X1, the resulting SPAG coherently shows a similar
340 gradient. The results are validated by the validation graph for almost the entire range of
341 probability values, except close to 0.8, where the SPAG slightly overestimate the probability of
342 presence of the genotype (the green line falls below the black one, i.e. the observed genotype
343 frequency is lower than what predicted by the SPAG). Figure 1C shows the intersection SPAG
344 for the genotype A0A0 of the three loci under selection. It indicates a very low probability of
345 finding them all simultaneously, except in the South-East of the simulated area (where all
346 environmental gradients considered had values close to 1). Figure 1D presents a union-SPAG
347 of three genotypes. It shows that the probability of finding at least one of them is higher than
348 0.5 in most of the area. The validation graph indicates that the model tends to underestimate
349 the probability of finding the genotypes for threshold values below 0.6 (the red line is above
350 the black one, i.e. the genotype frequency outside the thresholded SPAG is higher than what
351 predicted). Finally, Figure 1E depicts the probability of finding 70% of 4 genotypes of interest,
352 i.e. at least 3 of them. This probability is low, except in the South-East and North-West part.

353 Again, the validation graph indicates a good power of the SPAG to predict the probability of
354 finding a set of genotypes of interest.
355

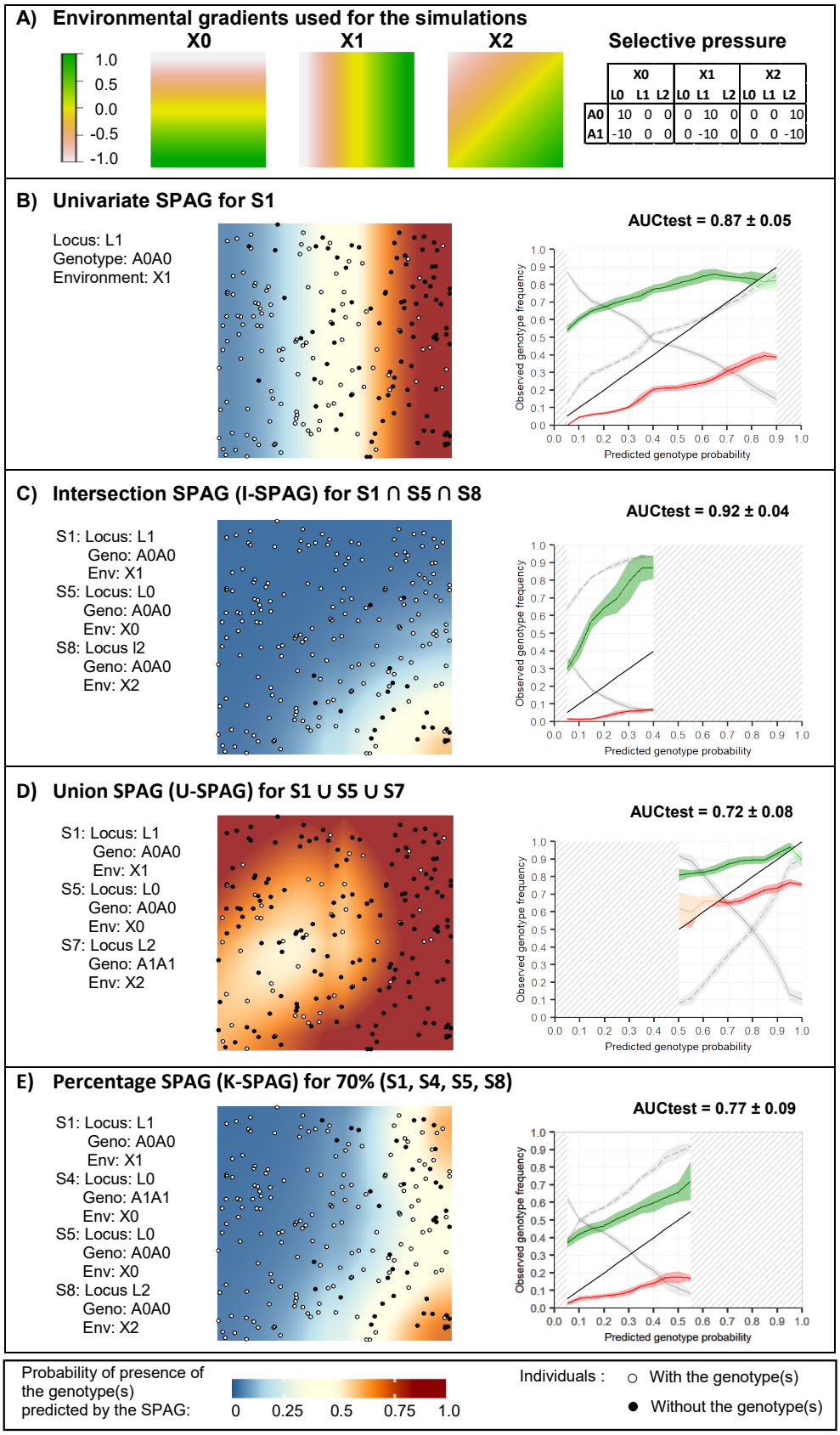


Figure 1 – SPAG – Simulated dataset

Univariate and Multivariate Spatial Areas of Genotypes Probability for the simulated dataset. The identifiers of the presented models (S1, S4, S5, S7, S8) refers to Table 1. Please refer to Box 1 to interpret the validation graphs shown on the right of each map.

361 **Moroccan goats**

362 Population structure

363 For the Moroccan dataset, the cumulated variance explained by the 10 first PCA components
364 on the SNP markers represents only 8.1% of the total variance and the increase in variance
365 explained is almost proportional to the number of components, which highlights that there is
366 no clear sub-structure. We therefore do not include any population structure on the subsequent
367 analysis and computed only logistic regressions without any covariates.

368 Logistic regressions

369 More than 483 million logistic association models were computed. After correction for false
370 discovery rate with a significant threshold of 5%, no model is significant according to the Wald
371 score, but seven models are significant according to the G score (Supp. File 5). Among them,
372 three models were strongly associated with the precipitation seasonality (bio15), which is a
373 measure of the variation of monthly precipitation over the year. Following this initial result, we
374 investigated in more details the adaptation to this bioclimatic variable. When considering only
375 the associations involving bio15 (25,447,348 models), 78 models are significant after FDR-
376 correction of G score, with a significant threshold of 5% (Supp. File 5). The SNPs involved in
377 these models are located on seven different genomic regions (Table 2), corresponding to four
378 annotated genes (DSG4, CDH2, KCTD1 and WRN) on the reference genome CHIR 1.0.
379

380 ***Table 2 – Significant models – Moroccan datasets – Bio15***
381 Significant models obtained for the analysis of Moroccan datasets with precipitation seasonality
382 (bio15) after FDR correction. Chr=Chromosome, Start=Start in base pairs of the region identified as
383 under selection, End=End in base pairs of the region, Peak SNP = SNP of the most significant model
384 in that region, Geno = corresponding SNP Genotype, GF=corresponding Genotype Frequency, β0 and
385 β1 = parameters of the logistic regression, G=G score (Log Likelihood ratio), qG=corresponding p-
386 value corrected for FDR, Genes = Annotated genes on the genomic region.

ID	Chr	Start (BP)	End (BP)	Peak (BP)	Geno	GF	G	qG	β0	β1	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

387 Spatial Areas of Genotype Probability

388 Figure2A shows the univariate SPAG for the genotype of model M1 presented in Table 2 (see
389 Supp. File 7 for the other univariate SPAGs). The predicted probability of presence of the
390 genotype is the highest in the extreme southwest of the country, near the Sahara desert. In
391 this region, the variations of precipitation are the greatest (the standard deviation of monthly
392 precipitation is more than 100% of the mean of monthly precipitation) and all goats carry the
393 genotype of interest. In coastal areas, the predicted probability of finding the genotype is close
394 to 0.5. In this area, the variations of precipitations are also high (more than 70%) and some of
395 the sampled goats carry the genotype, while others do not. Finally, in the Atlas Mountains and
396 in the northeast of the country, the probability of finding the genotype of model M1 is much
397 lower (<0.2 in most areas). In these regions, variations of precipitation are less important (35-
398 50%) and most of the goats sampled do not carry the genotype.

399 Two other markers positively correlated with bio15 were highlighted by models M3 and M5
400 (Table 2). Nevertheless, the I-SPAG presented in Figure 2B indicates that their simultaneous
401 presence is very unlikely (probability <0.1 for most of the territory). On the contrary, the

402 probability of finding at least one of the genotypes from the three models M1, M3 and M5, all
403 positively associated with the coefficient of precipitation, is very high in many parts of the
404 territory (U-SPAG, Figure 2C). Finally the K-SPAG presented in Figure 1D shows the
405 probability that goats carry at least 50% of the four variants positively associated with the
406 coefficient of precipitation (M1, M2, M3, M5), i.e. the probability to find at least two of them.
407 This map is the most contrasted, showing a very high probability of presence near the coast
408 and the Sahara desert (> 0.9) and a very low probability (<0.2) in the centre and northeast of
409 the country. For all these multivariate cases, the mean AUC value for the testing dataset over
410 the 10 runs is greater than 0.8. In addition, the validation graphs indicate that the SPAGs
411 computed with 25% of the individuals generally enable a correct estimate of the probability of
412 finding the genotype(s) of interest in the 75% remaining individuals. Only the U-SPAG (Figure
413 2C) tends to slightly overestimate the probability of presence since we observe a higher
414 presence of the genotypes in the individuals located outside the thresholded SPAG as
415 compare to what predicted by the SPAG (the red line is above the black line).
416

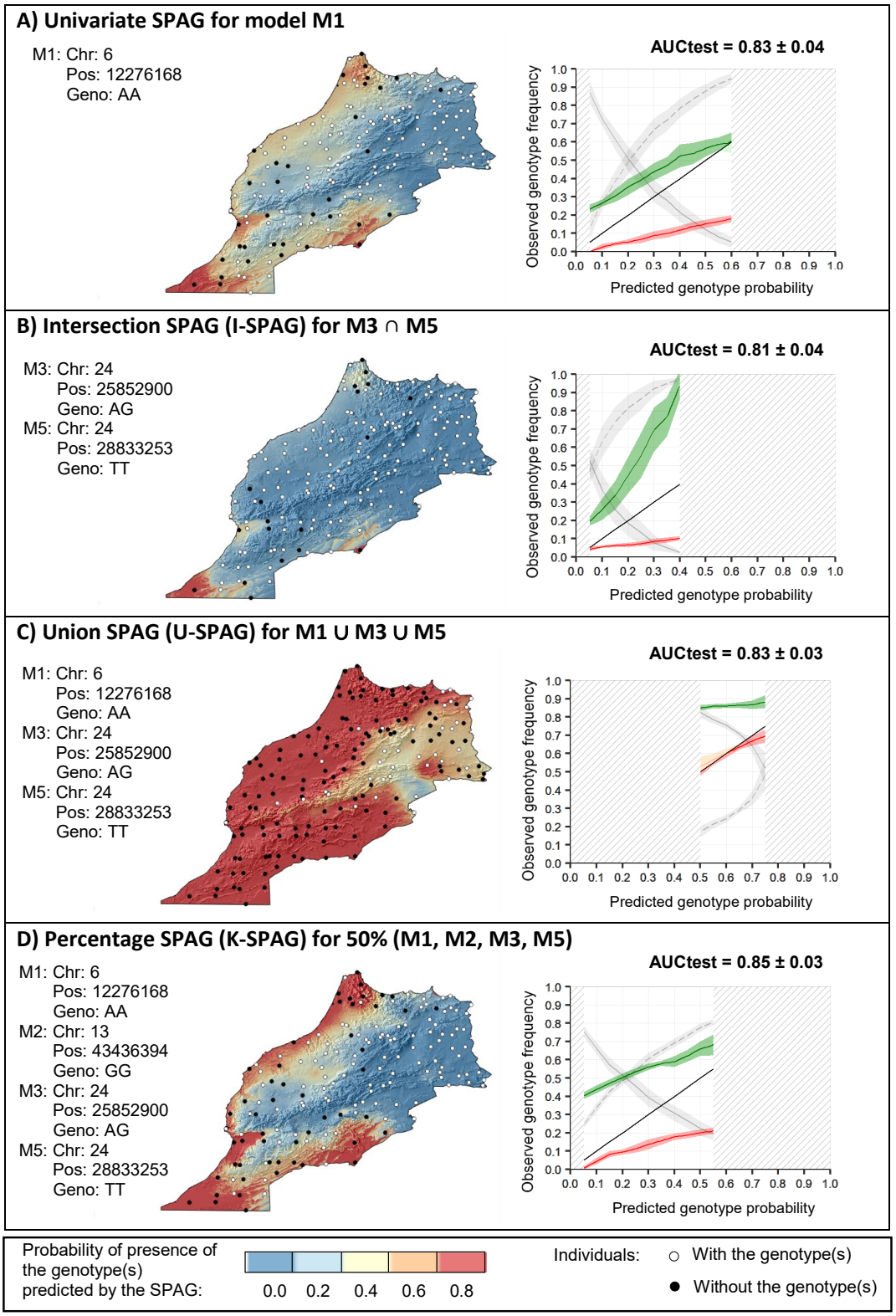


Figure 2– SPAG – Moroccan dataset

Univariate and Multivariate Spatial Areas of Genotypes Probability for the Moroccan dataset. The identifiers of the presented models (M1, M2, M3, M5) 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.

425 Projections under climate change

426 Figure 3 shows the differences between the current SPAGs presented in Figure 2 and their
427 corresponding projections for 2070. In Morocco, the precipitation seasonality (bio15) is
428 predicted to increase in the northwest of the country, with a maximum increase of 5 to 10% in
429 the extreme northwestern region (Tangier-Tetouan, see region's map in Supp. File 3) and to
430 decrease in other areas, especially in the Atlas Mountains and near the Sahara (from -10 to -
431 20%). The evolution of the univariate SPAG for model M1 (Figure 3A) consequently indicates
432 the highest risk in the Tangier-Tetouan area, where the mismatch between the current and
433 future SPAGs indicates that the probability of finding the genotype of interest should be 20%
434 higher to find individuals well adapted to future conditions. However, many individuals in this
435 area already carry the favourable genotype, and the risk for the population may thus be
436 reduced thanks to natural gene flow. Nevertheless, this is not the case in the southwest of this
437 area (Rabat, Casablanca) where the probability of finding the genotype should also be 10-20%
438 higher according to the SPAGs difference, and none of the goats sampled there currently carry
439 the adaptive variant. Similar observations can be made as regards the I-SPAG of M3 and M5
440 (Figure 3B), two other markers that may potentially confer an adaptation to high variations of
441 precipitation. However, the U-SPAG (Figure 3C) highlights no vulnerable areas, which
442 indicates that if the presence of at least one of the adaptive variants is sufficient to enable the
443 adaptation to high variations of precipitation, no population may be at risk. Finally the K-SPAG
444 (Figure 3D) also shows a risk area in the northwest of the country, where the probability of
445 carrying the adaptive variants should be approximately 20% higher. Again, individuals in the
446 northernmost part of this risk area may be less threatened due to the close presence of goats
447 already carrying the favourable genotypes, whereas the population from the Rabat-
448 Casablanca area may be more threatened due to the current much lower presence of the
449 adapted variants.

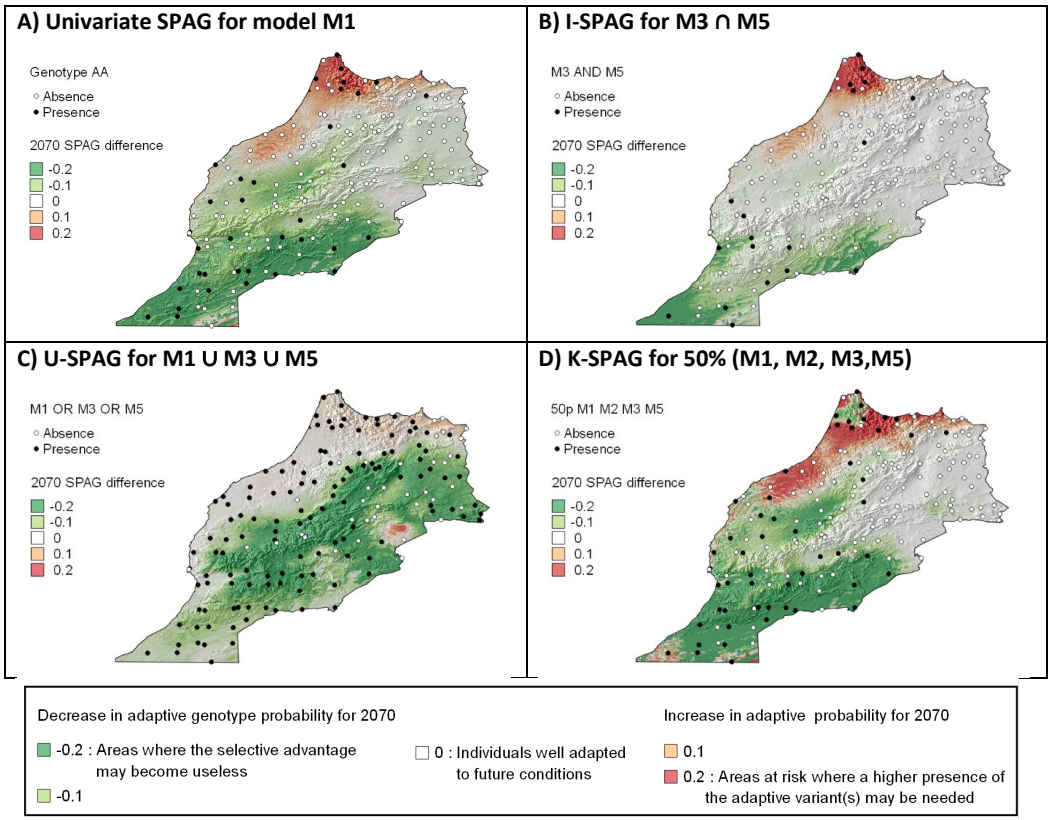


Figure 3– Moroccan goats - Predicted change in genotype probability for 2070
Predicted SPAG difference for 2070 considering the MPI-ESM-LR climate change scenario with RCP 8.5, for the Moroccan goats. The identifiers of the presented models (M1, M2, M3, M5) refer to Table 2. The maps show the average difference in probabilities of finding the genotype(s) based on the 10 runs computed with different selection of training sets.

European goats

Population structure

For the European goats, the first component of the PCA explains 6.2% of the total variance while the second, third and fourth components explain 2.0%, 1.7% and 1.6% respectively. The low variance explained by each PCA component indicates the absence of a clear population structure. However, since the variance explained by the first component is much higher than that explained by the next ones, it is possible that the first component is partially related to the population structure. We therefore computed logistic regressions without covariates and then logistic regressions with a covariate corresponding to the coordinates of goat individuals on the first component of the PCA.

466 Logistic regressions

467 More than 2.6 million logistic association models were computed, of which 4.9% were
468 significant both without covariate and with the first PCA-component as covariate, according to
469 both G score and Wald score corrected for a false positive rate of 5% (Supp. File 6). The ten
470 genomic regions associated with the strongest G scores when computed without covariate are
471 presented in Table 3. The corresponding models involved two bioclimatic variables related to
472 precipitation (bio13 = precipitation of the wettest month, bio18 = precipitation of the warmest
473 quarter) and two bioclimatic variables related to temperature (bio3 = isothermality, bio8 = mean
474 temperature of the wettest quarter). Seven annotated genes correspond exactly to one of the
475 SNPs identified: KRT12, CSN1S2, CACNB2, PRDM5, LOC102174324, PALM and NAV3.

476 ***Table 3– Significant models – European datasets***
477 Models corresponding to the 10 most significant genomic regions (based on G score of the model
478 without covariate) obtained for the analysis of the European dataset, considering all bioclimatic
479 variables. Chr=Chromosome, BP=Position in base pairs, GENO=Genotype of interest, GF=Genotype
480 Frequency, qGpop (resp. qWpop) = FDR-corrected p-values of Gscore (resp. Wald score) of the model
481 with the first PCA-component as covariate, qG0 (resp. qW0) = FDR-corrected p-values of Gscore (resp.
482 Wald score) of the models without any covariate, β0 and β1 = parameters of the logistic regression
483 without covariate, Genes = Annotated genes on the genomic region.

ID	ENV	CHR	BP	GENO	GF	qGpop	qWpop	qG0	qW0	β0	β1	Genes
E1a	bio18	19	40696776	GG	33.8	3.3E-09	3.9E-07	3.8E-17	1.2E-11	-1.41	1.30	KRT12
E1b	bio13	19	40696776	AA	40.3	4.3E-09	5.4E-07	1.4E-13	6.9E-10	-0.52	-1.07	KRT12
E2	bio18	1	38183832	AA	31.7	3.0E-09	1.3E-07	1.6E-15	1.2E-11	-0.97	1.13	-
E3a	bio3	6	86081075	CC	31.2	3.5E-11	1.2E-08	3.1E-14	4.2E-11	0.65	-1.06	CSN1S2
E3b	bio18	6	86081075	CC	31.2	9.6E-11	6.9E-08	5.7E-14	5.0E-10	0.71	1.12	CSN1S2
E4	bio8	13	32300758	GG	31.2	6.1E-10	9.0E-08	5.1E-14	5.6E-11	0.11	-1.02	CACNB2
E5	bio18	5	23213822	GG	39.0	6.0E-10	4.1E-08	7.3E-14	5.6E-11	-0.52	1.02	-
E6	bio8	6	4945809	AA	19.4	9.3E-07	7.9E-05	1.0E-13	9.2E-08	-2.06	1.55	PRDM5
E7	bio18	16	76397454	GG	22.8	1.1E-10	5.3E-07	1.0E-13	8.0E-09	1.30	1.29	LOC102174324
E8	bio18	7	67159272	CC	35.1	6.9E-10	1.3E-07	1.0E-13	2.3E-10	0.27	1.03	PALM
E9	bio3	5	7093719	GG	39.8	7.0E-11	1.8E-08	1.0E-13	9.0E-11	-0.63	1.02	NAV3
E10	bio18	14	85434737	AA	47.4	2.6E-09	1.6E-07	1.3E-13	2.2E-10	-0.16	-1.01	-

484

485 SPatial Areas of Genotype Probability

486 Figure4A shows the univariate SPAG for the genotype of model E10 presented in Table 3 (see
487 Supp. File 7 for the other univariate SPAGs). This genotype is negatively associated with
488 precipitation of the warmest quarter (bio18). The predicted probability of finding this genotype
489 is the lowest in the Swiss Alps (<0.1), slightly higher in the Jura, French Alps and Swiss Plateau
490 (between 0.2 and 0.4) and above 0.5 everywhere else, with a maximum around 0.8 on the
491 Mediterranean coast. The I-SPAG (Figure 5B) shows the probability of simultaneously finding
492 this genotype (model E10) and the genotype from model E1b, associated with low precipitation
493 during the wettest month (bio13). The two models E1b and E10 may therefore indicate an
494 adaptation to low values of precipitation. However, the I-SPAG indicates that their
495 simultaneous presence is not very likely (predicted frequency < 0.6 everywhere). The predicted
496 probability is the highest in the centre-north of France (regions Centre, Iles de France, East of
497 Pays de la Loire, Normandie and South of Hauts de France, see Supp. File 3 for regions' map)
498 and in the southern part (Occitanie and West of Provence), whereas in the Alps, Jura and most
499 of Switzerland, the probability of simultaneously finding these two genotypes is close to 0 and
500 none of the sampled goats carry them both. The probability of finding at least one of these two
501 variants, presented on the U-SPAG in Figure 4C, indicates a trend similar to the probability of
502 presence of E10 alone (Figure 4A), but with even stronger contrast between the Alps-Jura-
503 Switzerland area (frequencies < 0.3) and the rest of the territory (frequencies > 0.7). Finally
504 the K-SPAG (Figure 4D) shows the probability of finding at least 50% of five genotypes
505 negatively associated with precipitation in the warmest quarter (bio18), i.e. the probability of
506 finding at least three of them. Note that for the models positively associated with bio18 in Table
507 3, we used the alternative genotype that was the most significantly negatively correlated with
508 bio18 (indicated after the model ID). The resulting SPAG is very close to the I-SPAG of E1b
509 and E10 (Figure 4B). For all cases presented, the validation graphs indicate that the SPAGs
510 computed with 25% of the individuals generally correctly predict the genotype frequency of the
511 remaining 75% of individuals.

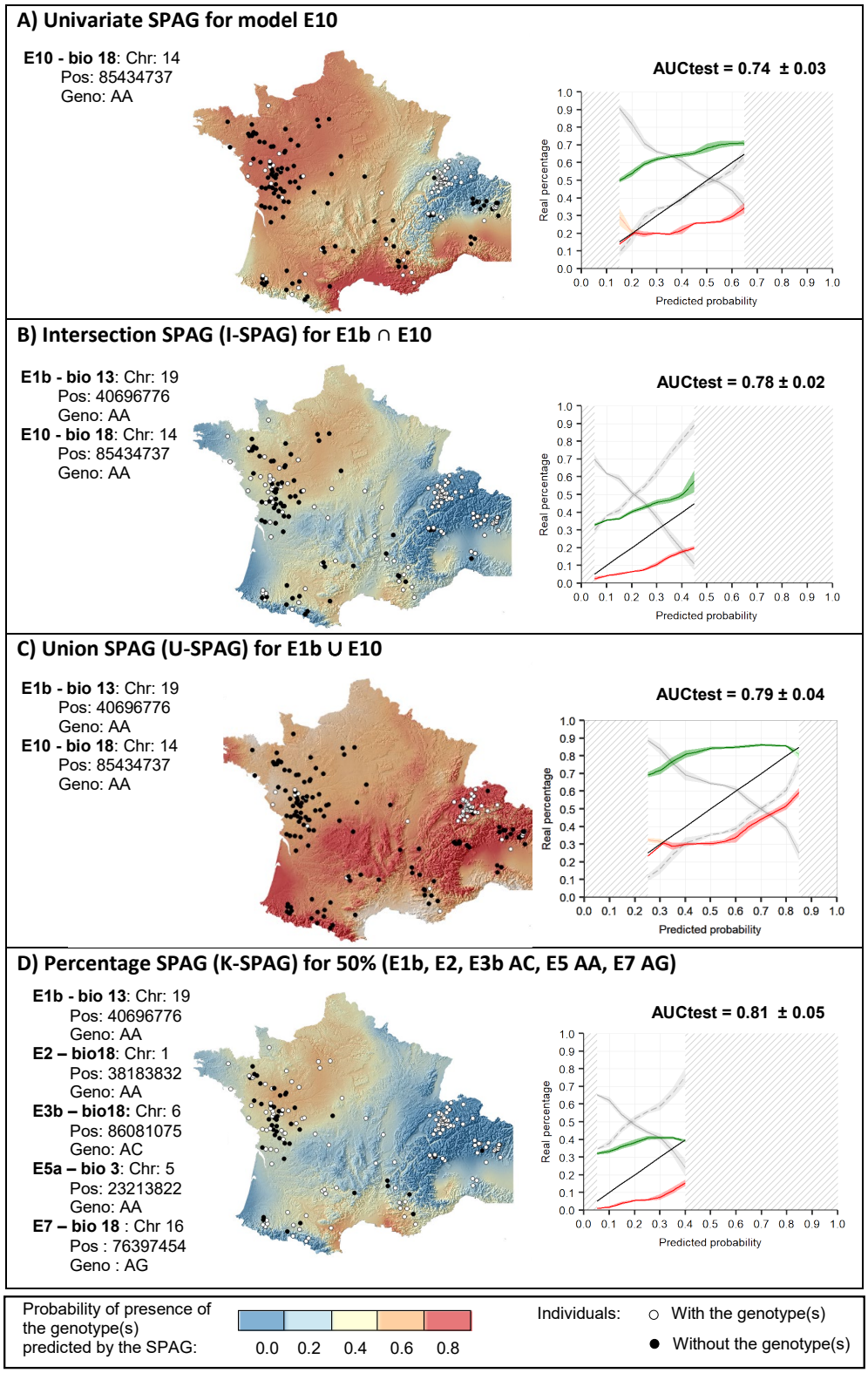


Figure 4 – SPAG – European dataset

Univariate and Multivariate Spatial Areas of Genotypes Probability for the European dataset. The identifiers of the presented models (E10, E1d, E2, E3b, E5, E7) refers to Table 3. 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. Note that since up to five individuals can be localised on the same site, a black dot indicates a presence if at least 50% of the individuals of the site carry the marker(s). Please refer to Box 1 to interpret the validation graphs shown on the right of each map.

521 Projection under climate change

522 Figure 6 shows the differences between the current SPAGs presented in Figure 5 and their
523 corresponding projections for 2070. The model E10 presented in Figure 5A was related to the
524 precipitation of the warmest quarter (bio 18), which is projected to decrease by 20 to 180 mm
525 over the whole study area until 2070. The projected precipitation loss is maximum in the Alps
526 (-120 to -180 mm), on the Mediterranean coast (-100 to -130) and in the centre-south of France
527 (West of Auvergne and East of Nouvelle Acquitaine, -90 to -120 mm, see region's map in Supp.
528 File 3). As a consequence, the mismatch between current and future SPAGs for model E10
529 (Figure 5A) indicates that all populations may be threatened by climate change, since the
530 probability of finding the adaptive genotype should be 10 to 20% higher everywhere. None of
531 the goats living in the Alps and Switzerland currently carry this adaptive genotype and these
532 populations may thus be particularly vulnerable. When considering the I-SPAG of E10 and E1b
533 (Figure 5B), a higher risk is highlighted in the centre-south of France (West of Auvergne and
534 its surrounding), northeast of France (Alsace in East of Grand-Est), Swiss Plateau, and
535 northwest of Italy. In centre-south of France and Swiss-Plateau, almost none of the goats
536 sampled carry simultaneously the two genotypes and the populations may therefore be
537 particularly threatened. The U-SPAG of the same genotypes (Figure 5C) shows results very
538 similar to the univariate SPAG for E10 alone (Figure 5A). If the presence of at least one of
539 these two genotypes may be sufficient to allow adaptation to low precipitation, then the most
540 threatened regions will again be the Alps and Switzerland, where both markers are currently
541 absent. Finally, the evolution of the K-SPAG (Figure 4D) shows a high risk in a large part of
542 France and northern Italy, where the SPAG's mismatch indicate that the probability of finding
543 the adaptive genotypes should be more than 20% higher.
544

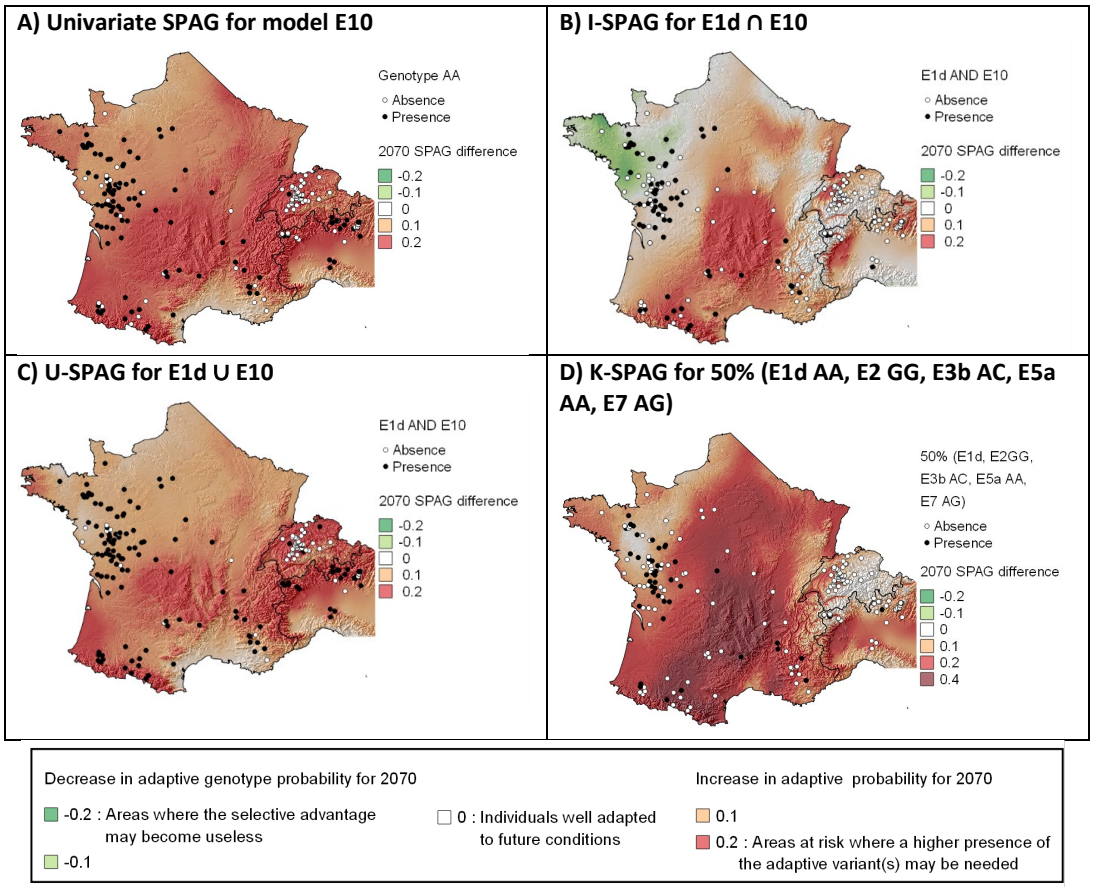


Figure 5– Predicted change in genotype probability for 2070
Predicted change in genotype probability for 2070 considering the MPI-ESM-LR climate change scenario with RCP 8.5, for the European goats. The identifiers of the presented models (E10, E1b, E2, E3b, E5a, E7) refer to Table 3. The maps show the average difference in probability of finding the genotype(s) based on the 10 runs computed with different random selection of training sets.

Discussion

Mapping genotype probabilities

The first utility of SPAGs is to quantify the current probability of finding beneficial genotypes or the expression of favourable traits in plant and animal populations, even in regions where no individuals have been sampled. Our results show that with few training individuals (i.e. 50 simulated individuals, 41 Moroccan or 96 European goats), a good estimate of the probability of finding genotype(s) of interest is possible. The univariate models presented here have already been applied to map the genotype frequencies of adaptive variants of the Scandinavian brown bears (Joost, 2006), Moroccan sheep (Rochat *et al.,* 2016) and coral reefs (Selmoni *et al.*, 2019). Multivariate models are presented here for the first time and, according

561 to the validation procedure applied on a simulated dataset and two case studies, they appear
562 to be powerful in estimating the combined probability of finding several genotypes potentially
563 correlated with different environmental variables. With the I-SPAG, the resulting probabilities
564 may rapidly become very low, but this model could be used when we suspect that the
565 simultaneous presence of some adaptive genotypes is needed to ensure the adaptation, or
566 when we would like to highlight the probability of simultaneously finding variants that may
567 confer an adaptive response to different environmental variables (e.g. low precipitation and
568 high temperature). At the other extreme, the U-SPAG may rapidly indicate high probabilities of
569 presence in most parts of the territory, but it can be used when it is suspected that the presence
570 of at least one of the genotypes may be sufficient to confer adaptive potential. Since it is
571 generally difficult to know whether the simultaneous presence of adaptive genotypes is needed
572 or whether an union is sufficient, K-SPAG offers an interesting compromise, allowing the
573 identification of populations that retain a given percentage of variants, which can help delineate
574 areas where there is the highest probability of finding individuals with high adaptive potential.

575 **From SPAG to conservation**

576 The study of the shift in SPAGs under climate change conditions can help identify 1) well-
577 adapted populations, where individuals currently show adaptive genotypes that seem to be
578 optimal under future conditions, 2) populations at risk where the current genotype frequency is
579 not optimal, but where the favourable genotypes are already present in the population, thus
580 potentially allowing a natural increase in genotype frequency through gene flow and 3)
581 threatened populations where optimal variants are currently lacking but would be needed to
582 ensure adaptation to future climate. These identifications may be of great value for
583 conservation planning. Indeed, when prioritising conservation areas, success could be
584 enhanced by choosing to preserve preadapted individuals that already carry functional variants
585 conferring them good adaptation to future climate (Orr and Unckless, 2008). Moreover SPAGs
586 can also be used to prevent the translocation of individuals that do not currently carry the
587 variants favourable for future conditions at the target site, which would result in a reduction or

588 loss of adaptive potential of the target populations (Weeks *et al.*, 2011). In addition,
589 conservation plans can be developed to increase the survival capacity of threatened
590 populations. This may involve assisted gene flow to import adaptive variants into a population
591 where they are lacking (Aitken and Whitlock, 2013; Kelly and Phillips, 2016) or artificial
592 selection of individuals already pre-adapted to future conditions (Hoffmann, 2010). However,
593 this has to be undertaken carefully since the selection of locally adapted individuals can result
594 in a loss of genetic diversity (Savage *et al.*, 2018), which may decrease the potential of
595 populations to adapt to new environmental changes. Kardos and Shafer (2018) therefore
596 proposed that gene-targeted conservation measures should only be taken with traits affecting
597 vital processes of the species and when phenotypic variation is large enough to ensure a high
598 probability of success. Finally, the SPAG maps presented could be integrated into decision-
599 making frameworks considering the adaptive potential when defining the vulnerability of
600 species (Bonin *et al.*, 2007; Williams *et al.*, 2008; Sgrò *et al.*, 2011; Dawson *et al.*, 2011;
601 Razgour *et al.*, 2018) or in more global decision frameworks that take into account other
602 vulnerability factors such as the predation level or habitat loss.

603 **The goat's example**

604 Several signatures of local adaptation where identified for the goats under study. In Morocco,
605 three of the genes identified (DSG4, KCTD1 and CDH2) may be related to the development of
606 hair (Kljuic *et al.*, 2003; Ling *et al.*, 2014; Wang *et al.*, 2017; Zhang *et al.*, 2019) or skin
607 properties (Hayashi *et al.*, 2007). These results suggest that goats confronted with high
608 variations of precipitation may have adapted through a modification of hair or skin traits, which
609 could for example ensure a better water repulsion. In Europe, two of the genes highlighted as
610 potentially conferring an adaptation to drought conditions (KRT12 and PRDM5) may be related
611 to properties of the cornea (Kao *et al.*, 1996; Burkitt Wright *et al.*, 2011). They could thus
612 potentially highlight an adaptation to higher UV-radiation associated with driest conditions. The
613 other genes identified are related the casein content of the milk (CSN1S2, Ramunno *et al.*,
614 2001), the calcium channel and energy pathway (CACNB2, Cardona *et al.*, 2014) or the skin

615 properties (PALM and NAV3, Kutzleb *et al.*, 1998; Karenko *et al.*, 2005). Many of the genes
616 highlighted on the two case studies may therefore be associated with a function that can be
617 influenced by climate, which reinforces the potential that they are true signatures of local
618 adaptation. However, it is known that logistic regressions such as implemented here, as most
619 of the other methods, may lead to the identification of false positive (Stucki *et al.*, 2017), and
620 the results should thus be confronted with other methods available to detect signatures of
621 natural selection. In addition, although previous studies show the power of genotype-
622 environment associations to predict phenotype (Lasky *et al.*, 2015; Vangestel *et al.*, 2018) or
623 fitness (Fournier-Level *et al.*, 2011; Hancock *et al.*, 2011), more investigations are needed to
624 verify that the genotypes identified are really conferring an adaptive advantage (Funk *et al.*,
625 2019).

626 The Moroccan case study highlighted that goat populations from the surroundings of Rabat
627 and Casablanca may lack adaptive genotypes potentially conferring an advantage to face high
628 variation of precipitation. If the adaptive role of these genotypes is validated, goat populations
629 in this region may be threatened. Because of the great economic and social importance of
630 goats in Morocco, it is crucial to preserve viable populations. Indeed, in this country, agriculture
631 contributes to 12 to 24% of the national GDP and employs 40% of the total active population
632 (Boujenane, 2005). Livestock farming, especially small ruminants, is the most important sector
633 of agriculture and goat farms account for 20% of the total number of farms (Boujenane, 2005).
634 It is therefore important to consider preserving or introducing the adaptive genotypes on each
635 vulnerable population. This could be done for example by favouring crossbreeding with
636 individuals from the southern or north-eastern part of the country, where adaptive genotypes
637 are currently well present, and by avoiding breeding or translocation with exotic goats or goats
638 from the Atlas or Oriental regions. In the northern part of Morocco (Tanger-Tetouan regions),
639 goat populations represent 12% of the national goat populations (Chentouf, 2014), and they
640 play an important role in preserving food security (Godber *et al.*, 2016). In this region,
641 crossbreeding with exotic breeds has been introduced to improve milk production (Boujenane,

642 2005; Godber *et al.*, 2016). However, our results show that the frequency of adaptive
643 genotypes should increase in the goat populations from this region, and that it is therefore
644 essential to maintain local individuals with the necessary adaptive genotypes.

645 **Limitations and perspectives**

646 The SPAG approach presented appears to be powerful for mapping the probability of finding
647 locally adapted genetic variants in a landscape. However, the adaptation process is complex
648 and often involves polygenic traits (Pritchard and Rienzo, 2010), for which the detection power
649 of the genotype-environment associations may be reduced (Villemereuil et al., 2014; Harrisson
650 et al., 2014). In this case, it may be advisable to use multivariate genotype-environment
651 association models (Forester et al., 2017) or to integrate other methods to identify SNPs
652 related to polygenic adaptation (Zhou et al., 2013; Lasky et al., 2015). In addition, since the
653 results of the shift under climate change may be highly dependent on the climate change
654 scenario considered, computations should be performed with various scenarios and less
655 weight should be given to the conclusions not consistent within scenarios (Reside *et al.*, 2018).
656 Finally, in order to assess the real vulnerability of populations, an analysis of connectivity
657 should be carried out to highlight the potential of natural gene flow to increase the probability
658 of finding favourable genotypes in threatened populations.

659 SPAGs could also be used to predict the presence of genotype(s) associated with other
660 pressures showing a spatial distribution, such as the presence of a parasite (Vajana *et al.*,
661 2018) or a predator (Cousyn *et al.*, 2001) or the urbanization level (Harris and Munshi-South,
662 2016). Very similar models can also be derived to predict allele frequencies instead of
663 genotypes frequencies or to integrate other covariates (e.g. to consider autocorrelation or to
664 use other indicators of population structure such as the Admixture coefficients). In addition,
665 SPAGs are provided as maps, which enables a visual identification of threatened populations
666 and could thus facilitate discussions between different conservation actors. SPAGs therefore

667 constitute a valuable tool to support conservation decisions, especially under current changing
668 climatic conditions.

669

670 ## *Conflict of Interest*

671 The authors declare that there is no conflict of interest regarding the publication of this paper.

672 ## *Code availability*

673 The R functions developed to compute the univariate and multivariate SPAGs are available at :
674 https://github.com/estellerochat/SPAG

675 ## *References*

676 Aitken SN, Whitlock MC (2013). Assisted Gene Flow to Facilitate Local Adaptation to Climate Change.
677 *Annual Review of Ecology, Evolution, and Systematics* **44**: 367–388.

678 Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, Villemereuil P de, *et al.* (2018).
679 Convergent genomic signatures of domestication in sheep and goats. *Nat Commun* **9**: 1–9.

680 Allendorf FW, Hohenlohe PA, Luikart G (2010). Genomics and the future of conservation genetics.
681 *Nat Rev Genet* **11**: 697–709.

682 Allendorf FW, Leary RF (1986). Heterozygosity and fitness in natural populations of animals.
683 *Conservation biology: the science of scarcity and diversity*: 57–76.

684 Benjamini Y, Hochberg Y (1995). Controlling the False Discovery Rate: A Practical and Powerful
685 Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*
686 *(Methodological)* **57**: 289–300.

687 Benjelloun B, Alberto FJ, Streeter I, Boyer F, Coissac E, Stucki S, *et al.* (2015). Characterizing neutral
688 genomic diversity and selection signatures in indigenous populations of Moroccan goats
689 (Capra hircus) using WGS data. *Front Genet* **6**.

690 Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, *et al.* (2018). Signatures of selection and
691 environmental adaptation across the goat genome post-domestication. *Genetics Selection*
692 *Evolution* **50**: 57.

693 Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, *et al.* (2017). Single-molecule
694 sequencing and chromatin conformation capture enable de novo reference assembly of the
695 domestic goat genome. *Nat Genet* **49**: 643–650.

696 Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P (2007). Population Adaptive Index: a New
697 Method to Help Measure Intraspecific Genetic Diversity and Prioritize Populations for
698 Conservation. *Conservation Biology* **21**: 697–708.

699 Boujenane I (2005). *Small Ruminant Breeds of Morocco*.

700 Burkitt Wright EMM, Spencer HL, Daly SB, Manson FDC, Zeef LAH, Urquhart J, *et al.* (2011). Mutations
701 in PRDM5 in Brittle Cornea Syndrome Identify a Pathway Regulating Extracellular Matrix
702 Development and Maintenance. *The American Journal of Human Genetics* **88**: 767–777.

703 Cardona A, Pagani L, Antao T, Lawson DJ, Eichstaedt CA, Yngvadottir B, *et al.* (2014). Genome-Wide
704 Analysis of Cold Adaptation in Indigenous Siberian Populations. *PLOS ONE* **9**: e98076.

705 Chentouf M (2014). Systèmes de production caprine au Nord du Maroc: Contraintes et propositions
706 d'amélioration. *Opt Médit A: Médit Semin* **108**: 25–32.

707 Cousyn C, Meester LD, Colbourne JK, Brendonck L, Verschuren D, Volckaert F (2001). Rapid, local
708 adaptation of zooplankton behavior to changes in predation pressure in the absence of
709 neutral genetic changes. *PNAS* **98**: 6256–6260.

710 Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM (2011). Beyond Predictions: Biodiversity
711 Conservation in a Changing Climate. *Science* **332**: 53–58.

712 Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, *et al.* (2013). Sequencing and automated whole-
713 genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature*
714 *Biotechnology* **31**: 135–141.

715 Duputié A, Rutschmann A, Ronce O, Chuine I (2015). Phenological plasticity will not help all species
716 adapt to climate change. *Global Change Biology* **21**: 3062–3073.

717 FAO (2015). The Second Report on the State of the World's Animal Genetic Resources for Food and
718 Agriculture.

719 Fielding AH, Bell JF (1997). A review of methods for the assessment of prediction errors in
720 conservation presence/absence models. *Environmental Conservation* **24**: 38–49.

721 Forester BR, Lasky JR, Wagner HH, Urban DL (2017). Using genotype-environment associations to
722 identify multilocus local adaptation. *bioRxiv*: 129460.

723 Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011). A Map of Local
724 Adaptation in Arabidopsis thaliana. *Science* **334**: 86–89.

725 Fox RJ, Donelson JM, Schunter C, Ravasi T, Gaitán-Espitia JD (2019). Beyond buying time: the role of
726 plasticity in phenotypic adaptation to rapid environmental change. *Philosophical*
727 *Transactions of the Royal Society B: Biological Sciences* **374**: 20180174.

728 Funk WC, Forester BR, Converse SJ, Darst C, Morey S (2019). Improving conservation policy with
729 genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act
730 decisions for conservation practitioners and geneticists. *Conserv Genet* **20**: 115–134.

731 Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012). Harnessing genomics for delineating
732 conservation units. *Trends in Ecology & Evolution* **27**: 489–496.

Garzón MB, Robson TM, Hampe A (2019). ΔTraitSDMs: species distribution models that account for local adaptation and phenotypic plasticity. *New Phytologist* **222**: 1757–1765.

Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Böttinger M, *et al.* (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*: 572–597.

Godber OF, Laroussi BF, Chentouf M, Wall R (2016). Intensification of Mediterranean Goat Production Systems: A Case Study in Northern Morocco. *Agriculture* **6**: 16.

Hällfors MH, Liao J, Dzurisin J, Grundel R, Hyvärinen M, Towle K, *et al.* (2016). Addressing potential local adaptation in species distribution models: implications for conservation under climate change. *Ecological Applications* **26**: 1154–1169.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, *et al.* (2011). Adaptation to Climate Across the Arabidopsis thaliana Genome. *Science* **334**: 83–86.

Harris SE, Munshi-South J (2016). Scans for positive selection reveal candidate genes and local adaptation of Peromyscus leucopus populations to urbanization. *bioRxiv*: 038141.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol Appl* **7**: 1008–1025.

Hayashi R, Yamato M, Sugiyama H, Sumide T, Yang J, Okano T, *et al.* (2007). N-Cadherin Is Expressed by Putative Stem/Progenitor Cells and Melanocytes in the Human Limbal Epithelial Stem Cell Niche. *STEM CELLS* **25**: 289–296.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.

754 Hijmans RJ, van Etten J (2012). raster: Geographic analysis and modeling with raster data. R package
755 version 2.0-12. http://CRAN.R-project.org/package=raster.

756 Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, *et al.* (2016). Finding the
757 Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The*
758 *American Naturalist* **188**: 379–397.

759 Hoelzel AR, Bruford MW, Fleischer RC (2019). Conservation of adaptive potential and functional
760 diversity. *Conserv Genet* **20**: 1–5.

761 Hoffmann I (2010). Climate change and the characterization, breeding and conservation of animal
762 genetic resources. *Animal Genetics* **41**: 32–46.

763 Hoffmann AA, Sgrò CM (2011). Climate change and evolutionary adaptation. *Nature* **470**: 479–485.

764 Hoffmann AA, Willi Y (2008). Detecting genetic responses to environmental change. *Nat Rev Genet* **9**:
765 421–432.

766 Hughes L (2000). Biological consequences of global warming: is the signal already apparent? *Trends in*
767 *Ecology & Evolution* **15**: 56–61.

768 Ikeda DH, Max TL, Allan GJ, Lau MK, Shuster SM, Whitham TG (2017). Genetically informed ecological
769 niche models improve climate change predictions. *Global Change Biology* **23**: 164–176.

770 IPBES (2019). Global assessment report on biodiversity and ecosystem services of the
771 Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. *E S*
772 *Brondizio, J Settele, S Díaz, and H T Ngo (editors)*.

773 Jombart T (2008). adegenet: a R package for the multivariate analysis of genetic markers.
774 *Bioinformatics* **24**: 1403–1405.

775 Joost S (2006). The geographical dimension of genetic diversity : a GIScience contribution for the
776 conservation of animal genetic resources.

777 Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, *et al.* (2007). A spatial analysis method
778 (SAM) to detect candidate loci for selection: towards a landscape genomics approach to
779 adaptation. *Molecular Ecology* **16**: 3955–3969.

780 Joost S, Vuilleumier S, Jensen JD, Schoville S, Leempoel K, Stucki S, *et al.* (2013). Uncovering the
781 genetic basis of adaptive change: on the intersection of landscape genomics and theoretical
782 population genetics. *Mol Ecol* **22**: 3659–3665.

783 Kao WW, Liu CY, Converse RL, Shiraishi A, Kao CW, Ishizaki M, *et al.* (1996). Keratin 12-deficient mice
784 have fragile corneal epithelia. *Invest Ophthalmol Vis Sci* **37**: 2572–2584.

785 Kardos M, Shafer ABA (2018). The Peril of Gene-Targeted Conservation. *Trends in Ecology & Evolution*
786 **33**: 827–839.

787 Karenko L, Hahtola S, Päivinen S, Karhu R, Syrjä S, Kähkönen M, *et al.* (2005). Primary Cutaneous T-
788 Cell Lymphomas Show a Deletion or Translocation Affecting NAV3, the Human UNC-53
789 Homologue. *Cancer Res* **65**: 8101–8110.

790 Kelly E, Phillips BL (2016). Targeted gene flow for conservation. *Conservation Biology* **30**: 259–267.

791 Kljuic A, Bazzi H, Sundberg JP, Martinez-Mir A, O'Shaughnessy R, Mahoney MG, *et al.* (2003).
792 Desmoglein 4 in Hair Follicle Differentiation and Epidermal Adhesion: Evidence from
793 Inherited Hypotrichosis and Acquired Pemphigus Vulgaris. *Cell* **113**: 249–260.

794 Kolmogorov A (1956). Foundations of the Theory of Probability. *Chelsea Pub Co*.

795 Kutzleb C, Sanders G, Yamamoto R, Wang X, Lichte B, Petrasch-Parwez E, *et al.* (1998). Paralemmin, a
796 Prenyl-Palmitoyl–anchored Phosphoprotein Abundant in Neurons and Implicated in Plasma
797 Membrane Dynamics and Cell Process Formation. *The Journal of Cell Biology* **143**: 795–813.

798 Landguth EL, Cushman SamuelA (2010). cdpop: A spatially explicit cost distance population genetics
799 program. *Molecular Ecology Resources* **10**: 156–161.

800 Landguth EL, Forester BR, Eckert AJ, Shirk AJ, Menon M, Whipple A, *et al.* (2020). Modelling
801 multilocus selection in an individual-based, spatially-explicit landscape genetics framework.
802 *Molecular Ecology Resources* **20**: 605–615.

803 Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, *et al.* (2015). Genome-
804 environment associations in sorghum landraces predict adaptive traits. *Science Advances* **1**:
805 e1400218.

806 Li J, Li H, Jakobsson M, Li S, Sjödin P, Lascoux M (2012). Joint analysis of demography and selection in
807 population genetics: where do we stand and where could we go? *Molecular Ecology* **21**: 28–
808 44.

809 Ling YH, Xiang, H, Zhang G, Ding JP, Zhang ZJ, Zhang YH, *et al.* (2014). Identification of complete
810 linkage disequilibrium in the DSG4 gene and its association with wool length and crimp in
811 Chinese indigenous sheep. *GMR | Genetics and Molecular Research | The Original by*
812 *FUNPEC-RP* **13**: 5617–5625.

813 Mable BK (2019). Conservation of adaptive potential and functional diversity: integrating old and
814 new approaches. *Conserv Genet* **20**: 89–100.

815 McGuire JL, Lawler JJ, McRae BH, Nuñez TA, Theobald DM (2016). Achieving climate connectivity in a
816 fragmented landscape. *PNAS* **113**: 7195–7200.

817 McManus C, Louvandini H, Gugel R, Sasaki LCB, Bianchini E, Bernal FEM, *et al.* (2011). Skin and coat
818 traits in sheep in Brazil and their relation with heat tolerance. *Trop Anim Health Prod* **43**:
819 121–126.

820 McManus C, Prescott E, Paludo GR, Bianchini E, Louvandini H, Mariante AS (2009). Heat tolerance in
821 naturalized Brazilian cattle breeds. *Livestock Science* **120**: 256–264.

822 Merilä J, Hendry AP (2014). Climate change, adaptation, and phenotypic plasticity: the problem and
823 the evidence. *Evolutionary Applications* **7**: 1–14.

824 Nicotra AB, Beever EA, Robertson AL, Hofmann GE, O'Leary J (2015). Assessing the components of
825 adaptive capacity to improve conservation and management efforts under global change.
826 *Conservation Biology* **29**: 1268–1278.

827 Noyes H, Brass A, Obara I, Anderson S, Archibald AL, Bradley DG, *et al.* (2011). Genetic and
828 expression analysis of cattle identifies candidate genes in pathways responding to
829 Trypanosoma congolense infection. *PNAS* **108**: 9304–9309.

830 Opdam P, Wascher D (2004). Climate change meets habitat fragmentation: linking landscape and
831 biogeographical scale levels in research and conservation. *Biological Conservation* **117**: 285–
832 297.

833 Orr HA, Unckless RL (2008). Population Extinction and the Genetics of Adaptation. *The American*
834 *Naturalist* **172**: 160–169.

835 Pauls SU, Nowak C, Bálint M, Pfenninger M (2013). The impact of global climate change on genetic
836 diversity within populations and species. *Mol Ecol* **22**: 925–946.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**: 904–909.

Pritchard JK, Rienzo AD (2010). Adaptation – not by sweeps alone. *Nat Rev Genet* **11**: 665–667.

R Development Core Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Ramunno L, Cosenza G, Pappalardo M, Longobardi E, Gallo D, Pastore N, *et al.* (2001). Characterization of two new alleles at the goat CSN1S2 locus. *Animal Genetics* **32**: 264–268.

Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibáñez C, *et al.* (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *PNAS* **116**: 10418–10423.

Razgour O, Taggart JB, Manel S, Juste J, Ibáñez C, Rebelo H, *et al.* (2018). An integrated framework to identify wildlife populations under threat from climate change. *Molecular Ecology Resources* **18**: 18–31.

Reed DH, Frankham R (2001). How Closely Correlated Are Molecular and Quantitative Measures of Genetic Variation? A Meta-Analysis. *Evolution* **55**: 1095–1103.

Reed TE, Schindler DE, Waples RS (2011). Interacting Effects of Phenotypic Plasticity and Evolution on Population Persistence in a Changing Climate. *Conservation Biology* **25**: 56–63.

Reich D, Price AL, Patterson N (2008). Principal component analysis of genetic data. *Nature Genetics* **40**: 491–492.

Reside AE, Butt N, Adams VM (2018). Adapting systematic conservation planning for climate change. *Biodivers Conserv* **27**: 1–29.

859 Rischkowsky B, Pilling D (2007). *The state of the world's animal genetic resources for food and*
860 *agriculture*. Food & Agriculture Org.

861 Rochat E, Leempoel K, Vajana E, Colli L, Ajmone-Marsan P, Joost S, *et al.* (2016). Map of genotype
862 frequency change in autochtonous Moroccan sheep breeds due to global warming.

863 Savage AE, Mulder KP, Torres T, Wells S (2018). Lost but not forgotten: MHC genotypes predict
864 overwinter survival despite depauperate MHC diversity in a declining frog. *Conserv Genet* **19**:
865 309–322.

866 Scherf B, Rischkowsky B, Hoffmann I, Wieczorek M, Montironi A, Cardellino R (2008). Livestock
867 Genetic Diversity in Dry Rangelands. In: Lee C, Schaaf T (eds) *The Future of Drylands*,
868 Springer Netherlands, pp 89–100.

869 Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S (2012). Adaptive Genetic
870 Variation on the Landscape: Methods and Cases. *Annual Review of Ecology, Evolution, and*
871 *Systematics* **43**: 23–43.

872 Selmoni O, Rochat E, Lecellier G, Berteaux-Lecellier V, Joost S (2019). Seascape genomics as a new
873 tool to empower coral reef conservation strategies: an example on north-western Pacific
874 Acropora digitifera. *bioRxiv*: 588228.

875 Sgrò CM, Lowe AJ, Hoffmann AA (2011). Building evolutionary resilience for conserving biodiversity
876 under climate change. *Evolutionary Applications* **4**: 326–337.

877 Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, *et al.* (2015). Genomics and
878 the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30**: 78–
879 87.

Silanikove N (2000). The physiological basis of adaptation in goats to harsh environments. *Small Ruminant Research* **35**: 181–193.

Stella A, Nicolazzi EL, Van Tassell CP, Rothschild MF, Colli L, Rosen BD, *et al.* (2018). AdaptMap: exploring goat diversity and adaptation. *Genet Sel Evol* **50**: 61.

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 BR, Duruz S, Colli L, Masembe C, *et al.* (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources* **17**: 1072–1089.

Thomassen HA, Fuller T, Buermann W, Milá B, Kieswetter CM, Jarrín-V. P, *et al.* (2011). Mapping evolutionary process: a multi-taxa approach to conservation prioritization. *Evolutionary Applications* **4**: 397–413.

Vajana E, Barbato M, Colli L, Milanesi M, Rochat E, Fabrizi E, *et al.* (2018). Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Front Genet* **9**.

Vandergast AG, Perry WM, Lugo RV, Hathaway SA (2011). Genetic landscapes GIS Toolbox: tools to map patterns of genetic divergence and diversity. *Molecular Ecology Resources* **11**: 158–161.

Vangestel C, Eckert AJ, Wegrzyn JL, St. Clair JB, Neale DB (2018). Linking phenotype, genotype and environment to unravel genetic components underlying cold hardiness in coastal Douglas-fir (Pseudotsuga menziesii var. menziesii). *Tree Genetics & Genomes* **14**: 10.

900 Villemereuil P de, Frichot É, Bazin É, François O, Gaggiotti OE (2014). Genome scan methods against
901 more complex models: when and how much should we trust them? *Molecular Ecology* **23**:
902 2006–2019.

903 Vitti JJ, Grossman SR, Sabeti PC (2013). Detecting Natural Selection in Genomic Data. *Annual Review*
904 *of Genetics* **47**: 97–120.

905 Wang S, Ge W, Luo Z, Guo Y, Jiao B, Qu L, *et al.* (2017). Integrated analysis of coding genes and non-
906 coding RNAs during hair follicle cycle of cashmere goat (Capra hircus). *BMC Genomics* **18**:
907 767.

908 Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, *et al.* (2011). Assessing the
909 benefits and risks of translocations in changing environments: a genetic perspective.
910 *Evolutionary Applications* **4**: 709–725.

911 Williams SE, Shoo LP, Isaac JL, Hoffmann AA, Langham G (2008). Towards an Integrated Framework
912 for Assessing the Vulnerability of Species to Climate Change. *PLOS Biol* **6**: e325.

913 Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR (2018). Rapid genetic adaptation
914 to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular*
915 *Ecology* **27**: 4041–4051.

916 Zhang Y, Wu K, Wang L, Wang Z, Han W, Chen D, *et al.* (2019). Comparative study on seasonal hair
917 follicle cycling by analysis of the transcriptomes from cashmere and milk goats. *Genomics*.

918 Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012). A high-performance computing
919 toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**:
920 3326–3328.

921 Zhou X, Carbonetto P, Stephens M (2013). Polygenic Modeling with Bayesian Sparse Linear Mixed
922 Models. *PLOS Genetics* **9**: e1003264.

923