RESOURCE ARTICLE: HBMs for distribution range shifts with GCC

A hierarchical Bayesian Beta regression approach to study the effects of geographic genetic structure and spatial autocorrelation on species distribution range shifts

Joaquín Martínez-Minaya[1], David Conesa[1], Marie-Josée Fortin[2], Carlos Alonso-Blanco[3], F. Xavier Picó[4] & Arnald Marcer[5,6]

[1]Departament d'Estadística i Investigació Operativa, Universitat de València, Burjassot, Valencia 46100, Spain

[2]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto (Ontario), M5S 3B2, Canada

[3]Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnología (CNB), Consejo Superior de Investigaciones Científicas (CSIC), 28049 Madrid, Spain

[4]Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), Sevilla 41092, Spain

[5]CREAF, E08193 Bellaterra (Cerdanyola del Vallès), Catalonia, Spain

[6]Universitat Autònoma de Barcelona, E08193 Bellaterra (Cerdanyola del Vallès), Catalonia, Spain

Correspondence: F. Xavier Picó, Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), Sevilla 41092, Spain. E-mail: xpico@ebd.csic.es

Abstract

Global climate change (GCC) may be imposing distribution range shifts in many organisms worldwide. Multiple efforts are currently focused on the development of models to better predict distribution range shifts due to GCC. We addressed this issue by including intra-specific genetic structure and spatial autocorrelation (SAC) of data in distribution range models. Both factors reflect the joint effect of eco-evolutionary processes on the geographic heterogeneity of populations. We used a collection of 301 geo-referenced accessions of the annual plant *Arabidopsis thaliana* in its Iberian Peninsula range, where the species shows a strong geographic genetic structure. We developed spatial and non-spatial hierarchical Bayesian models (HBMs) to depict current and future distribution ranges for the four genetic clusters detected. We also compared the performance of HBMs with `Maxent` (a presence-only model). `Maxent` and non-spatial HBM presented some shortcomings, such as the loss of accessions with high genetic admixture in the case of `Maxent` and the presence of residual SAC for both. As spatial HBMs removed residual SAC, these models showed higher accuracy than non-spatial HBMs and handled the spatial effect on model outcomes. The ease of modelling and the consistency among model outputs for each genetic cluster was conditioned by the sparseness of the populations across the distribution range. Our HBMs enrich the toolbox of software available to evaluate GCC-induced distribution range shifts considering both genetic heterogeneity and SAC, two inherent properties of any organism that should not be overlooked.

KEYWORDS: *Arabidopsis thaliana*, geographic genetic structure, global climate change, hierarchical Bayesian models, `Maxent`, spatial autocorrelation.

1. INTRODUCTION

Climate and land-use changes recorded in practically all Earth's bioclimatic zones are dramatically affecting the distribution of many terrestrial, aquatic and marine organisms. Since the turn of the century, various global meta-analyses have quantified the fingerprint that global climate change (GCC) has already left on distribution ranges (Parmesan & Yohe, 2003; Perry, Low, Ellis, & Reynolds, 2005; Parmesan, 2006; Chen, Hill, Ohlemüller, Roy, &Thomas, 2011; MacLean & Beissinger, 2017) and extinction rates (Urban, 2015; Wiens, 2016). At present, models are including some improvements to better predict changes in species distribution ranges due to GCC. Such improvements are aimed at considering all the possible organisms' responses to GCC, such as shifts in dispersal ability, phenology and physiology of life-history traits (Bellard, Bertelsmeier, Leadley, Thuiller, & Courchamp, 2012; Lenoir & Svenning, 2015). However, precise data on these responses are lacking for many organisms because of the intensive amount of labour and data needed to estimate them properly for multiple populations across large areas of the distribution range. Nonetheless, modelling approaches clearly have to go beyond presence-background models and related approaches (e.g. presence-absence models, random pseudo-absence point models) using current and future climatic conditions to increase their accuracy and reliability (Guisan, Thuiller, & Zimmermann, 2017). In this study, we address this issue by considering two important biological aspects that should be considered when modelling current distribution range of terrestrial organisms and their GCC-induced shifts.

Firstly, demographic processes (i.e. extinction/colonisation dynamics and dispersal ability) and adaptation to local environmental conditions determine the extent of population stratification, that is, geographically distributed allele frequencies depicting subpopulations or clusters at different spatial scales (Anderson et al., 2010). Molecular data are commonly used

to infer the number of genetically differentiated clusters and their degree of admixture (Pritchard, Stephens, & Donnelly, 2000; Falush, Stephens, & Pritchard, 2003). In addition, it is very informative to determine whether such genetic clusters are geographically distributed across the species distribution range (Rosenberg, Mahajan, Ramachandran, Zhao, Pritchard, & Feldman, 2005; Novembre et al., 2008). This is because we can interpret the number of genetic clusters, their geographic distribution and their degree of admixture as the result of all demographic processes and adaptive forces acting on populations. This paradigm has steadily gained ground in studies estimating future distribution range shifts due to GCC by means of species distribution models (SDMs; Bálint et al., 2011; Jay et al., 2012; D'Amen, Zimmermann, & Pearman, 2013; Yannic et al., 2014; Gotelli & Stanton-Geddes, 2015; Diniz-Filho et al., 2016; Marcer, Méndez-Vigo, Alonso-Blanco, & Picó, 2016; Ikeda, Max, Allan, Lau, Shuster, & Whitham, 2017; Lima, Ballesteros-Mejia, Lima-Ribeiro, & Collevatti, 2017; Milanesi, Caniglia, Fabbri, Puopolo, Galaverni, & Holderegger, 2018), stressing the need to consider the inherent genetic heterogeneity of organisms.

From a methodological viewpoint, working with intra-specific patterns of genetic diversity implies the combination of presence-only data, which commonly feed SDMs such as `Maxent`, with genetic structure data, which are mostly expressed as genetic cluster membership proportions (ranging between 0 and 1) that inform on the degree of genetic admixture (Serra-Varela, Alía, Ruiz Daniels, Zimmermann, Gonzalo-Jiménez, & Grivet, 2017). The problem arises when admixture information is lost because individuals have to be assigned to a single cluster, generally the one with the highest membership proportion according to some arbitrary threshold in order to run presence-only SDMs (Gotelli & Stanton-Geddes, 2015; Marcer et al., 2016; Ikeda et al., 2017). In doing that, the amount of data and information lost depends on the genetic structure of the study organism. For instance, species

96 with a pronounced genetic structure will likely have individuals with high genetic cluster
97 membership proportions, which facilitates the assignment of individuals to single genetic
98 clusters. In contrast, individual assignment to single genetic clusters will exhibit higher
99 uncertainty for weakly genetically structured organisms (e.g. high levels of individual genetic
100 admixture), posing problems for the development of SDMs to study the effects of GCC on
101 their patterns of geographic genetic structure. Either way, we lose valuable information that
102 may reduce the value and impact of GCC model outcomes and therefore our understanding of
103 the GCC effects on biodiversity.

104 The second biological aspect worth considering when modelling distribution ranges is
105 the presence of spatial autocorrelation (SAC) in data and the problems that SAC poses for
106 statistical and ecological interpretation. SAC can be defined as the dependence between close
107 observations in space (Legendre & Legendre, 2012), and it may be caused by exogenous
108 factors (e.g. historical processes and autocorrelated environmental variables) and/or
109 endogenous factors (e.g. dispersion) (Dale & Fortin, 2014). While variables representing
110 exogenous factors may be readily available to researchers, variables describing endogenous
111 factors representing important biological processes are more difficult to obtain (Belmaker et
112 al., 2015). Overall, SAC is recognised as a major challenge when predicting species
113 distributions (Dirnböck & Dullinger, 2004; de Oliveira, Rangel, Lima-Ribeiro, Terribile, &
114 Diniz-Filho, 2014) because it results in several modelling flaws, such as violation of the
115 assumption of independent error, inflated estimations of model performance, bias in model
116 selection, or inferential problems (Legendre, 1993; Dale & Fortin, 2002; Dormann, 2007;
117 Dormann et al., 2007; Fortin & Dale, 2009; Beale, Lennon, Yearsley, Brewer, & Elston,
118 2010; Swanson, Dobrowski, Finley, Thorne, & Schwartz, 2013; Wagner & Fortin, 2013). To
119 a certain extent, SAC can be dealt with data filtering, although often at a high cost of data

120 loss. For these reasons, taking SAC into account in GCC models is considered as mandatory
121 (Latimer, Wu, Gelfand, & Silander, 2006; Beguin, Martino, Rue, & Cumming, 2012; Record,
122 Fitzpatrick, Finley, Veloz, & Ellison, 2013; Swanson et al. 2013, Crase, Liedloff, Vesk,
123 Fukuda, & Wintle, 2014).
124 Here, we use hierarchical Bayesian models (HBMs), which account for the geographic
125 distribution of intra-specific genetic diversity and the presence of SAC, to analyse current
126 distribution range as well as the effect of GCC on its shifts. To that end, we use a collection of
127 301 natural populations of the annual plant *Arabidopsis thaliana* occurring in the Iberian
128 Peninsula, the region of the distribution range with the highest genomic diversity (The 1001
129 Genome Consortium, 2016). Genome-wide markers are used to infer Iberian *A. thaliana*'s
130 geographic genetic structure by estimating genetic cluster membership proportions. In order
131 to better understand the potential of our model, we compare three approaches, `Maxent` and
132 two Bayesian, representing a gradient of complexity in the treatment of intra-specific genetic
133 data and SAC. In particular, (1) presence-only SDMs (`Maxent`) that do not take SAC into
134 account and based on binary data for genetic cluster membership proportions, (2) non-spatial
135 hierarchical Bayesian models (HBMs) not accounting explicitly for SAC and based on
136 continuous data for genetic cluster membership proportions, and (3) spatially-explicit HBMs
137 accounting for SAC and based on continuous data for genetic cluster membership
138 proportions. Although HBMs represent well-established methods for statistical inference in
139 several research fields, the application of Beta regression to climate-driven shifts in species
140 distribution range is not common (see Martínez-Minaya, Cameletti, Conesa, & Pennino,
141 2018). In particular, we promote the use of Beta regressions where data fitting can be
142 achieved using integrated nested Laplace approximation (INLA) rather than Markov chain
143 Monte Carlo (MCMC) methods. We discuss our results in terms of the relevance of intra-

144 specific genetic variation and SAC to better interpret and contextualise the implications of
145 GCC on species distribution range shifts, but also identifying the limitations and caveats of
146 our approach.
147
148 ## 2. MATERIALS AND METHODS
149 ### 2.1. Source populations and genetic structure
150 We used a collection of 301 natural populations of the annual plant *Arabidopsis thaliana*
151 distributed across the Iberian Peninsula (ca. 800 × 700 km^2; 36.00° N – 43.48° N, 3.19° E –
152 9.30° W; Fig. 1a). This set of populations belongs to a long-term project pursuing a
153 permanent collection of natural populations from the western Mediterranean Basin (Spain,
154 Portugal and North Africa) intended to unravel *A. thaliana*'s evolutionary ecology, functional
155 genetics, and response to GCC (see Marcer et al., 2018, and references therein). Distance
156 among populations and altitudes had a range of 1 – 1,038 km (mean ± SD = 360.9 ± 200.2
157 km) and 1 – 2,662 m.a.s.l (mean ± SD = 786.5 ± 391.3 m.a.s.l.), respectively, including a
158 wide array of wild and humanised environments (Picó, Méndez-Vigo, Martínez-Zapater, &
159 Alonso-Blanco, 2008; Méndez-Vigo, Picó, Ramiro, Martínez-Zapater, & Alonso-Blanco,
160 2011; Manzano-Piedras, Marcer, Alonso-Blanco, & Picó, 2014).
161 Populations included in this study came from field surveys that spanned 12 years (2000 –
162 2011). We sampled seed from several individuals from each population. Every year and a few
163 months after every survey, field-collected seed was multiplied by the single seed descent
164 method in a glasshouse at the Centro Nacional de Biotecnología (CNB-CSIC) in Madrid.
165 Multiplied seeds were stored in dry, dark conditions in cellophane bags at room temperature,
166 storing conditions that can preserve *A. thaliana* seeds for years. In this study, we employed
167 one representative individual (accession hereafter) per population to analyse the genetic

structure of *A. thaliana* in the Iberian Peninsula. Importantly, accessions exhibited common phenotypes within their populations based on flowering time and/or the vernalization requirement for flowering, which are traits under strong selection in Iberian *A. thaliana* (Méndez-Vigo, Gomaa, Alonso-Blanco, & Picó, 2013) that appear to be mediated by variation in temperature (Méndez-Vigo et al., 2011; Vidigal et al., 2016). This procedure increased the odds of using accessions best suited to their local environments and, therefore, common genotypes in the populations.

Nuclear genetic data were obtained from 250 genome-wide single nucleotide polymorphisms (SNPs) previously used to genetically characterise Iberian *A. thaliana* (Picó et al. 2008, Gomaa, Montesinos-Navarro, Alonso-Blanco, & Picó, 2011; Méndez-Vigo et al., 2011; Manzano-Piedras et al., 2014; Marcer et al., 2016). In short, SNPs were selected based on their polymorphism shown in Central Europe, the Iberian Peninsula and in a worldwide collection of accessions, and genotyped using the SNPlex technique (Applied Biosystems, Foster City, CA, USA). These SNPs are located across the genome at putatively neutral regions spaced at 0.5 Mb average intervals (range = 0.11 Kb – 1.82 Mb). All accessions were genetically different from each other.

Genetic structure was assessed using the Bayesian model-based clustering algorithm implemented in `STRUCTURE` v.2.3.3 (Falush et al., 2003), as previously described (Méndez-Vigo et al., 2011; 2013). In brief, model settings included haploid multilocus genotypes, correlated allele frequencies between populations and a linkage model. Each run consisted of 50,000 burn-in MCMC iterations and 100,000 MCMC after-burning repetitions for parameter estimation. To determine the K number of ancestral populations and the ancestry membership proportions of each accession in each population, the algorithm was run 20 times for each defined number of groups (K value) from 1 to 10. The number of distinct genetic groups was

192 determined by testing the differences between the data likelihood for successive K values
193 using Wilcoxon tests for two related samples. The final K number was estimated as the largest
194 K value with significantly higher likelihood than that of K-1 runs (two-sided $P < 0.005$). This
195 was supported by a high similarity among the ancestry membership matrices from different
196 runs of the same K value (H'=0.99). The average symmetric similarity coefficient H' among
197 runs and the average matrix of ancestry membership proportions, derived from the 20 runs,
198 were calculated with `CLUMPP` v.1 (Jakobsson & Rosenberg, 2007). This analysis inferred four
199 genetic clusters in the Iberian Peninsula (Fig. 1), in agreement with previous studies on *A.*
200 *thaliana*'s genetic structure (Picó et al., 2008; Gomaa et al., 2011; Méndez-Vigo et al., 2011;
201 Manzano-Piedras et al., 2014; Marcer et al., 2016).
202
203 **2.2. Climatic variables and GCC scenarios**
204 We selected a total of eight bioclimatic predictors to define the climatic space: BIO1 (annual
205 mean temperature), BIO2 (mean diurnal range), BIO3 (isothermality), BIO4 (temperature
206 seasonality), BIO8 (mean temperature of the wettest quarter), BIO12 (annual precipitation),
207 BIO15 (precipitation seasonality) and BIO18 (precipitation of the warmest quarter). These
208 bioclimatic predictors were selected because their pairwise correlation coefficients were less
209 than 0.7, a threshold value commonly used to avoid unacceptable co-linearity among
210 independent variables (Pino, Font, Carbó, Jové, & Pallarès, 2005). We used the `dismo` R
211 package (Hijmans, Phillips, Leathwick, & Elith, 2017) to retrieve these climate layers from
212 the Digital Atlas of the Iberian Peninsula (http://www.opengis.uab.cat/wms/iberia/), which
213 provides interpolated surface layers of mean monthly data obtained from 2,285 weather
214 stations for the period 1951 – 1999. We refer to this time period as year 2000.

We chose the year 2070 as the scenario to evaluate *A. thaliana*'s distribution range shifts due to GCC. In order to use the most and least conservative GCC scenarios, we selected the representative concentration pathways (RCP) 2.6 and 8.5 (van Vuuren et al., 2011), respectively. In addition, we averaged four climate models: HadGEM2-ES (Met Office Hadley Centre, UK), MRI-CGCM3 (Meteorological Research Institute, Japan), MIROC-ESM (Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, The University of Tokyo and National Institute for Environmental Studies, Japan), and NorESM1-M (Norwegian Climate Centre, Norway). Data for 2070 were downloaded from the WorldClim Global Climate Database v.1.4 (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). The resolution of the climatic spaces for the years 2000 and 2070 was 1 km.

2.3. Species distribution models (SDMs)

SDMs link information on the presence/absence or abundance of a species to environmental variables to predict where it is likely to be present in un-sampled locations or time periods (Guisan & Thuiller, 2005; Elith & Leathwick, 2009). In the last years, the quantity and the quality of the datasets have substantially increased, resulting in a higher complexity of the statistical issues that have to be addressed when an SDM is created. As a result of this increasing complexity, the performance of the SDM inferential and predictive processes are becoming more challenging, forcing researchers to develop new sophisticated statistical techniques (see a review in Martínez-Minaya et al., 2018). Given the flood of methodologies developed to address this issue, we compared SDMs obtained with two contrasting approaches: a presence-only model (`Maxent`) and the hierarchical Bayesian Beta regressions (with and without a spatial term).

2.3.1. Maxent

We used Maxent v.3.3.3k (Phillips, Anderson, & Schapire, 2006; Elith et al., 2011) to model the current and future distribution of *A. thaliana*'s genetic clusters. As Maxent uses presence-only data, we assigned each of the 301 accessions to its predominant genetic cluster using a cut-off value of 0.5 to each genetic cluster membership proportion given by STRUCTURE (as in Marcer et al., 2016). As a result, the number of accessions per genetic cluster was reduced resulting in a low of 35 for genetic cluster C4 to a high of 103 accessions for genetic cluster C1 (Fig. 1b). The mean (± SE) membership proportions to each genetic cluster were 0.66 ± 0.01 (range = 0.51 – 0.92) for genetic cluster C1, 0.74 ± 0.02 (range = 0.52 – 0.96) for genetic cluster C2, 0.89 ± 0.02 (range = 0.56 – 0.97) for genetic cluster C3, and 0.77 ± 0.02 (range = 0.51 – 0.94) for genetic cluster C4. Eighty-two accessions (27.2%) did not have any genetic cluster membership proportion higher than 0.5 and could not be included in the Maxent models, stressing one of the limitations of this approach when dealing with accessions with high genetic admixture. We fitted all possible models determined by the set of combinations between the eight climatic predictors without considering interactions. We then ranked these models according to the five-fold cross-validated area-under-the-curve (AUC) metric and chose the most parsimonious one among the best five (Table S1). Then, we ran again the chosen model with all data points to obtain the final model. Maxent was used with default parameters with the exception of features, which were limited to the hinge type, making it similar to a Generalized Additive Model (Elith et al., 2011).

2.3.2. Hierarchical Bayesian Beta regression

Spatial and non-spatial HBMs were also used to model the current and future distribution of *A. thaliana*'s genetic clusters. In particular, spatial and non-spatial Beta regressions were conducted to estimate the genetic cluster membership probability, which in this particular context, can be thought of as the habitat suitability for each genetic cluster. In contrast to `Maxent`, Beta regressions allowed us to model each genetic cluster separately using all genetic information available, that is, the genetic cluster membership proportions of all 301 accessions. In other words, no data were excluded.

The class of Beta regression models is commonly used to model variables that assume values in the unit interval (between 0 and 1; Ferrari & Cribari-Neto, 2004), such as the case of membership probabilities of genetic clusters. A Beta distribution depends on two scaling parameters, $Beta(a, b)$, which can be parameterised in terms of its mean, μ, a dispersion parameter, $\phi = a + b$, and the variance, $\sigma^2 = \frac{\mu(1-\mu)}{1+\phi}$. This parameterisation better supports the truncated nature of the Beta distribution because the variance depends on the mean, which translates into maximum variance at the centre of the distribution and minimum variance at the edges. In addition, the dispersion of the distribution for a fixed μ decreases as ϕ increases. We did not transform the data to avoid the problems posed by extreme values, as proposed elsewhere (Cribari-Neto & Zeileis, 2010), because data fell far enough from the extremes of the Beta distribution.

As we were interested in depicting the relationship between the genetic cluster membership probabilities and the bioclimatic predictors, we linked the mean and the precision of the response variable to the linear bioclimatic predictors via suitable link functions. In particular, if Y_i represents the genetic cluster membership probability at location *i*, then its conditional distribution is $Y_i \mid \mu_i, \phi_i \sim Beta(\mu_i, \phi_i)$, where μ_i and ϕ_i are the Beta distribution parameters at location i. We used the *logit* and *log* links for μ_i and ϕ_i, respectively. The mean

was linked to climatic covariates (non-spatial term) and, in the case of spatial models, to a stochastic spatial effect (spatial term). The precision was assumed to be not dependent of any effect. The resulting model with a spatial term is known as a point-referenced spatial Beta regression (Paradinas et al., 2016; 2018). It is highly suitable for situations in which data are observed at continuous locations occurring within a defined spatial domain:

$$logit(\mu_i) = X_i\beta + W_i,$$

$$log(\phi_i) = \theta,$$

where β is the vector of regression coefficients ($\beta_0, \beta_1, \ldots, \beta_c$), X_i is the vector corresponding to the *i*th row of the design matrix whose first element is 1 (the one multiplying the intercept β_0), the covariate values at location *i* being the remaining elements, and W_i is the spatially structured random effect at each location *i*. W is assumed to be a multivariate Gaussian distribution whose covariance matrix, $\sigma_W^2 H(\varphi)$, depends on the distance between locations, and its parameters, σ_W^2 and φ, represent the variance and range of the spatial effect, respectively.

In the context of HBMs, parameters were treated as random variables and prior knowledge was incorporated using the corresponding prior distributions. These priors were specified in the second stage jointly with random effects. In the third and final level of the hierarchy, prior knowledge about the hyper-parameters was expressed. This hierarchical structure can also be considered as a latent Gaussian model (Rue & Held, 2005).

As posterior distributions for the parameters and hyper-parameters do not have an analytic expression, numerical approximations are usually needed. In the case of latent Gaussian models, integrated nested Laplace approximation (INLA) (Rue, Martino, & Chopin, 2009) is a computationally efficient alternative to the MCMC method. However, to fit and predict the particular case of continuously indexed Gaussian fields with INLA, W in our case,

an additional module is required. Lindgren, Rue & Lindström (2011) proposed an approach using an approximate stochastic weak solution to a stochastic partial differential equation (SPDE) as a Gaussian Markov random field (GMRF) approximation to continuous Gaussian fields with Matérn covariance structure, a highly flexible and general family of functions in spatial statistics (Rue & Held, 2005). The Markov property allowed the use of a precision sparse matrix, enabling efficient numerical algorithms. Under this approximation, the spatial effect is re-parameterised as follows:

$$W \sim N\big(0, Q(\kappa, \tau)\big).$$

Here, W depends on two different parameters, κ and τ, which determine the range of the effect and the total variance, respectively. More precisely, the range is approximately $\varphi = \sqrt{\frac{8}{\kappa}}$ and the variance is $\sigma_W^2 = \frac{1}{4\pi\kappa^2\tau^2}$ (Lindgren et al., 2011).

We specified prior distributions for the parameters and hyper-parameters. In particular, normal vague priors with mean 0 and precision 10^{-4} were used for the vector of regression coefficients. Although internally INLA works with parameters κ and τ, we specified the spatial effect in terms of φ and σ_W using the re-parameterisations $log(\varphi)$ and $log(\sigma_W)$ as independent normal vague distributions (Lindgren & Rue, 2015).

Overall, the full model was stated as follows:

$$Y_i \mid \mu_i, \phi_i \sim Beta(\mu_i, \phi_i)$$

$$logit(\mu_i) = X_i\beta + W_i$$

$$log(\phi_i) = \theta$$

$$\beta_0, \beta_1, \ldots, \beta_c \sim N(0, 10^{-4}), W \sim N\big(0, Q(\varphi, \sigma_W)\big)$$

$$log(\varphi) \sim N\big(m_\varphi, q_\varphi\big)$$

$$log(\sigma_W) \sim N\big(m_{\sigma_W}, q_{\sigma_W}\big)$$

$$\theta \sim logGamma(0,0.1),$$

where m_{φ} was automatically chosen so that the prior mean of φ was about 50% the diameter of the study geographic region, while m_{σ_W} was chosen in a way that the corresponding variance of the field was 1. For our analysis, this resulted in $m_{\varphi} = 13.517$ and $m_{\sigma_W} = 0$. Finally, the default *a priori* precisions for $log(\varphi)$ and $log(\sigma_w)$ distributions were $q_{\varphi} = 0.25$ and $q_{\sigma_W} = 0.25$, respectively.

These latter values, q_{φ} and q_{σ_W}, express the large uncertainty about the parameters before the analysis, resulting in quite non-informative hyper-priors. This is important because it allows the range to take values between 0 and the total diameter of the Iberian Peninsula. In contrast to `Maxent`, HBMs can take space into account when modelling distribution ranges, which gives the possibility to evaluate its mean effects and uncertainty. As mentioned above, once the inference is done, the main interest becomes to predict the response in un-sampled locations. To do that, we applied the SPDE by constructing a Delaunay triangulation (Hjelle & Daehlen, 2006) covering the whole Iberian Peninsula (Fig. S1).

2.3.3. Model selection, distribution range shifts and residual SAC

HBMs were run with and without the spatial component with the `R-INLA` package (Lindgren & Rue, 2015) in order to quantify its effects on distribution range shifts with GCC and to be compared with `Maxent` outcomes. We fitted all possible models given by the set of combinations among the eight climatic predictors without interactions. To select the best model, we used $LCPO = \frac{-1}{N}\sum_i^N log\ (CPO_i)$ as a summary statistic of the conditional predictive ordinate (CPO; Geisser, 1993), which gives an overall measure of predictive performance (Hooten & Hobbs, 2015). CPO is defined as the cross-validated predictive density at a given observation. CPO can be used to compute predictive measures, such as the

357 logarithmic score (Gneiting & Raftery, 2007) or the cross-validated mean Brier score
358 (Schmid, 2005). Among the best five models for each genetic cluster we selected the most
359 parsimonious one, that is, the one with the least number of predictors. Model quality
360 estimators, such as the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, &
361 Van Der Linde, 2002) and the Watanabe–Akaike information criterion (WAIC; Watanabe,
362 2010) were also computed. We also measured accuracy of spatial and non-spatial HBMs by
363 means of mean absolute error (MAE) and root mean squared error (RMSE). Lower values of
364 MAE and RMSE indicate better accuracy. The comparison between `Maxent` and HBMs in
365 terms of accuracy can be misleading because `Maxent` used sub-samples of data for each
366 genetic cluster whereas HBMs were always based on the entire data set (Fig. 1).
367 We compared distribution range shifts due to GCC when taking the spatial component
368 into account (spatial HBM), when excluding the spatial component (non-spatial HBM), and
369 `Maxent`. We added the probabilities calculated across the whole study area by each model
370 for each time frame and GCC scenario to quantify the geographic extent of the suitability of
371 each genetic cluster for each methodology. These probabilities were used to calculate the
372 percentage loss or gain of suitability for each model and GCC scenario.
373 All models mentioned above were checked for residual SAC. We calculated the residuals
374 for model predictions between observed and predicted values and tested for residual SAC
375 using the `spdep` R package (Bivand, Hauke, & Kossowski, 2013; Bivand & Piras, 2015). In
376 order to calculate residual SAC in `Maxent`, we followed the methodology used elsewhere
377 (De Marco, Diniz-Filho, & Bini, 2008; Václavík & Meentemeyer, 2009). Basically, we
378 estimated the Moran's *I* coefficient of autocorrelation with 10,000 MCMC iterations. Models
379 with *P*-values > 0.05 were considered as SAC free. As expected, spatial HBM did not show

residual SAC, while non-spatial HBM did. All Maxent models, except for genetic cluster C4, also retained residual SAC (Table S2).

3. RESULTS

3.1. Current distribution range

We compared the performance of three modelling approaches, Maxent and HBMs with and without the spatial component, to depict the current distribution range of four *A. thaliana*'s genetic clusters using eight selected bioclimatic predictors. Maxent models included between four and seven bioclimatic predictors per genetic cluster (Table 1). With these predictors, Maxent yielded a clear geographic distribution of genetic cluster ranges in the Iberian Peninsula (Fig. 2a), as found in earlier studies using the same approach but with different data (Marcer et al., 2016).

In the case of Bayesian models, spatial and non-spatial HBMs produced broadly similar geographic distributions of genetic clusters to those generated by Maxent models, particularly for genetic clusters C1 and C4 (Fig. 2a). In the case of genetic cluster C2, the spatial HBM depicted a rather continuous distribution in NE Spain, which clearly differed from those given by Maxent and non-spatial HBM showing the truncated distribution that this genetic cluster actually has in the wild (Fig. 2a). In general, spatial HBMs and Maxent models showed more compact distribution ranges than non-spatial HBMs, the former with more dramatic transitions between low and high probability values in all genetic clusters (Figs. 2a and S2). The exception was genetic cluster C3, whose predicted distribution range with non-spatial HBMs was rather blurred in comparison with that obtained with Maxent (Fig. 2a). In fact, it was not possible to fit spatial HBMs for genetic cluster C3 because the results were inconsistent. When using vague hyper-priors for the range of the spatial effect,

404 the resulting mean of the posterior distribution of the range was larger than the whole study
405 area. On the contrary, when using more informative priors, results were different and very
406 much conditioned by prior selections. Thus, the spatial effect for genetic cluster C3 did not
407 provide further explanation than what can already be explained by the bioclimatic predictors.
408 For all genetic clusters, the uncertainty of the predictive mean for non-spatial HBMs was
409 lower and more evenly distributed across space than that for spatial HBMs (Fig. 2b). The
410 main reason for this apparent reduction in uncertainty is that spatial models are reflecting the
411 intrinsic variability of the Beta-distributed data, variability that is not reflected by non-spatial
412 models. As a result, the distribution of means and standard deviations for spatial HBMs was
413 more pronounced than those for non-spatial HBMs (Fig. 2 and Table S3). Nonetheless, the
414 values of the mean of the posterior distribution of the precision parameter, which are
415 inversely proportional to the variance of the data, were larger in the spatial HBMs, reflecting
416 their acceptable behaviour (Table S4). In the case of spatial HBMs, these models allowed the
417 visualisation of the spatial effects, which clearly were more intense at the centre of the genetic
418 cluster distribution ranges (Fig. 3). Uncertainty of the mean of the spatial effect was greater
419 for genetic cluster C4 than for the other two genetic clusters (Fig. 3). Overall, spatial HBMs
420 selected less bioclimatic predictors than non-spatial HBMs and `Maxent` models to define the
421 distribution range of the four genetic clusters (Tables 1 and S5), particularly for genetic
422 cluster C4. Finally, the combination of bioclimatic predictors used by `Maxent` models and
423 spatial and non-spatial HBMs was quite different (Table 1).
424 Mean absolute error (MAE) and root mean squared error (RMSE) were lower for spatial
425 than for non-spatial HBMs for all genetic clusters in which the comparison was possible
426 (Table 2). This indicated that spatial HBMs had lower average model prediction errors in the
427 response variable.

3.2. Distribution range shifts with GCC

`Maxent` models and HBMs were also used to quantify distribution range shifts of *A. thaliana*'s genetic clusters under different GCC models and scenarios. The three modelling approaches yielded different GCC predictions for each genetic cluster based on suitability shifts in distribution range projections (Table 3, Fig. 4 and S3). Overall, `Maxent` showed a trend of predicting more dramatic changes in distribution range due to GCC for all genetic clusters compared to spatial and non-spatial HBMs (Table 3 and Fig. 4). For genetic cluster C1, important reductions in distribution range were predicted for the two GCC scenarios with `Maxent` and non-spatial HBMs, whereas spatial HBMs predicted slight increases (Table 3 and Fig. 4). For genetic cluster C2, `Maxent` predicted increasing and decreasing distribution ranges with RCP 2.6 and RCP 8.5, respectively, whereas both HBMs predicted small fluctuations in distribution range in both GCC scenarios (Table 3 and Fig. 4). For genetic cluster C3, `Maxent` showed very large increases in distribution range, particularly for the RCP 8.5 scenario, whilst non-spatial HBMs predicted slight fluctuations in distribution range in both GCC scenarios (Table 3 and Fig. 4). Finally, for genetic cluster C4, all approaches predicted increases in distribution range in both GCC scenarios. `Maxent` gave higher increases in RCP 2.6 than in RCP 8.5 and vice-versa for both HBMs (Table 3 and Fig. 4).

4. DISCUSSION

Distribution range shifts represent the most important effect of GCC on biodiversity because of their ecological implications and the potentially detrimental socio-economic impact on society. GCC models for distribution range shifts have to increase their sophistication by adding realism to the model outcomes, yet without losing interpretability or increasing

452 uncertainty. In this study, we address this issue by developing hierarchical Bayesian models
453 (HBMs) for the annual plant *Arabidopsis thaliana* incorporating two of these elements, which
454 are inherent to practically all organisms: the geographic distribution of intra-specific genetic
455 variation and the spatial autocorrelation in data. Importantly, both geographic genetic
456 structure and spatial autocorrelation can be considered as indicators of eco-evolutionary
457 forces shaping species' distribution range, such as colonization/extinction dynamics, dispersal
458 ability, local adaptation and historical factors.
459
460 **4.1. Current distribution range**
461 The selection of the modelling approach may have significant repercussions when considering
462 a species as a genetically heterogeneous organism, whose geographic distribution of its
463 genetic variation may have major implications for understanding the effects of GCC on its
464 distribution range. In the case of `Maxent`, and of any modelling technique dealing with
465 binary response variables, a major problem is the loss of data resulting from the conversion of
466 a continuous variable (i.e. the genetic cluster membership proportions characterising each
467 individual) into a binary variable (i.e. the assignment of each individual to the genetic cluster
468 with the highest membership proportion) (Gotelli & Stanton-Geddes, 2015; Marcer et al.,
469 2016). In our study, binarization of genetic data had an important cost in terms of data loss as
470 82 of 301 accessions did not reach the minimum membership proportion of 0.5 required to be
471 assigned to a genetic cluster. As a result, the number of accessions per genetic cluster used in
472 `Maxent` was reduced (Fig. 1b). In contrast, HBMs did not have this limitation and used the
473 whole set of 301 accessions also including genetic admixture among the four genetic clusters
474 detected in the Iberian Peninsula. It must be emphasised that accessions exhibiting genetic
475 admixture are quite relevant in biological terms. For example, they may indicate the existence

476 of contact zones between genetic lineages where important processes affecting the distribution
477 range may take place, such as the balance between selection and dispersal against hybrids
478 (Barton & Hewitt, 1985). For this reason, HBMs represent a better choice to model
479 distribution ranges of intra-specific genetic lineages if it is undesirable to discard accessions
480 with too much genetic admixture.

481 Broadly speaking, `Maxent` and the Bayesian modelling approaches were consistent in
482 depicting the current geographic distribution of the four genetic clusters of Iberian *A. thaliana*
483 (Figs 1 and 2). The exception was the genetic cluster C3, in which non-spatial HBM blurred
484 the distribution range and spatial HBM was not able to produce results due to unacceptable
485 outcomes in a Bayesian framework. Interestingly, the genetic cluster C3 is strongly
486 differentiated from the rest of clusters found in the Iberian Peninsula, as well as across the
487 whole species' distribution range. In fact, the genetic cluster C3 is considered as the relict
488 cluster with a long evolutionary history (Picó et al., 2008; Brennan et al., 2014; The 1001
489 Genomes Consortium, 2016; Durvasula et al., 2017). The relict nature of the genetic cluster
490 C3 is also supported by its scattered distribution across the Iberian Peninsula, a geographic
491 distribution that is interpreted as the result of Iberian glacial refugia (Picó et al., 2008;
492 Brennan et al., 2014; Marcer et al., 2016), whereas the rest of the genetic clusters exhibit
493 geographically marked distributions, likely as a result of more recent demographic histories.
494 Overall, this result indicates that modelling the distribution range of genetic clusters or
495 species with scattered distributions may be difficult no matter what modelling approach is
496 applied. For the particular case of genetic cluster C3, characterised by the high genetic
497 membership proportions of their accessions and the relatively low admixture with other
498 genetic clusters, `Maxent` predicts its distribution best.

For the rest of genetic clusters with marked geographic distributions (NW, NE and SW Iberian Peninsula for genetic clusters C1, C2 and C4, respectively), `Maxent` and Bayesian approaches were able to model their current distribution ranges. However, they exhibited some differences among genetic clusters. For example, genetic cluster C2 exhibits a disjunct distribution due to a major geographic barrier (i.e. the Ebro river valley in NE Spain), which was clearly depicted by `Maxent` (Fig. 2). It is worth noting that such disjunt distribution is not a sampling problem, but the result of the low occurrence of the species confirmed after repeated field campaigns in the region. Bayesian Beta regression approaches, particularly the spatial HBM, blurred, albeit not totally erasing, the disjunct distribution of this genetic cluster. In contrast, `Maxent` and Bayesian Beta regression approaches were more consistent for genetic clusters C1 and C4, which exhibited more compact distributions. Overall, we conclude that the continuity of clusters' distribution range increases its suitability to be modelled by alternative means.

Spatial HBMs, along with `Maxent` for genetic cluster C4, did not show residual spatial autocorrelation, which is a desirable property to avoid inaccurate parameter estimates and inadequate quantification of uncertainty (Latimer et al., 2006; Beguin et al., 2012; Record et al., 2013; Crase et al., 2014). In addition, spatial HBMs exhibited lower average model prediction errors than non-spatial HBMs. Hence, and from a purely statistical viewpoint, the higher rigour of spatial HBMs, in terms of higher accuracy and efficient removal of residual spatial autocorrelation, confers them a clear advantage (Swanson et al., 2013; Crase et al., 2014). Spatial HBMs also allowed the assessment of the spatial effects on distribution range, which were quite compact and with high intensities at the distribution range centre (Fig. 3). Such patterns may account for the lower number of bioclimatic predictors selected by spatial HBMs in comparison with non-spatial HBMs and `Maxent`. As a matter of fact, the reduction

of predictors represents a common shortcoming of spatial distribution models that, in some cases, may jeopardise the biological interpretation of the environmental factors underlying current distribution ranges (Beale et al., 2010; Swanson et al., 2013). We want to emphasise, however, that the five best spatial HBMs for each genetic cluster included additional variables compared to the best model, and all models were quite similar in terms of DIC, WAIC and LCPO values (Table S5). Thus, we have different options to identify environmental drivers of current distribution ranges. Furthermore, the reduction of predictors in spatial models may not reduce the models' interpretability.

4.2 Distribution range shifts with GCC

Taking spatial effects into account had a profound effect on the predictions of distribution range shifts due to GCC for *A. thaliana*'s genetic clusters in the Iberian Peninsula. In general, spatial HBMs exhibited more conservative patterns of change compared to `Maxent` and non-spatial HBMs (Fig. 4). This result is in agreement with other research suggesting that environment-only models forecast substantially greater range shifts compared to models incorporating spatial effects (Swanson et al., 2013; Crase et al., 2014). The rationale is that organisms exhibiting a high spatial autocorrelation in the environmental drivers accounting for their distribution ranges will have larger areas with similar climates, which will also make GCC effects more predictable and homogeneous across space (Nadeau, Urban, & Bridle, 2017). For this reason, genetic clusters or species with continuous distributions not only increase the ease of modelling, but also facilitate the assessment of the spatial autocorrelation on distribution range shifts. Overall, there is no denying that considering spatial autocorrelation adds realistic biological elements for understanding the long-term effects of GCC on biodiversity (De Marco et al., 2008; Swanson et al., 2013; Crase et al., 2014;

547 Cardador, Sardà-Palomera, Carrete, & Mañosa, 2014). Nevertheless, it must be noted that we
548 are assuming that spatial autocorrelation and its underlying forces remain relatively constant
549 during climate change. Clearly, this assumption, although beyond the scope of this study, will
550 need to be addressed in the future.
551 In general, the outcomes generated by the three modelling approaches for GCC scenarios
552 were quite different. Such a disparity in model outcomes may indicate the differential effects
553 of environmental drivers and the sources of spatial autocorrelation on the GCC response of
554 genetic clusters, but also the effect of geographic distribution of each genetic cluster on model
555 performance. In fact, the problems affecting the modelling of current distribution ranges,
556 namely the disjunct and scattered geographic distributions of genetic clusters C2 and C3,
557 respectively, also affected the predictions of distribution range shifts with GCC. For example,
558 in the case of genetic cluster C2, spatial HBM predicted a rather continuous distribution in NE
559 Spain when it is more reasonable to expect that the barrier separating the two major nuclei of
560 populations both sides of the Ebro river valley will be expanded with warming, as predicted
561 by `Maxent`. Furthermore, the GCC effects on genetic cluster C3 are the tougher to predict.
562 Although `Maxent` increased the potential distribution range of this genetic cluster over the
563 Iberian Peninsula, as relict organisms exhibiting scattered distributions, the future of the C3
564 cluster may simply depend on the effect of GCC on the preservation of its habitats as they are
565 today. In fact, populations of genetic cluster C3 may exhibit strong local adaptation (Méndez-
566 Vigo et al. 2013), constraining the ability of relict genotypes to colonise novel habitats.
567 In contrast, interpreting the problems of the GCC effects on distribution range shifts for
568 the other two genetic clusters with continuous distributions was totally different. For example,
569 genetic cluster C1 has a continuous distribution across the NE Iberian Peninsula, which is
570 characterised by Atlantic and continental climates. GCC models excluding spatial effects, i.e.

571 `Maxent` and non-spatial HBMs, indicate that GCC will restrict *A. thaliana* to northern and
572 mountainous areas, which is a typical scenario of migration towards environments that will
573 probably retain similar characteristics in the future. In contrast, spatial HBMs yielded a totally
574 different outcome, indicating that genetic cluster C1 will maintain and even increase its
575 current distribution range (Fig. 4). The strong spatial effects detected by spatial HBMs for
576 genetic cluster C1 may account for this result, as the response of *A. thaliana* to GCC is also
577 expected to be more homogeneous. Recent experimental data from transplant experiments
578 using accessions from this genetic cluster indicate that this scenario may be plausible, as *A.*
579 *thaliana* performed well in warmer environments, highlighting the potential of this genetic
580 cluster to cope with warming (Exposito-Alonso, Brennan, Alonso-Blanco, & Picó, 2018). The
581 same applies to genetic cluster C4, which is also distributed continuously in the typically
582 Mediterranean SW Iberian Peninsula. In this case, however, all modelling approaches
583 predicted its expansions with GCC, although some discrepancies among modelling
584 approaches were recorded (Fig. 4). Although we lack experimental evidence of the effects of
585 warming on performance of C4 *A.thaliana* accessions, we believe that such expansion with
586 GCC is highly probable, as the genetic cluster C4 mostly occupies the warmest Iberian region.
587
588 **4.3 Conclusions**
589 We developed hierarchical Bayesian Beta regression models to explore the current
590 distribution range and its GCC-induced shifts for an organism with a marked geographic
591 genetic structure, which represents the outcome of historical, ecological and evolutionary
592 forces probably acting in concert. For this reason, the effects of GCC have to be understood as
593 a mosaic of responses varying in extent and intensity determined by the complexity of the
594 geographic genetic structure exhibited by study organisms. Rather than predicting mere

595 contractions or expansions for a single organism, we should expect a reshuffling of the
596 genetic diversity and its geographic structure with GCC, which is obviously more difficult to
597 predict. The HBMs developed here enrich the toolbox of software available to deal with such
598 expectation.

599 From a statistical viewpoint, our HBMs allow the modelling of each genetic cluster and
600 avoid the binarization of genetic cluster membership proportions, required by `Maxent`, which
601 may imply an important data loss. This has the advantage that populations with high
602 admixture can be included in HBMs. In addition, our HBMs can take the spatial
603 autocorrelation of data into account, which not only improves the statistical properties of the
604 model (i.e. removal of residual SAC) but also adds realism, as spatial autocorrelation may
605 represent the result of eco-evolutionary processes shaping distribution ranges. Despite such
606 desirable properties, our simulations of current and future distribution ranges of the four
607 genetic clusters of Iberian *A. thaliana* indicated that the ease of modelling is strongly related
608 to the continuity of their distributions. Furthermore, the biological knowledge of the study
609 organism (e.g. the identification of relict genetic lineages based on whole-genome markers,
610 the detection of void areas after years of extensive field sampling, and the experimental
611 quantification of plant performance with warming) emerges as an essential element in the
612 understanding of model outputs. Finally, we believe that further work should also be
613 conducted to validate model outputs by independent means (i.e. the assignment of new *A.*
614 *thaliana* populations to genetic clusters based on model predictions).

615 We conclude by stressing the importance of developing better models to forecast the
616 effects of GCC on organisms' distribution range worldwide. Such predictive tools, and the
617 comparison thereof, may lead to the mitigation of the inevitable impact of GCC on
618 biodiversity. However, we have to keep increasing our comprehension of the evolutionary

(e.g. physiological adaptive responses) and demographic (e.g. extinction/colonization dynamics and dispersal ability) factors accounting for the response of organisms to environmental changes imposed by GCC.

ACKNOWLEDGEMENTS

J.M.-M. thanks Generalitat Valenciana and the European Social Fund (ESF) for support via VALi+d grant ACIF/2016/455 (ESF invests in your future). D.C. was funded by grant MTM2016-77501-P (AEI/FEDER, UE). F.X.P. and A.M. were funded by grant CGL2016-77720-P (AEI/FEDER, UE). A.M. and M.J.F. acknowledge the NEWFORESTS project (PIRSES-GA-2013-612645) from the European 7FP. A.M. also acknowledges the grant 2017-SGR-1006 from the Agency for Management of University and Research Grants of the Generalitat de Catalunya. The authors thank the Subject Editor, Frédéric Austerlitz, and seven anonymous reviewers for their useful comments on earlier versions of this manuscript.

DATA ACCESSIBILITY

The data that support the findings of this study are openly available in "Zenodo" at http://doi.org/10.5281/zenodo.2552025, reference number 2552025.

AUTHOR CONTRIBUTION

F.X.P. and A.M. conceived and designed the study. C.A.-B. and F.X.P. collected the samples and analyzed genetic data. J.M.-M., D.C. M.-J.F., and A.M. developed models and performed simulations. F.X.P. led the writing and all authors made significant contributions to its final version through several revisions.

643 ORCID ID: Joaquín Martínez-Minaya (0000-0002-1016-8734), David Conesa (0000-0002-
644 5442-5691), Marie-Josée Fortin (0000-0002-9935-1366), Carlos Alonso-Blanco (0000-0002-
645 4738-5556), F. Xavier Picó (0000-0003-2849-4922), Arnald Marcer (0000-0002-6532-7712).
646

647 # REFERENCES

648 Anderson, C. D., Epperson, B. K., Fortin, M.-J., Holderegger, R., James, P. M. A.,
649 Rosenberg, M. S., ... Spear, S. (2010). Considering spatial and temporal scale in landscape-
650 genetic studies of gene flow. *Molecular Ecology*, 19, 3565–3575. doi: 10.1111/j.1365-
651 294X.2010.04757.x
652 Bálint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., ... Nowak, C.
653 (2011). Cryptic biodiversity loss linked to global climate change. *Nature Climate Change*,
654 1, 313–318. doi: 10.1038/NCLIMATE1191
655 Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology*
656 *and Systematics*, 16, 113–148. https://www.jstor.org/stable/2097045
657 Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010).
658 Regression analysis of spatial data. *Ecology Letters*, 13, 246–264. doi: 10.1111/j.1461-
659 0248.2009.01422.x
660 Beguin, J., Martino, S., Rue, H., & Cumming, S. G. (2012). Hierarchical analysis of spatially
661 autocorrelated ecological data using integrated nested Laplace approximation. *Methods in*
662 *Ecology and Evolution*, 3, 921–929. doi: 10.1111/j.2041-210X.2012.00211.x
663 Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of
664 climate change on the future of biodiversity. *Ecology Letters*, 15, 365–377. doi:
665 10.1111/j.1461-0248.2011.01736.x

666 Belmaker, J., Zarnetske, P., Tuanmu, M. -N., Zonneveld, S., Record, S., Strecker, A., &
667 Beaudrot, L. (2015). Empirical evidence for the scale dependence of biotic interactions.
668 *Global Ecology and Biogeography*, 24, 750–761. doi: 10.1111/geb.12311
669 Bivand, R. S., & Piras, G. (2015). Comparing implementations of estimation methods for
670 spatial econometrics. *Journal of Statistical Software*, 63, 18. URL:
671 https://www.jstatsoft.org/article/view/v063i18
672 Bivand, R. S., Hauke, J., & Kossowski, T. (2013). Computing the Jacobian in Gaussian
673 spatial autoregressive models: An illustrated comparison of available methods.
674 *Geographical Analysis*, 45, 150–179. doi: 10.1111/gean.12008
675 Brennan, A. C., Méndez-Vigo, B., Haddioui, A., Martínez-Zapater, J. M., Picó, F. X., &
676 Alonso-Blanco, C. (2014). The genetic structure of *Arabidopsis thaliana* in the south-
677 western Mediterranean range reveals a shared history between North Africa and southern
678 Europe. *BMC Plant Biology*, 14, 17. doi:10.1186/1471-2229-14-17
679 Cardador, L., Sardà-Palomera, F., Carrete, M., & Mañosa, S. (2014). Incorporating spatial
680 constraints in different periods of the annual cycle improves species distribution model
681 performance for a highly mobile bird species. *Diversity and Distributions*, 20, 515–528.
682 DOI: 10.1111/ddi.12156
683 Chen, I. -C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range
684 shifts of species associated with high levels of climate warming. *Science*, 333, 1024–1026.
685 doi: 10.1126/science.1206432
686 Crase, B., Liedloff, A., Vesk, P. A., Fukuda, Y., & Brendan, A. (2014). Incorporating spatial
687 autocorrelation into species distribution models alters forecasts of climate-mediated range
688 shifts. *Global Change Biology*, 20, 2566–2579. doi: 10.1111/gcb.12598
689 Cribari-Neto, F. & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*,

690 34, 1–24. doi: 10.18637/jss.v034.i02.
691 D'Amen, M., Zimmermann, N. E., & Pearman, P. B. (2013). Conservation of
692 phylogeographic lineages under climate change. *Global Ecology and Biogeography*, 22,
693 93–104. doi: 10.1111/j.1466-8238.2012.00774.x
694 Dale, M. R. T., & Fortin, M. -J. (2002). Spatial autocorrelation and statistical tests in ecology.
695 *Ecoscience*, 9, 162–167. doi: 10.1080/11956860.2002.11682702
696 Dale, M. R. T., Fortin, M. -J. (2014). *Spatial analysis – A guide for ecologists* (2nd ed.),
697 Cambridge: Cambridge University Press.
698 De Marco Jr., D., Diniz-Filho, J. A. F., & Bini, L. M. (2008). Spatial analysis improves
699 species distribution modelling during range expansion. *Biology Letters*, 4, 577–580.
700 doi:10.1098/rsbl.2008.0210
701 de Oliveira, G., Rangel, T. F., Lima-Ribeiro, M. S., Terribile, L. C., & Diniz-Filho, J. A. F.
702 (2014). Evaluating, partitioning, and mapping the spatial autocorrelation component in
703 ecological niche modeling: a new approach based on environmentally equidistant records.
704 *Ecography*, 37, 637–647. doi: 10.1111/j.1600-0587.2013.00564.x
705 Diniz-Filho, J. A. F., Barbosa, A. C. O. F., Collevatti, R. G., Chaves, L. J., Terribile, L. C.,
706 Lima-Ribeiro, M. S., & Telles, M. P. C. (2016). Spatial autocorrelation analysis and
707 ecological niche modelling allows inference of range dynamics driving the population
708 genetic structure of a Neotropical savanna tree. *Journal of Biogeography*, 43, 167–177.
709 doi:10.1111/jbi.12622
710 Dirnböck, T., & Dullinger, S. (2004). Habitat distribution models, spatial autocorrelation,
711 functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation*
712 *Science*, 15, 77–84. doi: 10.1111/j.1654-1103.2004.tb02239.x

713 Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of
714 species distribution data. *Global Ecology and Biogeography*, 16, 129–138. doi:
715 10.1111/j.1466-8238.2006.00279.x
716 Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., ...
717 Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species
718 distributional data: a review. *Ecography*, 30, 609–628. doi: 10.1111/j.2007.0906-
719 7590.05171.x
720 Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., ...
721 Hancock, A. M. (2017). African genomes illuminate the early history and transition to
722 selfing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the*
723 *United States of America*, 114, 5213–5218. doi: 10.1073/pnas.1616736114
724 Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and
725 prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*,
726 40, 677–697. doi: 10.1146/annurev.ecolsys.110308.120159.
727 Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical
728 explanation of Maxent for ecologists. *Diversity and Distributions*, 17, 43–57. doi:
729 10.1111/j.1472-4642.2010.00725.x
730 Exposito-Alonso, M., Brennan, A.C., Alonso-Blanco, C., & Picó, F.X. (2018). Spatio-
731 temporal variation in fitness responses to contrasting environments in *Arabidopsis*
732 *thaliana*. *Evolution*, 72, 1570–1586. doi: 10.1111/evo.13508.
733 Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using
734 multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164,
735 1567–1587. PMCID: PMC1462648
736 Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions.

Journal of Applied Statistics, 31, 799–815. doi: 10.1080/0266476042000214501

Fortin, M. -J., & Dale, M. R. T. (2009). Spatial autocorrelation in ecological studies: a legacy of solutions and myths. *Geographical Analysis*, 41, 392–397. doi: 10.1111/j.1538-4632.2009.00766.x

Geisser, S. (1993). *Predictive inference* (1st ed.). London: Chapman and Hall.

Gneiting, R., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102, 359–378. doi: 10.1198/016214506000001437

Gomaa, N. H., Montesinos-Navarro, A., Alonso-Blanco, C., & Picó, F. X. (2011). Temporal variation in genetic diversity and effective population size of Mediterranean and subalpine *Arabidopsis thaliana* populations. *Molecular Ecology*, 20, 3540–3554. doi: 10.1111/j.1365-294X.2011.05193.x

Gotelli, N. J., & Stanton-Geddes, J. (2015). Climate change, genetic markers and species distribution modeling. *Journal of Biogeography*, 42, 1577–1585. doi:10.1111/jbi.12562

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993–1009. doi: 10.1111/j.1461-0248.2005.00792.x

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models*. Cambridge: Cambridge University Press.

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. doi: 10.1002/joc.1276

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). Package 'dismo': Species distribution modeling. https://CRAN.R-project.org/package=dismo

760 Hjelle, Ø., & Daehlen, M. (2006). *Triangulations and Applications (Mathematics and*
761 *Visualization)*. Heidelberg: Springer-Verlag GmbH.
762 Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists.
763 *Ecological Monographs*, 85, 3–28. doi: 10.1890/14-0661.1
764 Ikeda, D. H., Max, T. L., Allan, G. J., Lau, M. K., Shuster, S. M., & Whitham, T. G. (2017).
765 Genetically informed ecological niche models improve climate change predictions. *Global*
766 *Change Biology*, 23, 164–176. doi: 10.1111/gcb.13470
767 Jakobsson, M., & Rosenberg, N. A. 2007. CLUMPP: a cluster matching and permutation
768 program for dealing with label switching and multimodality in analysis of population
769 structure. *Bioinformatics*, 23, 1801–1806. doi:10.1093/bioinformatics/btm233
770 Jay, F., Manel, S., Alvarez, N., Durand, E. Y., Thuiller, W., Holderegger, R., ... François, O.
771 (2012). Forecasting changes in population genetic structure of alpine plants in response to
772 global warming. *Molecular Ecology*, 21, 2354–2368. doi: 10.1111/j.1365-
773 294X.2012.05541.x
774 Latimer, A. M., Wu, S., Gelfand, A. E., & Silander Jr., J. A. (2006). Building statistical
775 models to analyze species distributions. *Ecological Applications*, 16, 33–50. doi:
776 10.1890/04-0609
777 Legendre, P. (1993). Spatial autocorrelation–trouble or new paradigm. *Ecology*, 74, 1659–
778 1673. doi: 10.2307/1939924
779 Legendre, P., & Legendre, L. (2012). *Numerical ecology* (3rd ed.). Toronto: Elsevier Science.
780 Lenoir, J., & Svenning, J. -C. (2015). Climate-related range shifts – a global multidimensional
781 synthesis and new research directions. *Ecography*, 38, 15–28. doi: 10.1111/ecog.00967

782 Lima, J. S. Ballesteros-Mejia, L., Lima-Ribeiro, M. S., & Collevatti, R. G. (2017). Climatic
783 changes can drive the loss of genetic diversity in a Neotropical savanna tree species.
784 *Global Change Biology*, 23, 4639–4650. doi: 10.1111/gcb.13685
785 Lindgren, F., & Rue, H. 2015. Bayesian Spatial Modelling with R-INLA. *Journal of*
786 *Statistical Software*, 63, 19. URL: https://www.jstatsoft.org/article/view/v063i19
787 Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and
788 Gaussian Markov random fields: the stochastic partial differential equation approach.
789 *Journal of the Royal Statistical Society B*, 73, 423–498. doi: 1369–7412/11/73423
790 MacLean, S. A., & Beissinger, S. R. (2017). Species' traits as predictors of range shifts under
791 contemporary climate change: a review and meta-analysis. *Global Change Biology*, 23,
792 4094–4105. doi: 10.1111/gcb.13736
793 Manzano-Piedras, E., Marcer, A., Alonso-Blanco, C., & Picó, F. X. (2014). Deciphering the
794 adjustment between environment and life history in annuals: lessons from a
795 geographically-explicit approach in *Arabidopsis thaliana*. *PLoS ONE*, 9, e87836. doi:
796 10.1371/journal.pone.0087836
797 Marcer, A., Méndez-Vigo, B., Alonso-Blanco, C., & Picó, F. X. (2016). Tackling
798 intraspecific genetic structure in distribution models better reflects species geographical
799 range. *Ecology and Evolution*, 6, 2084–2097. doi: 10.1002/ece3.2010
800 Marcer, A., Vidigal, D. S., James, P. M. A., Fortin, M. -J., Méndez-Vigo, B., Hilhorst, H. W.
801 M., ... Picó, F. X. (2018). Temperature fine-tunes Mediterranean *Arabidopsis thaliana* life-
802 cycle phenology geographically. *Plant Biology*, 20, 148–156. doi: 10.1111/plb.12558
803 Martínez-Minaya, J., Cameletti, M., Conesa, D., & Pennino, M. G. (2018). Species
804 distribution modeling: a statistical review with focus in spatio-temporal issues. *Stochastic*
805 *Environmental Research and Risk Assessment,* 1-18. doi: 10.1007/s00477-018-1548-7

[dataset] Martínez-Minaya, J., Conesa, D., Fortin, M. J., Alonso-Blanco, C. Picó, F. X., & Marcer, A; 2019; A hierarchical Bayesian Beta regression approach to study the effects of geographic genetic structure and spatial autocorrelation on species distribution range shifts; Zenodo; http://doi.org/10.5281/zenodo.2552025.

Méndez-Vigo, B., Picó, F. X., Ramiro, M., Martínez-Zapater, J. M., & Alonso-Blanco, C. (2011). Altitudinal and climatic adaptation is mediated by flowering traits and *FRI*, *FLC* and *PHYC* genes in *Arabidopsis*. *Plant Physiology*, 157, 1942–1955. doi: 10.1104/pp.111.183426

Méndez-Vigo, B., Gomaa, N. H., Alonso-Blanco, C., & Picó, F. X. (2013). Among- and within-population variation in flowering time of Iberian *Arabidopsis thaliana* estimated in field and glasshouse conditions. *New Phytologist*, 197, 1332–1343. doi: 10.1111/nph.12082

Milanesi, P., Caniglia, R., Fabbri, E., Puopolo, F., Galaverni, M., & Holderegger, R. (2018). Combining Bayesian genetic clustering and ecological niche modeling: Insights into wolf intraspecific genetic structure. *Ecology and Evolution*, 8, 11224–11234. doi: 10.1002/ece3.4594

Nadeau, C. P., Urban, M. C., & Bridle, J. R. (2017). Climates past, present, and yet-to-come shape climate change vulnerabilities. *Trends in Ecology and Evolution*, 32, 786–800. doi: 10.1016/j.tree.2017.07.012

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., ... Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, 456, 98–101. doi: 10.1038/nature07331

Paradinas, I., Marín, M., Pennino, M. G., López-Quílez, A., Conesa, D., Barreda, D., González, M. & Bellido J.M. (2016). Identifying the best fishing-suitable areas under the

new European discard ban. *ICES Journal of Marine Science*, 73, 2479–2487. doi: 10.1093/icesjms/fsw114.

Paradinas, I., Pennino, M. G., López-Quílez, A., Marín, M., Bellido, J. M., & Conesa, D. (2018). Modelling spatially sampled proportion processes. *Revstat – Statistical Journal*, 16, 71–86.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution and Systematics*, 37, 637–669. doi: 1543-592X/06/1201-0637

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42. doi: 10.1038/nature01286

Perry, A. L., Low, P. J., Ellis, J. R., & Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. *Science*, 308, 1912–1915. doi: 10.1126/science.1111322

Phillips, S. J., Anderson, R, P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259. doi:10.1016/j.ecolmodel.2005.03.026

Picó, F. X., Méndez-Vigo, B., Martínez-Zapater, J. M., & Alonso-Blanco, C. (2008). Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. *Genetics*, 180, 1009–1021. doi: 10.1534/genetics.108.089581

Pino, J., Font, X., Carbó, J., Jové, M., & Pallarès, L. 2005. Large-scale correlates of alien plant invasion in Catalonia (NE of Spain). *Biological Conservation*, 122, 339–350. doi:10.1016/j.biocon.2004.08.006

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959. PMCID: PMC1461096

854 Record, S., Fitzpatrick, M. C., Finley, A. O., Veloz, S., & Ellison, A. M. (2013). Should
855 species distribution models account for spatial autocorrelation? A test of model projections
856 across eight millennia of climate change. *Global Ecology and Biogeography*, 22, 760–771.
857 doi: 10.1111/geb.12017
858 Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M.
859 W. (2005). Clines, clusters, and the effect of study design on the inference of human
860 population structure. *PLoS Genetics*, 1, e70. doi: 10.1371/journal.pgen.0010070
861 Rue, H., & Held, L. (2005). *Gaussian Markov random fields: theory and applications* (1st
862 ed.). London: Chapman and Hall.
863 Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent
864 Gaussian models using integrated nested Laplace approximations. *Journal of the Royal*
865 *Statistical Society B*, 71, 319–392. doi: 1369–7412/09/71319
866 Schmid, C. H. (2005). Multivariate classification rules: calibration and discrimination. In:
867 Armitage, P., & Colton, T. (Eds.), *Encyclopedia of Biostatistics*. Chichester: John Wiley &
868 Sons.
869 Serra-Varela, M. J., Alía, R., Ruiz Daniels, R., Zimmermann; N. E., Gonzalo-Jiménez, J., &
870 Grivet, D. (2017). Assessing vulnerability of two Mediterranean conifers to support genetic
871 conservation management in the face of climate change. *Diversity and Distributions*, 23,
872 507–516. doi: 10.1111/ddi.12544
873 Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian
874 measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64: 583–
875 639. doi: 10.1111/1467-9868.00353
876 Swanson, A. K., Dobrowski, S. Z., Finley, A. O., Thorne, J. H., & Schwartz, M. K. (2013).
877 Spatial regression methods capture prediction uncertainty in species distribution model

projections through time. *Global Ecology and Biogeography*, 22, 242–251. doi: 10.1111/j.1466-8238.2012.00794.x

The 1001 Genomes Consortium. (2016). 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, 166, 481–491. doi: 10.1016/j.cell.2016.05.063

Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, 348, 571–573. doi: 10.1126/science.aaa4984

Václavík, T., & Meentemeyer, R. K. (2009). Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling*, 220, 3248–3258. doi: 10.1016/j.ecolmodel.2009.08.013

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., ... Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, 109, 5–31. doi: 10.1007/s10584-011-0148-z

Vidigal, D. S., Marques, A. C. S. S., Willems, L. A. J., Buijs, G., Méndez-Vigo, B., Hilhorst, H. W. M., ... Alonso-Blanco, C. (2016). Altitudinal and climatic associations of seed dormancy and flowering traits evidence adaptation of annual life cycle timing in *Arabidopsis thaliana*. *Plant, Cell and Environment*, 39, 1737–1748. doi: 10.1111/pce.12734

Wagner, H. H., & Fortin, M. -J. (2013). A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics*, 14, 253–261. doi: 10.1007/s10592-012-0391-5

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594. URL: http://www.jmlr.org/papers/v11/watanabe10a.html

902 Wiens, J. J. (2016). Climate-related local extinctions are already widespread among plant and
903 animal species. *PLoS Biology*, 14, e2001104. doi: 10.1371/journal.pbio.2001104
904 Yannic, G., Pellissier, L., Ortego, J., Lecomte, N., Couturier, S., Cuyler, C., ... Côté, S. D.
905 (2014). Genetic diversity in caribou linked to past and future climate change. *Nature*
906 *Climate Change*, 4, 132–137. doi: 10.1038/NCLIMATE2074
907

Table 1. Bioclimatic variable percentage contributions to the fit of the best `Maxent` models and β coefficients of the best non-spatial and spatial HBMs for the distribution range of each genetic cluster of *A. thaliana* in the Iberian Peninsula. Bioclimatic variables: BIO1; Annual mean temperature, BIO2; Mean diurnal range, BIO3; Isothermality, BIO4; Temperature seasonality, BIO8; Mean temperature of the wettest quarter, BIO12; Annual precipitation, BIO15; Precipitation seasonality, and BIO18; Precipitation of the warmest quarter. For `Maxent`, the number of occurrence points was 103, 43, 38, and 35 for genetic clusters C1, C2, C3 and C4, respectively. For non-spatial and spatial HBMs, models included all 301 occurrence points.

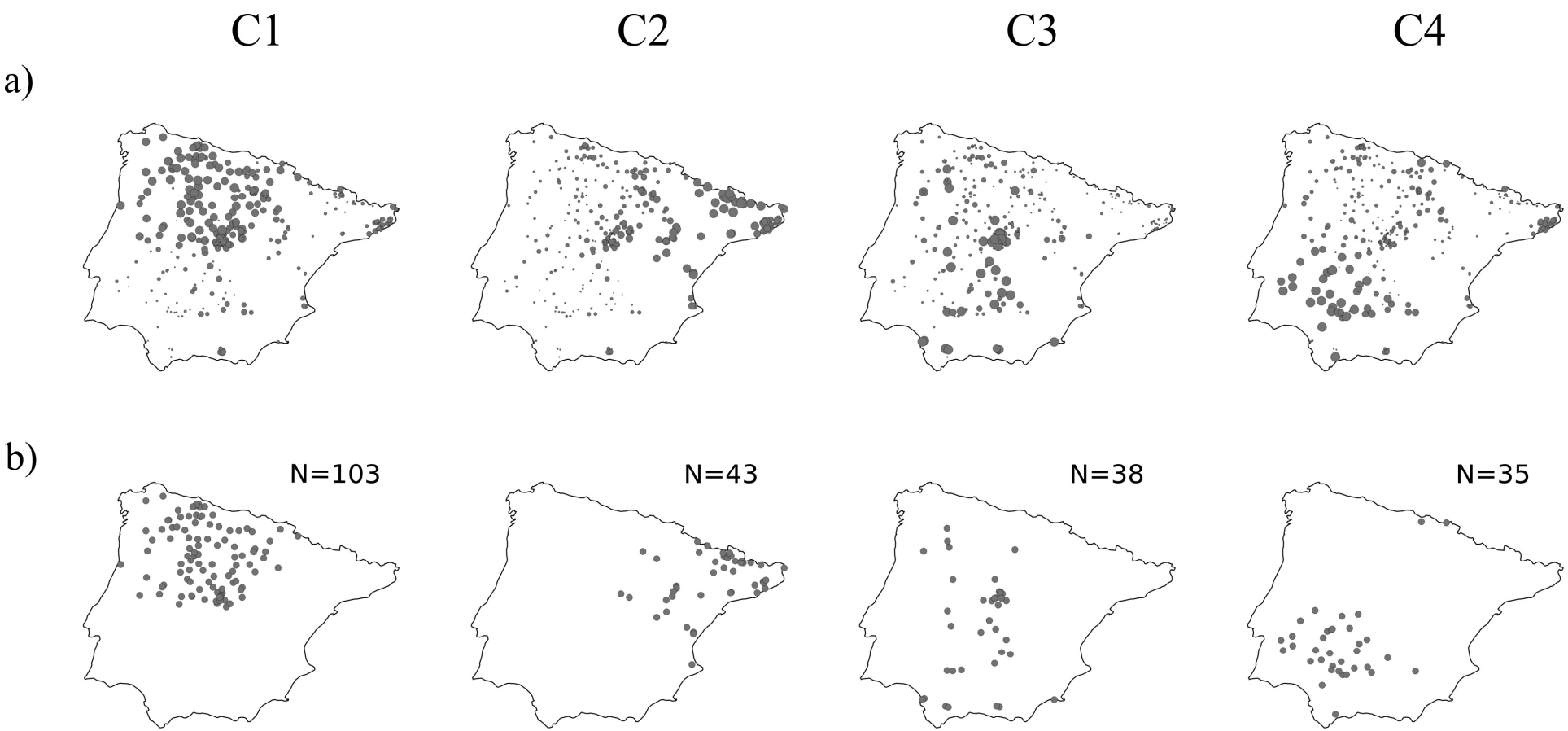
Cluster	Model	Bioclimatic predictors							
		BIO1	BIO2	BIO3	BIO4	BIO8	BIO12	BIO15	BIO18
C1	`Maxent`	64.07	–	–	9.73	17.30	–	1.74	7.16
	Non-spatial HBMs	–	–	5.773	-0.565	-0.104	–	-0.050	-0.009
	Spatial HBMs	0.147	–	–	–	-0.071	–	–	–
C2	`Maxent`	–	3.01	–	20.06	7.50	–	–	69.43
	Non-spatial HBMs	-0.112	–	–	0.373	0.104	-0.001	–	0.014
	Spatial HBMs	–	–	–	–	-0.044	-0.002	0.025	0.010
C3	`Maxent`	19.76	–	33.11	–	–	–	38.42	8.70
	Non-spatial HBMs	–	–	–	0.288	–	0.001	–	-0.006
	Spatial HBMs	–	–	–	–	–	–	–	–
C4	`Maxent`	22.65	2.52	–	11.54	–	20.18	–	43.11
	Non-spatial HBMs	0.197	–	–	–	–	–	0.021	0.004
	Spatial HBMs	0.164	–	–	–	–	–	–	–

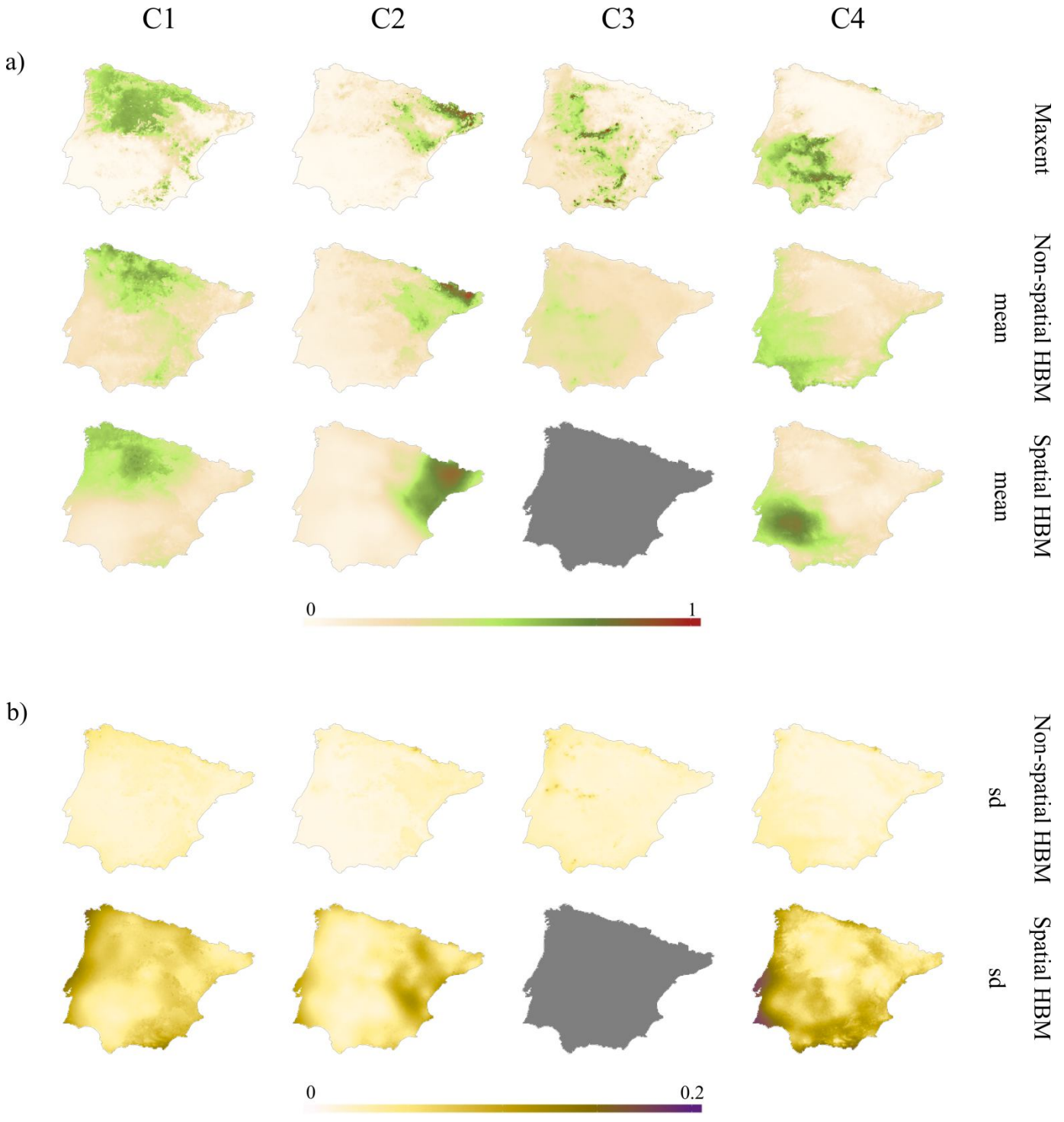
Table 2. Mean absolute error (MAE) and root mean squared error (RMSE) for spatial and non-spatial HBMs applied to each genetic cluster of *A. thaliana* in the Iberian Peninsula. The spatial effect term (W) is also indicated in spatial HBMs.

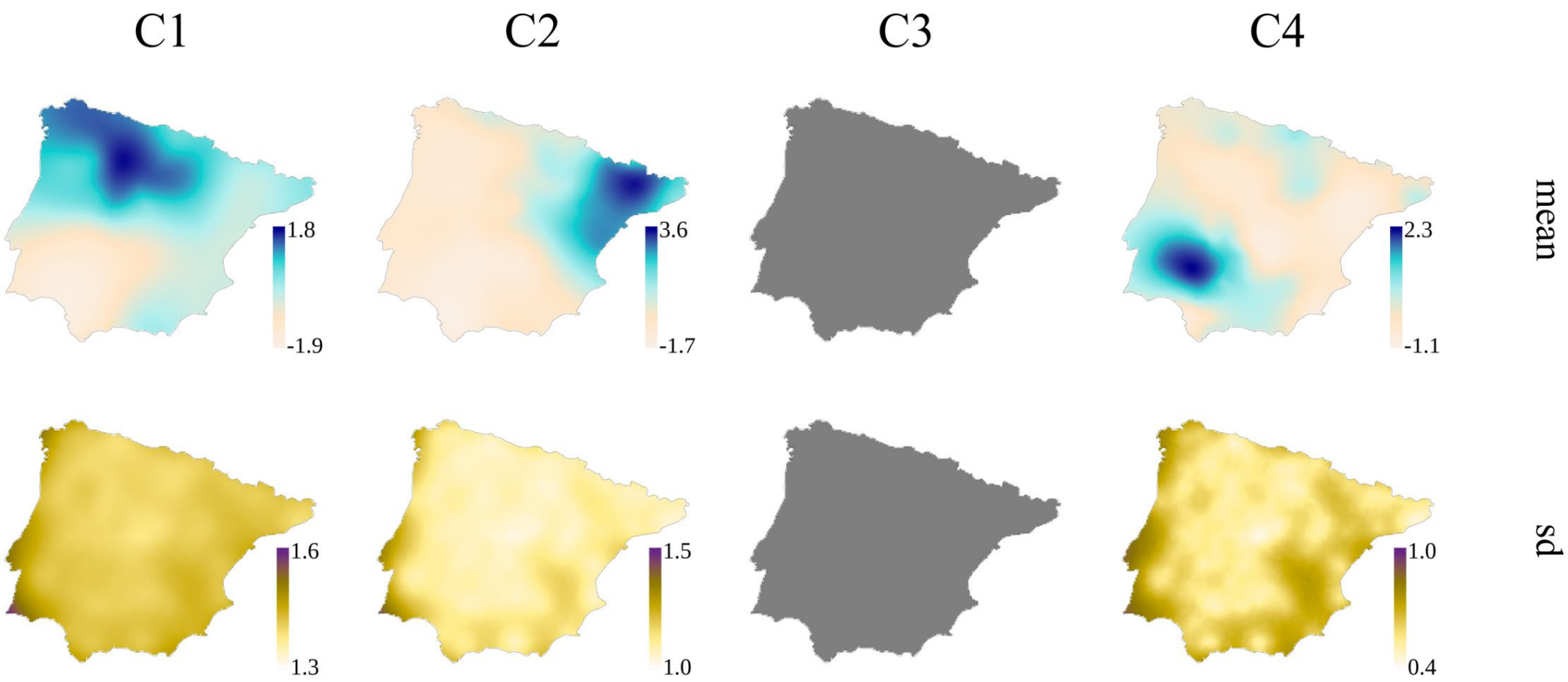
Cluster	HBM	Model	MAE	RMSE
C1	Non-spatial	$Y \sim \beta_0 + BIO3 + BIO4 + BIO8 + BIO15 + BIO18$	0.174	0.210
	Spatial	$Y \sim \beta_0 + BIO1 + BIO8 + W$	0.134	0.171
C2	Non-spatial	$Y \sim \beta_0 + BIO1 + BIO4 + BIO8 + BIO12 + BIO18$	0.116	0.153
	Spatial	$Y \sim \beta_0 + BIO8 + BIO12 + BIO15 + BIO18 + W$	0.070	0.095
C3	Non-spatial	$Y \sim \beta_0 + BIO4 + BIO12 + BIO18$	0.217	0.268
	Spatial	–	–	–
C4	Non-spatial	$Y \sim \beta_0 + BIO1 + BIO15 + BIO18$	0.148	0.189
	Spatial	$Y \sim \beta_0 + BIO1 + W$	0.096	0.128

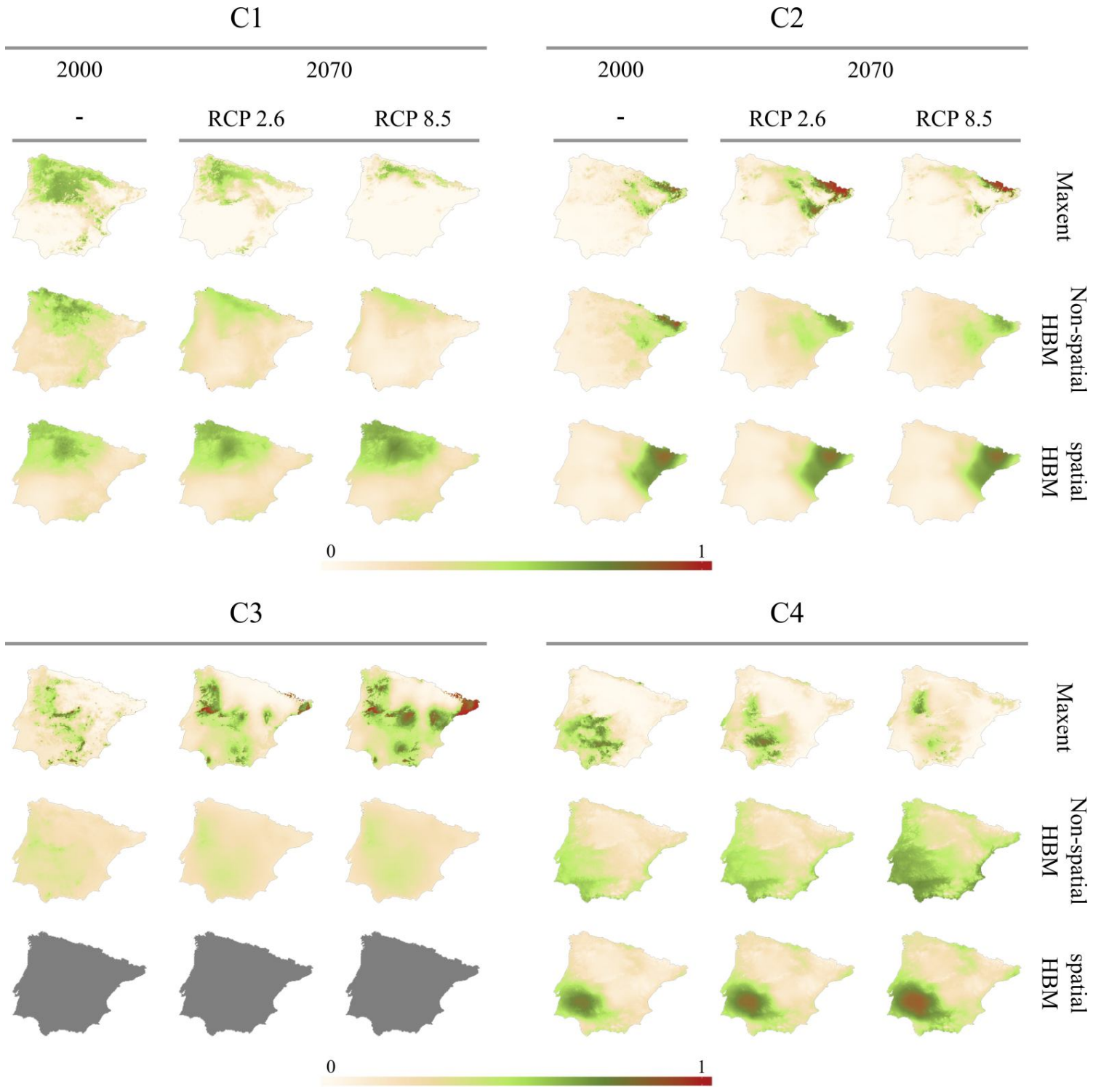
Table 3. Predicted cumulative probabilities for the entire Iberian Peninsula and percentage change, with respect to values in 2000, per genetic cluster, GCC scenario and modelling approach for each of the genetic clusters of *A. thaliana* in the Iberian Peninsula.

Cluster	GCC	Maxent		Non-spatial HBMs		Spatial HBMs	
		Cum. Prob.	% Change	Cum. Prob.	% Change	Cum. Prob.	% Change
C1	2000	5255.60	–	7277.59	–	7106.33	–
	RCP 2.6	3155.53	-39.96	6063.70	-16.68	7505.95	5.62
	RCP 8.5	1250.86	-76.20	4003.50	-44.99	8468.49	19.17
C2	2000	3020.12	–	4917.93	–	5318.48	–
	RCP 2.6	3746.01	24.04	5072.67	3.15	4953.02	-6.87
	RCP 8.5	2726.87	-9.71	4961.37	0.88	5194.97	-2.32
C3	2000	4540.23	–	6092.25	–	–	–
	RCP 2.6	6167.06	35.83	5984.34	-1.77	–	–
	RCP 8.5	8372.54	84.41	6343.11	4.12	–	–
C4	2000	4473.78	–	6698.21	–	6428.20	–
	RCP 2.6	5939.25	32.76	7894.27	17.86	7527.49	17.10
	RCP 8.5	4916.06	9.89	10378.67	54.95	8975.59	39.63


C1
C2
C3
C4
a)
b)
N=103
N=43
N=38
N=35


C1
C2
C3
C4
a)
Maxent
Non-spatial HBM
mean
Spatial HBM
mean
0
1
b)
Non-spatial HBM
sd
Spatial HBM
sd
0
0.2


C1
C2
C3
C4
mean
sd
1.8
-1.9
3.6
-1.7
2.3
-1.1
1.6
1.3
1.5
1.0
1.0
0.4


C1
2000
2070
-
RCP 2.6
RCP 8.5
C2
2000
2070
-
RCP 2.6
RCP 8.5
Maxent
Non-spatial HBM
spatial HBM
0
1
C3
C4
Maxent
Non-spatial HBM
spatial HBM
0
1

MOLECULAR ECOLOGY RESOURCES

Supplemental Information for:

A hierarchical Bayesian approach to study the effects of geographic genetic structure and spatial autocorrelation on species distribution range shifts

Joaquín Martínez-Minaya, David Conesa, Marie-Josée Fortin, Carlos Alonso-Blanco,

F. Xavier Picó, & Arnald Marcer

MOLECULAR ECOLOGY RESOURCES

Table S1. The best five `Maxent` models for each genetic cluster according to five-fold cross-validated AUC. We provide the model formula, the mean area under the curve (AUC) and its standard deviation (SD). The best model among the best five according to parsimony is indicated in bold. The number of occurrences after applying a threshold cut value of 0.5 was 103, 43, 38 and 35 for genetic cluster C1, C2, C3 and C4, respectively.

Cluster	Model	AUC	SD
C1	Y ~ BIO1 + BIO3 + BIO4 + BIO8 + BIO15 + BIO18	0.812	0.028
C1	**Y ~ BIO1 + BIO4 + BIO8 + BIO15 + BIO18**	**0.812**	**0.029**
C1	Y ~ BIO1 + BIO3 + BIO4 + BIO8 + BIO12 + BIO15	0.811	0.029
C1	Y ~ BIO1 + BIO2 + BIO4 + BIO8 + BIO12 + BIO15	0.811	0.029
C1	Y ~ BIO1 + BIO3 + BIO4 + BIO8 + BIO15	0.811	0.029
C2	Y ~ BIO1 + BIO2 + BIO4 + BIO8 + BIO18	0.910	0.039
C2	Y ~ BIO1 + BIO2 + BIO4 + BIO8 + BIO12 + BIO18	0.909	0.039
C2	Y ~ BIO1 + BIO2 + BIO3 + BIO4 + BIO8 + BIO18	0.908	0.040
C2	Y ~ BIO2 + BIO3 + BIO4 + BIO8 + BIO18	0.908	0.039
C2	**Y ~ BIO2 + BIO4 + BIO8 + BIO18**	**0.908**	**0.039**
C3	Y ~ BIO1 + BIO2 + BIO3 + BIO8 + BIO15 + BIO18	0.809	0.073
C3	Y ~ BIO1 + BIO2 + BIO3 + BIO15 + BIO18	0.808	0.073
C3	Y ~ BIO1 + BIO3 + BIO8 + BIO15 + BIO18	0.808	0.073
C3	**Y ~ BIO1 + BIO3 + BIO15 + BIO18**	**0.808**	**0.072**
C3	Y ~ BIO1 + BIO2 + BIO3 + BIO8 + BIO12 + BIO15 + BIO18	0.808	0.073
C4	**Y ~ BIO1 + BIO2 + BIO4 + BIO12 + BIO18**	**0.864**	**0.047**
C4	Y ~ BIO1 + BIO2 + BIO4 + BIO8 + BIO12	0.863	0.044
C4	Y ~ BIO1 + BIO2 + BIO3 + BIO8 + BIO12	0.863	0.049
C4	Y ~ BIO1 + BIO2 + BIO4 + BIO8 + BIO12 + BIO18	0.863	0.047
C4	Y ~ BIO1 + BIO2 + BIO3 + BIO4 + BIO12 + BIO18	0.863	0.048

MOLECULAR ECOLOGY RESOURCES

Table S2. Results for the Moran's *I* test on residual spatial autocorrelation (SAC) for each modelling approach and genetic cluster. Models with *P*-value > 0.05 are considered as residual SAC free. Spatial HBM for C3 is indicated by dashes because it did not produce acceptable results.

Method	Cluster	Moran's *I*	*P*-value	SAC free
`Maxent`	C1	0.0747	0.0158	No
`Maxent`	C2	0.3553	0.0001	No
`Maxent`	C3	0.7953	0.0001	No
`Maxent`	C4	0.0544	0.0626	Yes
Non-spatial HBM	C1	0.0775	0.0001	No
Non-spatial HBM	C2	0.0795	0.0001	No
Non-spatial HBM	C3	0.0417	0.0094	No
Non-spatial HBM	C4	0.1021	0.0001	No
Spatial HBM	C1	0.0147	0.1011	Yes
Spatial HBM	C2	0.0141	0.0982	Yes
Spatial HBM	C3	–	–	–
Spatial HBM	C4	-0.0015	0.4148	Yes

MOLECULAR ECOLOGY RESOURCES

Table S3. Summary of posterior distributions for the best non-spatial (A) and spatial HBMs (B) for each genetic cluster according to the logarithmic conditional predictive ordinates (LCPO). The mean, the standard deviation (SD), the quantiles (0.025, 0.5 and 0.975) and the mode of posterior distributions are given. Results of spatial HBM for C3 is not given as it did not produce acceptable results.

(A) Non-spatial HBMs

C1	Mean	SD	0.025 q	0.5 q	0.975 q	Mode
β_0	4.3073	1.8921	0.5817	4.3108	8.0106	4.3178
BIO3	5.7731	2.8855	0.1521	5.7570	11.4781	5.7247
BIO4	-0.5653	0.1232	-0.8077	-0.5652	-0.3240	-0.5648
BIO8	-0.1040	0.0156	-0.1348	-0.1040	-0.0733	-0.1039
BIO15	-0.0501	0.0060	-0.0620	-0.0501	-0.0383	-0.0500
BIO18	-0.0091	0.0015	-0.0120	-0.0091	-0.0063	-0.0091

C2	Mean	SD	0.025 q	0.5 q	0.975 q	Mode
β_0	-3.6781	0.8329	-5.3233	-3.6750	-2.0516	-3.6687
BIO1	-0.1117	0.0328	-0.1757	-0.1118	-0.0469	-0.1121
BIO4	0.3729	0.0916	0.1944	0.3724	0.5540	0.3713
BIO8	0.1042	0.0244	0.0563	0.1042	0.1519	0.1042
BIO12	-0.0011	0.0003	-0.0018	-0.0011	-0.0004	-0.0011
BIO18	0.0137	0.0019	0.0100	0.0136	0.0174	0.0136

C3	Mean	SD	0.025 q	0.5 q	0.975 q	Mode
β_0	-3.1345	0.8826	-4.8767	-3.1314	-1.4107	-3.1252
BIO4	0.2875	0.1184	0.0568	0.2869	0.5213	0.2857
BIO12	0.0012	0.0003	0.0006	0.0012	0.0018	0.0012
BIO18	-0.0062	0.0014	-0.0089	-0.0062	-0.0035	-0.0061

C4	Mean	SD	0.025 q	0.5 q	0.975 q	Mode
β_0	-4.9265	0.4764	-5.8607	-4.9270	-3.9904	-4.9281
BIO1	0.1972	0.0290	0.1405	0.1971	0.2543	0.1970
BIO15	0.0212	0.0059	0.0096	0.0212	0.0327	0.0212
BIO18	0.0036	0.0012	0.0012	0.0036	0.0059	0.0036

MOLECULAR ECOLOGY
RESOURCES

(B) Spatial HBMs

C1	Mean	SD	0.025 q	0.5 q	0.975 q	Mode
β_0	-2.2260	1.5085	-5.4395	-2.2065	0.9344	-2.1654
BIO1	0.1472	0.0507	0.0501	0.1464	0.2490	0.1447
BIO8	-0.0713	0.0317	-0.1343	-0.0710	-0.0100	-0.0704

C2	Mean	SD	0.025 q	0.5 q	0.975 q	Mode
β_0	-1.9800	1.0121	-4.1034	-1.9534	-0.0309	-1.9139
BIO8	-0.0442	0.0209	-0.0854	-0.0441	-0.0033	-0.0440
BIO12	-0.0015	0.0005	-0.0025	-0.0015	-0.0005	-0.0015
BIO15	0.0255	0.0111	0.0039	0.0254	0.0474	0.0253
BIO18	0.0095	0.0032	0.0033	0.0095	0.0160	0.0094

C4	Mean	SD	0.025 q	0.5 q	0.975 q	Mode
β_0	-4.9265	0.4764	-5.8607	-4.9270	-3.9904	-4.9281
BIO1	0.1972	0.0290	0.1405	0.1971	0.2543	0.1970

MOLECULAR ECOLOGY RESOURCES

Table S4. Mean posterior distribution for the hyper-parameter $\phi_i = \exp(\theta)$ for each of the best non-spatial and spatial HBMs for each genetic cluster (C1, C2, C3 and C4).

Model	C1	C2	C3	C4
Non-spatial HBMs	3.645	6.232	2.171	3.879
Spatial HBMs	4.651	13.228	–	6.891

MOLECULAR ECOLOGY
RESOURCES

Table S5. The best five non-spatial (A) and spatial HBMs (B) for each genetic cluster according to LCPO. We provide the model formula, the deviance information criterion (DIC), the Watanabe-Akaike information criterion (WAIC), and the logarithmic conditional predictive ordinates (LCPO). For each genetic cluster, the best model among the best five according to parsimony is indicated in bold. For spatial models, the spatial effect term (W) is also indicated in the formula. Spatial HBM for C3 is indicated by dashes because it did not produce acceptable results.

(A) Non-spatial HBMs

Cluster	Model	DIC	WAIC	LCPO
C1	**Y ~ β_0+ BIO3 + BIO4 + BIO8 + BIO15 + BIO18**	**-212.81**	**-213.39**	**-0.354**
C1	Y ~ β_0+ BIO3 + BIO4 + BIO8 + BIO12 + BIO15 + BIO18	-213.01	-213.38	-0.354
C1	Y ~ β_0+ BIO2 + BIO4 + BIO8 + BIO12 + BIO15 + BIO18	-212.73	-212.87	-0.353
C1	Y ~ β_0+ BIO2 + BIO4 + BIO8 + BIO15 + BIO18	-212.43	-212.80	-0.353
C1	Y ~ β_0+ BIO2 + BIO3 + BIO8 + BIO15 + BIO18	-212.02	-212.60	-0.353
C2	Y ~ β_0+ BIO1 + BIO4 + BIO8 + BIO12 + BIO15 + BIO18	-456.02	-454.53	-0.755
C2	**Y ~ β_0+ BIO1 + BIO4 + BIO8 + BIO12 + BIO18**	**-454.12**	**-452.62**	**-0.752**
C2	Y ~ β_0+ BIO1 + BIO3 + BIO4 + BIO8 + BIO12 + BIO15 + BIO18	-454.01	-452.61	-0.752
C2	Y ~ β_0+ BIO1 + BIO2 + BIO4 + BIO8 + BIO12 + BIO15 + BIO18	-454.04	-452.59	-0.752
C2	Y ~ β_0+ BIO1 + BIO2 + BIO3 + BIO8 + BIO12 + BIO15 + BIO18	-453.03	-451.67	-0.750
C3	**Y ~ β_0+ BIO4 + BIO12 + BIO18**	**-356.48**	**-356.84**	**-0.593**
C3	Y ~ β_0+ BIO1 + BIO4 + BIO12 + BIO18	-356.06	-356.72	-0.593
C3	Y ~ β_0+ BIO1 + BIO2 + BIO4 + BIO12 + BIO18	-355.14	-355.93	-0.591
C3	Y ~ β_0+ BIO1 + BIO3 + BIO12 + BIO18	-355.27	-355.80	-0.591
C3	Y ~ β_0+ BIO1 + BIO3 + BIO4 + BIO12 + BIO18	-354.78	-355.56	-0.591
C4	**Y ~ β_0+ BIO1 + BIO15 + BIO18**	**-379.14**	**-377.35**	**-0.627**
C4	Y ~ β_0+ BIO1 + BIO8 + BIO15 + BIO18	-378.19	-376.57	-0.625
C4	Y ~ β_0+ BIO1 + BIO2 + BIO3 + BIO4 + BIO15 + BIO18	-378.23	-376.15	-0.625
C4	Y ~ β_0+ BIO1 + BIO4 + BIO15 + BIO18	-377.62	-375.81	-0.624
C4	Y ~ β_0+ BIO1 + BIO3 + BIO15 + BIO18	-377.42	-375.71	-0.624

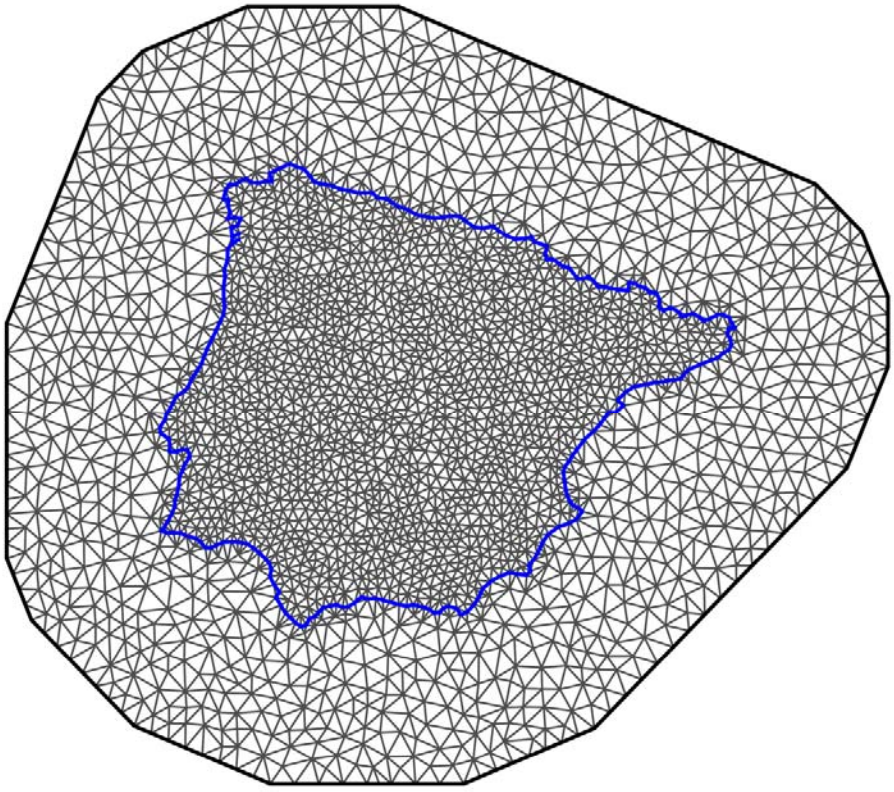
MOLECULAR ECOLOGY
RESOURCES

(B) Spatial HBMs

Cluster	Model	DIC	WAIC	LCPO
C1	Y ~ β_0+ BIO1 + BIO3 + BIO8 + W	-262.07	-262.46	-0.434
C1	**Y ~ β_0+ BIO1 + BIO8 + W**	**-261.30**	**-261.71**	**-0.434**
C1	Y ~ β_0+ BIO1 + BIO2 + BIO8 + W	-261.25	-261.52	-0.433
C1	Y ~ β_0+ BIO1 + BIO3 + BIO4 + BIO8 + W	-261.09	-261.47	-0.433
C1	Y ~ β_0+ BIO1 + BIO2 + BIO3 + BIO8 + W	-261.07	-261.32	-0.432
C2	Y ~ β_0+ BIO2 + BIO8 + BIO12 + BIO15 + BIO18 + W	-616.66	-616.29	-1.014
C2	**Y ~ β_0+ BIO8 + BIO12 + BIO15 + BIO18 + W**	**-616.07**	**-615.55**	**-1.012**
C2	Y ~ β_0+ BIO1 + BIO2 + BIO8 + BIO12 + BIO15 + BIO18 + W	-615.45	-614.96	-1.011
C2	Y ~ β_0+ BIO4 + BIO8 + BIO12 + BIO15 + BIO18 + W	-615.11	-614.83	-1.010
C2	Y ~ β_0+ BIO3 + BIO8 + BIO12 + BIO15 + BIO18 + W	-615.62	-613.85	-1.010
C3	–	–	–	–
C3	–	–	–	–
C3	–	–	–	–
C3	–	–	–	–
C3	–	–	–	–
C4	**Y ~ β_0+ BIO1 + W**	**-492.78**	**-475.08**	**-0.779**
C4	Y ~ β_0+ BIO1 + BIO2 + W	-493.30	-475.13	-0.778
C4	Y ~ β_0+ BIO1 + BIO3 + W	-492.88	-474.65	-0.777
C4	Y ~ β_0+ BIO1 + BIO4 + W	-492.30	-474.26	-0.777
C4	Y ~ β_0+ BIO1 + BIO2 + BIO15 + W	-491.88	-474.07	-0.777

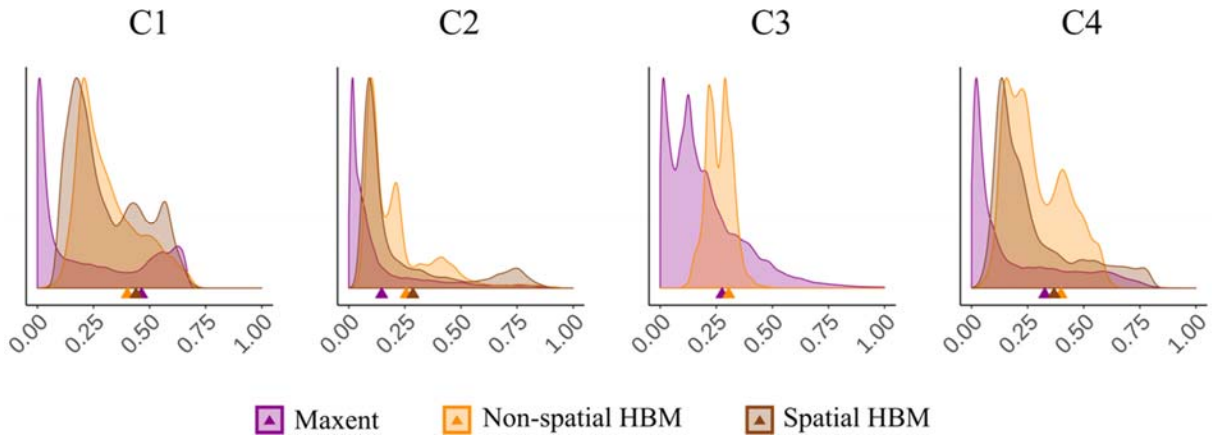
MOLECULAR ECOLOGY
RESOURCES

Figure S1. Delaunay triangulation used in HBMs to predict the response variable in un-sampled locations.



MOLECULAR ECOLOGY
RESOURCES

Figure S2. Density distributions of predicted genetic cluster membership probabilities for the whole of the study area for each modelling approach. `Maxent` densities must be interpreted as suitability for populations with a higher than 0.5 cluster coefficient. Small coloured triangles indicate the 0.75 percentile of the corresponding coloured distribution.



MOLECULAR ECOLOGY
RESOURCES

Figure S3. Density plots of predicted current (year 2000) and future distributions (year 2070) of suitability values across the Iberian Peninsula for each genetic cluster and modelling approach. Small coloured triangles indicate the 0.75 percentile of the corresponding coloured distribution.