INTRODUCTION

Climate and land-use changes recorded in practically all of Earth's bioclimatic zones are dramatically affecting the distribution of many terrestrial, aquatic and marine organisms. Since the turn of the 21st century, various global meta-analyses have quantified the fingerprint that global climate change (GCC) has already left on distribution ranges (Chen, Hill, Ohlemüller, Roy, & Thomas, 2011; MacLean & Beissinger, 2017; Parmesan, 2006; Parmesan & Yohe, 2003; Perry, Low, Ellis, & Reynolds, 2005) and extinction rates (Urban, 2015; Wiens, 2016). At present, models include some improvements to better predict changes in species distribution ranges due to GCC.
Such improvements are aimed at considering all the possible organism 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 ranges of terrestrial organisms and their GCC-induced shifts.

First, demographic processes (i.e., extinction/colonization dynamics and dispersal ability) and adaptation to local environmental conditions determine the extent of population stratification, namely 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 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). In addition, it is very informative to determine whether such genetic clusters are geographically distributed across the species distribution range (Novembre et al., 2008; Rosenberg et al., 2005). This is because we can interpret the number of genetic clusters, their geographical 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; D’Amen, Zimmermann, & Pearman, 2013; Diniz-Filho et al., 2016; Gotelli & Stanton-Geddes, 2015; Ikeda et al., 2017; Jay et al., 2012; Lima, Ballesteros-Mejía, Lima-Ribeiro, & Collevatti, 2017; Mercè, Méndez-Vigo, Alonso-Blanco, & Picó, 2016; Milanesi et al., 2018; Yannic et al., 2014), stressing the need to consider the inherent genetic heterogeneity of organisms.

From a methodological viewpoint, working with intraspecific patterns of genetic diversity implies the combination of presence-only data, which commonly feed SDMs such as Maxent, with genetic structure data, which are generally expressed as genetic cluster membership proportions (ranging between 0 and 1) that inform on the degree of genetic admixture (Serra-Varela et al., 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 to run presence-only SDMs (Gotelli & Stanton-Geddes, 2015; Ikeda et al., 2017; Mercè et al., 2016). In doing that, the amount of data and information lost depends on the genetic structure of the study organism. For instance, species with a pronounced genetic structure will probably have individuals with high genetic cluster membership proportions, which facilitates the assignment of individuals to single genetic clusters. In contrast, individual assignment to single genetic clusters will exhibit higher uncertainty for weakly genetically structured organisms (e.g., high levels of individual genetic admixture), posing problems for the development of SDMs to study the effects of GCC on their patterns of geographic genetic structure. Either way, we lose valuable information that may reduce the value and impact of GCC model outcomes and therefore our understanding of the GCC effects on biodiversity.

The second biological aspect worth considering when modelling distribution ranges is the presence of spatial autocorrelation (SAC) in data and the problems that SAC poses for statistical and ecological interpretation. Spatial autocorrelation can be defined as the dependence between close observations in space (Legendre & Legendre, 2012), and it may be caused by exogenous factors (e.g., historical processes and autocorrelated environmental variables) and/or endogenous factors (e.g., dispersion) (Dale & Fortin, 2014). While variables representing exogenous factors may be readily available to researchers, variables describing endogenous factors representing important biological processes are more difficult to obtain (Belmaker et al., 2015). Overall, SAC is recognized as a major challenge when predicting species distributions (Dírnböck & Dullinger, 2004; de Oliveira, Rangel, Lima-Ribeiro, Terribile, & Diniz-Filho, 2014) because it results in several modelling flaws, such as violation of the assumption of independent error, inflated estimations of model performance, bias in model selection or inferential problems (Beale, Lennon, Yearsley, Brewer, & Elston, 2010; Dale & Fortin, 2002; Dormann, 2007; Dormann et al., 2007; Fortin & Dale, 2009; Legendre, 1993; Swanson, Dobrowski, Finley, Thorne, & Schwartz, 2013; Wagner & Fortin, 2013). To a certain extent, SAC can be dealt with via data filtering, although often at a high cost of data loss. For these reasons, taking SAC into account in GCC models is considered as mandatory (Belmaker et al., 2015; Crase, Liedloff, Vesk, Fukuda, & Wintle, 2014; Latimer, Wu, Gelfand, & Silander, 2006; Record, Fitzpatrick, Finley, Veloz, & Ellison, 2013; Swanson et al., 2013).

Here, we use hierarchical Bayesian models (HBMs), which account for the geographical distribution of intraspecific genetic diversity and the presence of SAC, to analyse current distribution range as well as the effect of GCC on its shifts. To that end, we use a collection of 301 natural populations of the annual plant Arabidopsis thaliana occurring in the Iberian Peninsula, the region of the distribution range with the highest genomic diversity (The 1001 Genomes Consortium, 2016). Genome-wide markers are used to infer Iberian A. thaliana’s geographical genetic structure by estimating genetic cluster membership proportions. To better understand the potential of our model, we compare three approaches, Maxent and two Bayesian, representing a gradient of complexity in the treatment of intraspecific genetic data and SAC, in particular (a) presence-only SDMs (Maxent) that do not take SAC into account and are based on binary data for genetic cluster membership proportions; (b) nonspatial HBMs not accounting explicitly for SAC and based on continuous data for genetic cluster membership proportions; and (c) spatially explicit HBMs accounting for SAC and based on continuous data for genetic cluster membership...
proportions. Although HBMs represent well-established methods for statistical inference in several research fields, the application of Beta regression to climate-driven shifts in species distribution range is not common (see Martínez-Minaya, Cameletti, Conesa, & Pennino, 2018). In particular, we promote the use of Beta regressions where data fitting can be achieved using integrated nested Laplace approximation (INLA) rather than Markov chain Monte Carlo (MCMC) methods. We discuss our results in terms of the relevance of intraspecific genetic variation and SAC to better interpret and contextualize the implications of GCC on species distribution range shifts, but also identifying the limitations and caveats of our approach.

2 | MATERIALS AND METHODS

2.1 | Source populations and genetic structure

We used a collection of 301 natural populations of the annual plant *A. thaliana* distributed across the Iberian Peninsula (~800 × 700 km²; 36.00°–43.48°N, 3.19°–9.30°W; Figure 1a). This set of populations belongs to a long-term project pursuing a permanent collection of natural populations from the western Mediterranean Basin (Spain, Portugal and North Africa) intended to unravel *A. thaliana*’s evolutionary ecology, functional genetics, and response to GCC (see Marcer et al., 2018, and references therein). Distance among populations and altitudes had a range of 1–1,038 km (mean ± SD = 360.9 ± 200.2 km) and 1–2,662 m above sea level (a.s.l) (mean ± SD = 786.5 ± 391.3 m a.s.l.), respectively, including a wide array of natural and modified environments (Manzano-Piedras, Marcé, Alonso-Blanco, & Picó, 2014; Méndez-Vigo, Picó, Ramiro, Martínez-Zapater, & Alonso-Blanco, 2011; Picó, Méndez-Vigo, Martínez-Zapater, & Alonso-Blanco, 2008).

**FIGURE 1** (a) Geographical position of the 301 *Arabidopsis thaliana* accessions of study for the four genetic clusters detected in the Iberian Peninsula. Dot size is proportional to genetic cluster membership proportion. For each accession, the four genetic cluster membership proportions sum to 1. (b) Geographical position of selected accessions after applying the membership proportion threshold of 0.5. The number of accessions included per genetic cluster is also indicated.

Populations included in this study came from field surveys that spanned 12 years (2000–2011). We sampled seed from several individuals from each population. Every year and a few months after every survey, field-collected seed was multiplied by the single seed descent method in a glasshouse at the Centro Nacional de Biotecnología (CNB-CSIC) in Madrid. Seeds that had multiplied were stored in dry, dark conditions in cellophane bags at room temperature, storing conditions that can preserve *A. thaliana* seeds for years. In this study, we used one representative individual (hereafter accession) 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 characterize Iberian *A. thaliana* (Gomaa, Montesinos-Navarro, Alonso-Blanco, & Picó, 2011; Manzano-Piedras et al., 2014; Marcé et al., 2016; Méndez-Vigo et al., 2011; Picó et al., 2008). In short, SNPs were selected based on their polymorphism shown in Central Europe, the Iberian Peninsula and in a worldwide collection of accessions, and were genotyped using the SNPlex technique (Applied Biosystems). 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* version 2.3.3 (Falush et al., 2003), as previously described (Méndez-Vigo et al., 2013,
2011). 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 determined by testing the differences between the data likelihood for successive K values using Wilcoxon tests for two related samples. The final K number was estimated as the largest K value with significantly higher likelihood than that of K = 1 runs (two-sided p < 0.005). This was supported by a high similarity among the ancestry membership matrices from different runs of the same K value (H’ = 0.99). The average symmetric similarity coefficient H’ among runs and the average matrix of ancestry membership proportions, derived from the 20 runs, were calculated with CLUMPP version 1 (Jakobsson & Rosenberg, 2007). This analysis inferred four genetic clusters in the Iberian Peninsula (Figure 1), in agreement with previous studies on A. thaliana’s genetic structure (Gomaa et al., 2011; Manzano-Piedras et al., 2014; Marcer et al., 2016; Méndez-Vigo et al., 2011; Picó et al., 2008).

2.2 | Climatic variables and GCC scenarios

We selected a total of eight bioclimatic predictors to define the climatic space: annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), mean temperature of the wettest quarter (BIO8), annual precipitation (BIO12), precipitation seasonality (BIO15) and precipitation of the warmest quarter (BIO18). These bioclimatic predictors were selected because their pairwise correlation coefficients were <0.7, a threshold value commonly used to avoid unacceptable colinearity among independent variables (Pino, Font, Carbó, Jové, & Pallarès, 2005). We used the dismo R package (Hijmans, Phillips, Leathwick, & Elith, 2017) to retrieve these climate layers from the Digital Atlas of the Iberian Peninsula (http://www.opengis.uab.cat/wms/iberia/), which provides interpolated surface layers of mean monthly data obtained from 2,285 weather 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 (RCPs) 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 version 1.4 (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). Resolution of the climatic spaces for the years 2000 and 2070 was 1 km.

2.3 | Species distribution models

Species distribution models link information on the presence/absence or abundance of a species to environmental variables to predict where it is likely to be present in unsampled locations or time periods (Elith & Leathwick, 2009; Guisan & Thuiller, 2005). In recent years, the quantity and the quality of the data sets have increased substantially, 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 (reviewed by 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 version 3.3.3k (Elith et al., 2011; Phillips, Anderson, & Schapire, 2006) 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 accessions for genetic cluster C4 to a high of 103 for genetic cluster C1 (Figure 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). We then reran 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 nonspatial HBMs were also used to model the current and future distribution of A. thaliana’s genetic clusters. In particular, spatial and nonspatial Beta regressions were conducted to estimate the probability of genetic cluster membership, 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, \( \text{Beta}(a, b) \), which can be parameterized in terms of its mean, \( \mu \), a dispersion parameter, \( \phi = a + b \), and the variance, \( \sigma^2 = \frac{a(1-a)}{1+\phi} \). This parameterization 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 \( \mu \) decreases as \( \phi \) 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 probabilities of genetic cluster membership 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 \( Y_i | \mu_i, \phi_i \sim \text{Beta}(\mu_i, \phi_i) \), where \( \mu_i \) and \( \phi_i \) are the Beta distribution parameters at location \( i \). We used the logit and log links for \( \mu_i \) and \( \phi_i \), respectively. The mean was linked to climatic covariates (nonspatial term) and, in the case of spatial models, to a stochastic spatial effect (spatial term). Precision was linked to climatic covariates (nonspatial term) and, in the case of spatial models, to a stochastic spatial effect (spatial term). Precision was assumed to be not dependent on 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:

\[
\begin{align*}
\logit(\mu_i) &= X_i \beta + W_i, \\
\log(\phi_i) &= \theta,
\end{align*}
\]

where \( \beta \) is the vector of regression coefficients \( (\beta_0, \beta_1, \ldots, \beta_p) \), \( X_i \) is the vector corresponding to the \( i \)th row of the design matrix whose first element is 1 (the one multiplying the intercept \( \beta_0 \)), the covariate values at location \( i \) being the remaining elements, and \( W_i \) is the spatially structured random effect at each location \( i \). \( W_i \) is assumed to be a multivariate Gaussian distribution whose covariance matrix, \( \sigma^2_WH(\phi) \), depends on the distance between locations, and its parameters, \( \sigma^2_W \) and \( \phi \), 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 hyperparameters was expressed. This hierarchical structure can also be considered as a latent Gaussian model (Rue & Held, 2005).

As posterior distributions for the parameters and hyperparameters do not have an analytical expression, numerical approximations are usually needed. In the case of latent Gaussian models, 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, and 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 reparameterized as follows:

\[
W \sim N(0, Q(\kappa, \tau))
\]

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

We specified prior distributions for the parameters and hyperparameters. 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 \( \kappa \) and \( \tau \), we specified the spatial effect in terms of \( \varphi \) and \( \sigma_W \) using the reparameterizations \( \log(\varphi) \) and \( \log(\sigma_W) \) as independent normal vague distributions (Lindgren & Rue, 2015).

Overall, the full model was stated as follows:

\[
Y_i | \mu_i, \phi_i \sim \text{Beta}(\mu_i, \phi_i) \]

\[
\begin{align*}
\logit(\mu_i) &= X_i \beta + W_i, \\
\log(\phi_i) &= \theta,
\end{align*}
\]

\[
\beta_0, \beta_1, \ldots, \beta_p \sim N(0, 10^{-4}), \ W \sim N(0, Q(\varphi, \sigma_W))
\]

\[
\begin{align*}
\log(\varphi) &\sim N(m_\varphi, a_\varphi), \\
\log(\sigma_W) &\sim N(m_{\sigma_W}, a_{\sigma_W}), \\
\theta &\sim \log\text{Gamma}(0.1)
\end{align*}
\]

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

These latter values, $q_p$ and $q_{np}$, express the large uncertainty about the parameters before the analysis, resulting in quite non-informative hyperpriors. 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 provides the possibility to evaluate its mean effects and uncertainty. As mentioned above, once the inference is made, the main interest becomes to predict the response in unsampled locations. To do that, we applied the SPDE by constructing a Delaunay triangulation (Hjelle & Daehlen, 2006) covering the whole Iberian Peninsula (Figure S1).

2.3.3 | Model selection, distribution range shifts and residual SAC

Hierarchical Bayesian models 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 $LCPO = -\frac{1}{N} \sum_i \log(\text{CPO}_i)$ as a summary statistic of the conditional predictive ordinate (CPO; Geisser, 1993), which gives an overall measure of predictive performance (Hooten & Hobbs, 2015). Conditional predictive ordinate is defined as the cross-validated predictive density at a given observation. Conditional predictive ordinate can be used to compute predictive measures, such as the logarithmic score (Gneiting & Raftery, 2007) or the cross-validated mean Brier score (Schmid & Griffith, 2005). Among the best five models for each genetic cluster we selected the most parsimonious one, namely the one with the least number of predictors. Model quality estimators, such as the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) and the Watanabe–Akaike information criterion (WAIC; Watanabe, 2010) were also computed. We also measured accuracy of spatial and nonspatial HBMs by means of mean absolute error (MAE) and root mean squared error (RMSE). Lower values of MAE and RMSE indicate better accuracy. The comparison between Maxent and HBMs in terms of accuracy can be misleading because Maxent used subsamples of data for each genetic cluster whereas HBMs were always based on the entire data set (Figure 1).

We compared distribution range shifts due to GCC when taking the spatial component into account (spatial HBM), when excluding the spatial component (nonspatial HBM) and Maxent. We added the probabilities calculated across the whole study area by each model for each time frame and GCC scenario to quantify the geographical extent of the suitability of each genetic cluster for each methodology. These probabilities were used to calculate the percentage loss or gain of suitability for each model and GCC scenario.

All models mentioned above were checked for residual SAC. We calculated the residuals for model predictions between observed and predicted values and tested for residual SAC using the spdep R package (Bivand, Hauke, & Kossowski, 2013; Bivand & Piras, 2015). To calculate residual SAC in Maxent, we followed the methodology used elsewhere (De Marco, Diniz-Filho, & Bini, 2008; Václavík & Meentemeyer, 2009). Basically, we estimated the Moran’s I coefficient of autocorrelation with 10,000 MCMC iterations. Models with $p > 0.05$ were considered as SAC-free. As expected, spatial HBM did not show residual SAC, while nonspatial 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 genetic clusters of A. thaliana 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 geographical distribution of genetic cluster ranges in the Iberian Peninsula (Figure 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 nonspatial HBMs produced broadly similar geographic distributions of genetic clusters to those generated by Maxent models, particularly for genetic clusters C1 and C4 (Figure 2a). In the case of genetic cluster C2, the spatial HBM depicted a rather continuous distribution in northeast Spain, which clearly differed from those given by Maxent and nonspatial HBM, showing the truncated distribution that this genetic cluster actually has in the wild (Figure 2a). In general, spatial HBMs and Maxent models showed more compact distribution ranges than nonspatial HBMs, the former with more dramatic transitions between low and high probability values in all genetic clusters (Figure 2a; Figure S2).

The exception was genetic cluster C3, whose predicted distribution range with nonspatial HBMs was rather blurred in comparison with that obtained with Maxent (Figure 2a). In fact, it was not possible to fit spatial HBMs for genetic cluster C3 because the results were inconsistent. When using vague hyperpriors for the range of the spatial effect, the resulting mean of the posterior distribution of the range was larger than the whole study area. By contrast, when using more informative priors, results were different and very much conditioned by prior selections. Thus, the spatial effect for genetic cluster C3 did not provide explanation further to what can already be explained by the bioclimatic predictors.

For all genetic clusters, the uncertainty of the predictive mean for nonspatial HBMs was lower and more evenly distributed across space than that for spatial HBMs (Figure 2b). The main reason for this apparent reduction in uncertainty is that spatial models reflect...
the intrinsic variability of the Beta-distributed data, variability that is not reflected by nonspatial models. As a result, the distribution of means and standard deviations for spatial HBM was more pronounced than those for nonspatial HBM (Figure 2; Table S3). Nonetheless, mean values of the posterior distribution of the precision parameter, which are inversely proportional to the variance of the data, were larger in the spatial HBM, reflecting their acceptable behaviour (Table S4). In the case of spatial HBM, these models allowed the spatial effects to be visualized, which clearly were more intense at the centre of the genetic cluster distribution ranges (Figure 3). Uncertainty of the mean of the spatial effect was greater for genetic cluster C4 than for the other two genetic clusters (Figure 3). Overall, spatial HBMs selected fewer bioclimatic predictors than nonspatial HBMs and Maxent models to define the distribution range of the four genetic clusters (Table 1; Table S5), particularly for genetic cluster C4. Finally, the combination of bioclimatic predictors used by Maxent models and spatial and nonspatial HBMs was quite different (Table 1).

Mean absolute error and RMSE were lower for spatial than for nonspatial HBM for all genetic clusters in which the comparison was possible (Table 2). This indicated that spatial HBM had lower average model prediction errors in the 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, Figure 4; Figure 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 nonspatial HBMs (Table 3, Figure 4). For genetic cluster C1, important reductions in distribution range were predicted for the two GCC scenarios with Maxent and nonspatial HBMs, whereas spatial HBMs predicted slight increases (Table 3, Figure 4). For genetic cluster C2, Maxent predicted increasing and decreasing distribution ranges with RCPs 2.6 and 8.5, respectively, whereas both HBMs predicted small fluctuations in distribution range in both GCC scenarios (Table 3, Figure 4). For genetic cluster C3, Maxent showed very large increases in distribution range, particularly for the RCP 8.5 scenario, whilst nonspatial HBMs predicted slight fluctuations in distribution range in both GCC scenarios (Table 3, Figure 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, Figure 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. Global climate change models for distribution range shifts have to increase their sophistication by adding realism to the model outcomes, yet without losing interpretability or without increasing uncertainty. In this study, we address this issue by developing HBMs for the annual plant A. thaliana incorporating two of these elements, which are
inherent to practically all organisms: the geographical distribution of intraspecific genetic variation and the SAC in the data. Importantly, both geographical genetic structure and SAC can be considered as indicators of eco-evolutionary forces shaping species’ distribution ranges, such as colonization/extinction dynamics, dispersal ability, local adaptation and historical factors.

4.1 Current distribution range

The selection of the modelling approach may have significant repercussions when considering a species as a genetically heterogeneous organism, whose geographical distribution of its genetic variation may have major implications for understanding the effects of GCC.

**FIGURE 2** (a) Predicted current distributions (year 2000) for each genetic cluster of Arabidopsis thaliana and methodology (Maxent, nonspatial and spatial hierarchical Bayesian models [HBMs]). Darker and lighter intensities indicate higher and lower suitability, respectively. (b) Uncertainty of nonspatial and spatial HBMs. Darker and lighter intensities indicate higher and lower uncertainty, respectively. Grey maps for genetic cluster C3 indicate that the spatial HBMs were not acceptable.
on its distribution range. In the case of Maxent, and of any modeling technique dealing with binary response variables, a major problem is the loss of data resulting from the conversion of a continuous variable (i.e., the genetic cluster membership proportions characterizing each individual) into a binary variable (i.e., the assignment of each individual to the genetic cluster with the highest membership proportion) (Gotelli & Stanton-Geddes, 2015; Marcer et al., 2016). In our study, binarization of genetic data had an important cost in terms of data loss as 82 of 301 accessions did not reach the minimum membership proportion of 0.5 required to be assigned to a genetic cluster. As a result, the number of accessions per genetic cluster used in Maxent was reduced (Figure 1b). In contrast, HBMs did not have this limitation and used the whole set of 301 accessions, and also including genetic admixture among the four genetic clusters detected in the Iberian Peninsula. It must be emphasized that accessions exhibiting genetic admixture are quite relevant in biological terms. For example, they may indicate the existence of contact zones between genetic lineages where important processes affecting the distribution range may take place, such as the balance between selection and dispersal against hybrids (Barton & Hewitt, 1985). For this reason, HBMs represent a better choice to model distribution ranges of intraspecific genetic lineages if it is undesirable to discard accessions with too much genetic admixture.

Broadly speaking, Maxent and the Bayesian modelling approaches were consistent in depicting the current geographical distribution of the four genetic clusters of Iberian A. thaliana (Figures 1 and 2). The exception was genetic cluster C3, in which nonspatial HBM blurred the distribution range and spatial HBM was not able to produce results due to unacceptable outcomes in a Bayesian framework. Interestingly, genetic cluster C3 is strongly differentiated from the other clusters found in the Iberian Peninsula, as well as across the whole species’ distribution range. In fact, genetic cluster C3 is considered as the relict cluster with a long evolutionary history (Brennan et al., 2014; Durvasula et al., 2017; Picó et al., 2008; The 1001 Genomes Consortium, 2016). The relict nature of genetic cluster C3 is also supported by its

### TABLE 2

| Cluster | HBM   | Model                        | MAE  | RMSE |
|---------|-------|------------------------------|------|------|
| C1      | Nonspatial | $Y \sim \beta_0 + \text{BIO}3 + \text{BIO}4 + \text{BIO}8 + \text{BIO15} + \text{BIO18}$ | 0.174 | 0.210 |
|         | Spatial | $Y \sim \beta_0 + \text{BIO}1 + \text{BIO}8 + W$ | 0.134 | 0.171 |
| C2      | Nonspatial | $Y \sim \beta_0 + \text{BIO}1 + \text{BIO}4 + \text{BIO}8 + \text{BIO12} + \text{BIO18}$ | 0.116 | 0.153 |
|         | Spatial | $Y \sim \beta_0 + \text{BIO}8 + \text{BIO}12 + \text{BIO15} + \text{BIO18} + W$ | 0.070 | 0.095 |
| C3      | Nonspatial | $Y \sim \beta_0 + \text{BIO}4 + \text{BIO12} + \text{BIO18}$ | 0.217 | 0.268 |
|         | Spatial | – | – | – |
| C4      | Nonspatial | $Y \sim \beta_0 + \text{BIO}1 + \text{BIO15} + \text{BIO18}$ | 0.148 | 0.189 |
|         | Spatial | $Y \sim \beta_0 + \text{BIO}1 + W$ | 0.096 | 0.128 |
scattered distribution across the Iberian Peninsula, a geographical distribution that is interpreted as the result of Iberian glacial refugia (Brennan et al., 2014; Marcer et al., 2016; Picó et al., 2008), whereas the rest of the genetic clusters exhibit geographically marked distributions, probably as a result of more recent demographic histories. Overall, this result indicates that modelling the distribution range of genetic clusters or species with scattered distributions may be difficult no matter what modelling approach is applied. For the particular case of genetic cluster C3, characterized by the high genetic membership proportions of their accessions and the relatively low admixture with other genetic clusters, Maxent predicts its distribution best.

For the remaining genetic clusters with marked geographical 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 geographical barrier (i.e., the Ebro river valley in NE Spain), which was clearly depicted by Maxent (Figure 2). Note that such a disjunct distribution is not a sampling problem, but is 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 (but did not totally erase) 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 a cluster’s distribution range increases its suitability to be modelled by alternative means.

Spatial HBMs, along with Maxent for genetic cluster C4, did not show residual SAC, which is a desirable property to avoid inaccurate parameter estimates and inadequate quantification of uncertainty (Beguin et al., 2012; Crase et al., 2014; Latimer et al., 2006; Record et al., 2013). In addition, spatial HBMs exhibited lower average model prediction errors than nonspatial HBMs. Hence, and from a purely statistical viewpoint, the higher rigour of spatial HBMs, in terms of higher accuracy and efficient removal of residual SAC, gives them a clear advantage (Crase et al., 2014; Swanson et al., 2013). Spatial HBMs also allowed us to assess the spatial effects on distribution range, which were quite compact and with high intensities at the distribution range centre (Figure 3). Such patterns may account for the lower number of bioclimatic predictors selected by spatial HBMs in comparison with nonspatial HBMs and Maxent. In fact, the reduction of predictors represents a common shortcoming of SDMs that, in some cases, may jeopardize the biological interpretation of the environmental factors underlying current distribution ranges (Beale et al., 2010; Swanson et al., 2013). We want to emphasize, 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 nonspatial HBMs (Figure 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 (Crase et al., 2014; Swanson et al., 2013). The rationale is that organisms exhibiting a high SAC 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 assessment of the SAC on distribution range shifts. Overall, it is clear that considering SAC adds realistic biological elements for understanding the long-term effects of GCC on biodiversity (Cardador, Sardà-Palomera, Carrete, & Mañosa, 2014; Crase et al., 2014; De Marco et al., 2008; Swanson et al., 2013). Nevertheless, it must be noted that we are assuming that SAC and its underlying forces remain relatively constant during climate change. Clearly, this assumption, although beyond the scope of this study, will need to be addressed in the future.

In general, the outcomes generated by the three modelling approaches for GCC scenarios were quite different. Such a disparity in model outcomes may indicate the differential effects of environmental drivers and the sources of SAC on the GCC response of genetic clusters, but also the effect of the geographical distribution of each genetic cluster on model performance. In fact, the problems affecting the modelling of current distribution ranges,
namely the disjunct and scattered geographical distributions of genetic clusters C2 and C3, respectively, also affected the predictions of distribution range shifts with GCC. For example, in the case of genetic cluster C2, spatial HBM predicted a rather continuous distribution in NE Spain when it is more reasonable to expect that the barrier separating the two major nuclei of populations on both sides of the Ebro river valley will be expanded with warming, as predicted by Maxent. Furthermore, the GCC effects on genetic cluster C3 are more difficult to predict. Although Maxent increased the potential distribution range of this genetic cluster over the Iberian Peninsula, as relict organisms exhibiting scattered distributions, the future of cluster C3 may simply depend on the effect of GCC on the preservation of its habitats as they are today. In fact, populations of genetic cluster C3 may exhibit strong local adaptation (Méndez-Vigo et al., 2013), constraining the ability of relict genotypes to colonize novel habitats.

In contrast, interpreting the problems of the GCC effects on distribution range shifts for the other two genetic clusters with continuous distributions was totally different. For example, genetic cluster C1 has a continuous distribution across the NE Iberian Peninsula, which is characterized by Atlantic and continental climates. Global climate change models excluding spatial effects, namely Maxent and nonspatial HBMs, indicate that GCC will restrict A. thaliana to northern and mountainous areas, which is a typical scenario of migration towards environments that will probably retain similar characteristics in the future. In contrast, spatial HBMs yielded a totally different outcome, indicating that genetic cluster C1 will maintain and even increase its current distribution range (Figure 4). The strong spatial effects detected by spatial HBMs for genetic cluster C1 may account for this result, as the response of A. thaliana to GCC is also expected to be more homogeneous. Recent experimental data from transplant experiments using accessions from this genetic cluster indicate that this scenario may be plausible, as A. thaliana performed well in warmer environments, highlighting the potential of this genetic cluster to cope with warming (Exposito-Alonso, Brennan, Alonso-Blanco, & Picó, 2018). The same applies to genetic cluster C4, which is also distributed continuously in the typically Mediterranean SW Iberian Peninsula. In this case, however, all modelling approaches predicted its expansions with GCC, although some discrepancies among modelling approaches were recorded (Figure 4). Although we lack experimental evidence of the effects of warming on the performance of C4 A. thaliana accessions, we believe that such expansion with GCC is highly probable, as genetic cluster C4 mostly occupies the warmest Iberian region.

5 | CONCLUSIONS

We developed hierarchical Bayesian Beta regression models to explore the current distribution range and its GCC-induced shifts for an organism with a marked geographical genetic structure, which represents the outcome of historical, ecological and evolutionary forces probably acting in concert. For this reason, the effects of GCC have to be understood as a mosaic of responses varying in extent and intensity determined by the complexity of the geographical genetic structure exhibited by study organisms. Rather than predicting mere contractions or expansions for a single organism, we should expect a reshuffling of the genetic diversity and its geographical structure with GCC, which is obviously more difficult to predict. The HBMs developed here enrich the toolbox of software available to deal with such expectation.

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

We conclude by stressing the importance of developing better models to forecast the effects of GCC on organisms’ distribution ranges worldwide. Such predictive tools, and the comparison thereof, may lead to mitigation of the inevitable impact of GCC on biodiversity. However, we need to increase our understanding 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 the manuscript.

AUTHOR CONTRIBUTIONS

F.X.P. and A.M. conceived and designed the study. C.A.-B. and F.X.P. collected the samples and analysed the 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.

**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.

**ORCID**

Joaquín Martínez-Minaya https://orcid.org/0000-0002-1016-8734
David Conesa https://orcid.org/0000-0002-5442-5691
Marie-Josée Fortin https://orcid.org/0000-0002-9935-1366
Carlos Alonso-Blanco https://orcid.org/0000-0002-4738-5556
F. Xavier Picó https://orcid.org/0000-0003-2849-4922
Arnald Marcer https://orcid.org/0000-0002-6532-7712

**REFERENCES**

Anderson, C. D., Epperson, B. K., Fortin, M. J., Holderegger, R., James, P. M. A., Rosenberg, M. S., ... Spear, S. (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology, 19*, 3565–3575. https://doi.org/10.1111/j.1365-294X.2010.04757.x

Bálint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., ... Nowak, C. (2011). Cryptic biodiversity loss linked to global climate change. *Nature Climate Change, 1*, 313–318. https://doi.org/10.1038/NCLIMATE1191

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics, 16*, 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010). Regression analysis of spatial data. *Ecology Letters, 13*, 246–264. https://doi.org/10.1111/j.1466-824X.2009.01422.x

Beguin, J., Martino, S., Rue, H., & Cumming, S. G. (2012). Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. *Methods in Ecology and Evolution, 3*, 921–929. https://doi.org/10.1111/j.2041-210X.2012.00211.x

Bellard, C., Bertelmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters, 15*, 365–377. https://doi.org/10.1111/j.1466-0248.2011.01736.x

Belmaker, J., Zarnetske, P., Tuanmu, M.-N., Zonneveld, S., Record, S., Strecker, A., & Beaudrot, L. (2015). Empirical evidence for the scale dependence of biotic interactions. *Global Ecology and Biogeography, 24*, 750–761. https://doi.org/10.1111/geb.12311

Bivand, R. S., Hauke, J., & Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis, 45*, 150–179. https://doi.org/10.1111/gean.12008

Bivand, R. S., & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software, 63*, 18. https://doi.org/10.18637/jss.v063.i18

Brennan, A. C., Méndez-Vigo, B., Haddoui, A., Martínez-Zapater, J. M., Picó, F. X., & Alonso-Blanco, C. (2014). The genetic structure of *Arabidopsis thaliana* in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC Plant Biology, 14*, 17. https://doi.org/10.1186/1471-2229-14-17

Cardador, L., Sardà-Palomera, F., Carrete, M., & Mañosa, S. (2014). Incorporating spatial constraints in different periods of the annual cycle improves species distribution model performance for a highly mobile bird species. *Diversity and Distributions, 20*, 515–528. https://doi.org/10.1111/dad.12156

Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science, 333*, 1024–1026. https://doi.org/10.1126/science.1206432

Crase, B., Liedloff, A., Ves, P. A., Fukuda, Y., & Wintle, B. A. (2014). Incorporating spatial autocorrelation into species distribution models alters forecasts of climate-mediated range shifts. *Global Change Biology, 20*, 2566–2579. https://doi.org/10.1111/gcb.12598

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in *R*. *Journal of Statistical Software, 34*, 1–24. https://doi.org/10.18637/jss.v034.i02

Dale, M. R. T., & Fortin, M.-J. (2002). Spatial autocorrelation and statistical tests in ecology. *Ecology Science, 9*, 162–167. https://doi.org/10.1080/11956860.2002.11682702

Dale, M. R. T., & Fortin, M.-J. (2014). Spatial analysis – A guide for ecologists (2nd ed.). Cambridge, UK: Cambridge University Press.

D’Amen, M., Zimmermann, N. E., & Pearman, P. B. (2013). Conservation of phylogeographic lineages under climate change. *Global Ecology and Biogeography, 22*, 93–104. https://doi.org/10.1111/j.1466-8238.2012.00774.x

De Marco, P., Diniz-Filho, J. A. F., & Bini, L. M. (2008). Spatial analysis improves species distribution modelling during range expansion. *Biological Letters, 4*, 577–580. https://doi.org/10.1098/rsbl.2008.0210

de Oliveira, G., Rangel, T. F., Lima-Ribeiro, M. S., Terribile, L. C., & Diniz-Filho, J. A. F. (2014). Evaluating, partitioning, and mapping the spatial autocorrelation component in ecological niche modeling: A new approach based on environmentally equidistant records. *Ecography, 37*, 637–647. https://doi.org/10.1111/1600-0587.2013.00564.x

Diniz-Filho, J. A. F., Barbosa, A. C. O. F., Collevatti, R. G., Chaves, L. J., Terribile, L. C., Lima-Ribeiro, M. S., & Telles, M. P. C. (2016). Spatial autocorrelation analysis and ecological niche modelling allows inference of range dynamics driving the population genetic structure of a Neotropical savanna tree. *Journal of Biogeography, 43*, 167–177. https://doi.org/10.1111/jbi.12622

Dirnböck, T., & Dullinger, S. (2004). Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation Science, 15*, 77–84. https://doi.org/10.1111/j.1654-1103.2004.tb02239.x

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography, 16*, 129–138. https://doi.org/10.1111/j.1466-8238.2006.00279.x

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., ... Hancock, A. M. (2017). African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America, 114*, 5213–5218. https://doi.org/10.1073/pnas.161763114

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics, 40*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of Maxent for ecologists. *Diversity and Distributions, 17*, 43–57. https://doi.org/10.1111/j.1472-444X.2010.00725.x

Exposito-Alonso, M., Brennan, A. C., Alonso-Blanco, C., & Picó, F. X. (2018). Spatio-temporal variation in fitness responses to contrasting environments in *Arabidopsis thaliana*. *Evolution, 72*, 1570–1586. https://doi.org/10.1111/evo.13508
