

Geographical restructuring of *Arabidopsis thaliana*'s genetic makeup in the Iberian Peninsula due to climate change based on genetic cluster membership

Arnald Marcer^{a, b}, Marie-Josée Fortin^c and F. Xavier Picó^d

^aCREAF, Cerdanyola del Vallès 08193, Spain

^bUniv. Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain

^cDept. of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 3G5, Canada

^dDepartamento de Ecología Integrativa, Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), Sevilla, Spain

CORRESPONDENCE AUTHOR

Name: Arnald Marcer

Address: CREA, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain.

Email: arnald.marcer@uab.cat

RUNNING TITLE: *Arabidopsis thaliana* genetic restructuring in the Iberian Peninsula

ABSTRACT

Aim To assess the effects of climate change on genetic lineages of *Arabidopsis thaliana* at the *admixed* population level by directly modelling genetic cluster membership values to predict potential genetic cluster memberships across the Iberian Peninsula.

Location Iberian Peninsula

Methods We used a dataset of 274 accessions structured in four genetic clusters as inferred from 250 nuclear single-nucleotide polymorphisms with Bayesian clustering methods. We predicted the change in percentages of genetic cluster membership at a population level and the changes in potential suitability across the study area by combining parametric (Beta regression) and non-parametric (Recursive trees) methods.

Results Climate change will affect genetic lineages of *Arabidopsis thaliana* differently. Genetic clusters GC1 and GC2 will suffer a substantive reduction of their respective suitable areas while GC3 and GC4 will expand northward. At the population level, except for GC4, the rest of the lineages will undergo a genetic turnover for many of their populations.

Main conclusions *A. thaliana* in the Iberian Peninsula will undergo a major internal genetic restructuring and range change due to climate change. Genetic lineages of *Arabidopsis thaliana* in the Iberian Peninsula will be affected differently which reinforce the need for taking into account intraspecific genetic variation when modelling species distribution. Despite limited predictive power of individual statistical models, the combination of distinct models can compensate this shortcoming.

KEYWORDS

Arabidopsis thaliana, intraspecific genetic diversity, climate change, distribution modelling, beta regression, recursive trees, Iberian Peninsula, genetic clusters

INTRODUCTION

The intraspecific response of species to climate change is a major research objective in biogeography, evolutionary ecology and conservation (Bellard et al., 2012; Cardinale et al., 2012; Hoffmann et al., 2015) and a necessary step to design better species conservation actions and strategies. Yet, we need to factor in intraspecific genetic variation since a species is an aggregate of different genetic lineages which differ in their adaptation to environmental conditions. However, the effects of climate change on genetic diversity are just beginning to be explored (Bellard et al., 2012). Indeed, climate change can profoundly affect species distributions (Bakkenes et al., 2002; Thomas et al., 2004; Thuiller et al., 2005) as it has already been documented (Davis & Shaw, 2001; Parmesan & Yohe, 2003; Parmesan, 2006). As temperature is expected to rise over the 21st century under all studied emission scenarios (IPCC, 2014), it will alter species distribution (Loarie et al., 2009). We need to predict these changes if we want to take adequate actions to mitigate future losses of biodiversity (Pfenninger et al., 2012). There are abundant examples in the scientific literature of tools and methodologies aimed at predicting species distributions in the face of global climate change (e.g. Aitken et al., 2008; Guisan & Zimmermann, 2000; Franklin et al., 2012; Pfenninger et al., 2012; Razgour 2015). However, most studies have focused on forecasting distributions at the species level and not taking into account for species internal genetic variability (Bálint et al., 2011; Benito-Garzón et al., 2011; Alsos et al., 2012; Pauls et al., 2013). Within a species, this genetic variability manifests itself as an assemblage of lineages or clusters differing in their degree of adaptation (Sork et al., 2010; Neiva et al., 2015). Such genetic clusters can be inferred from genetic markers such as microsatellites or single-nucleotide polymorphisms (SNPs) using clustering algorithms such as STRUCTURE (Pritchard et al., 2000). Hence, distribution models at the species level will be necessarily less precise than models at the intraspecific genetic level. The knowledge of this intraspecific genetic diversity may give quite a different vision on the projected distributions of species under climate change (Jump et al., 2009; Benito-Garzón et al., 2011; Yannic et al., 2014; Marcer et al., 2016) and be fundamental in conservation.

Data on intraspecific genetic lineages come as proportions (i.e. percentages of genetic cluster or lineage membership for admixed populations) instead of the presence/absence data used in traditional SDM techniques. However, despite the existence of notable studies that try to infer species distributions by working at the genetic level (e.g. Razgour et al., 2013; Valladares et al., 2014; Yannic et al., 2014), few studies (e.g. Jay et al., 2012) have directly used these proportion-type data to model the distribution of intraspecific units as suggested by Gotelli & Stanton-Geddes (2015). More studies are much in need in this respect.

A model species well suited for exploring this issue is the annual plant *Arabidopsis thaliana* (L.) Heynh (*Brassicaceae*). Indeed, for *A. thaliana*, an extensive georeferenced database has already been collected for the Iberian Peninsula (Picó et al., 2008; Manzano-Piedras et al., 2014) which is a major area of genetic diversity in the global distribution of this species (Brennan et al., 2014). The Iberian Peninsula is an environmentally diverse region where *A. thaliana* shows local adaptation to the different environments as this region was a glacial refugium (Picó et al., 2008). It comprises a wide range of climates and *A. thaliana* shows great variation in response to these different conditions (Bouchapke et al., 2008; Picó et al., 2008; Hancock et al., 2011; Fournier-Level et al., 2011; Manzano-Piedras et al., 2014). This local adaptation is reflected in the intraspecific genetic clusters of this species (Hancock et al., 2011; Fournier-Level et al., 2011).

Here, we analyse how the distribution of *Arabidopsis thaliana* could be affected by climate change at the intraspecific genetic level in the Iberian Peninsula. We used *admixed* genetic lineages inferred from a set of 250 SNPs analysed in 274 accessions spread across the Iberian Peninsula. Our objectives are *a*) to assess the effects of climate change on genetic lineages at the *admixed* population level by directly modelling genetic cluster membership values and *b*) to predict potential genetic cluster membership values across the Iberian Peninsula. Our work complements and expands comparable studies (Jay et al., 2012; Gotelli & Stanton-Geddes, 2015) and aims at modelling genetic diversity directly from values of membership to genetic clusters.

MATERIALS AND METHODS

Study area

The Iberian Peninsula is located in the western part of Eurasia (ca. between 10 W and 4 E, and 36 S and 44 N). It is a mountainous region of circa 580 000 km² and one of the major biodiversity hotspots in the Mediterranean. It harbors around 50% of the European plant species and 31% of the European endemic plants (Williams et al., 2000; Araújo et al., 2007) and has been a major Pleistocene glacial refugium (Hewitt, 2001; Gomez and Lunt, 2006).

Species, accessions and genetic data

We used a dataset of 274 accessions of *Arabidopsis thaliana* (L.) Heynh. derived from the one described in Marcer et al. (2016) and which includes additional climate predictors. Genetic data consist of percentages of cluster membership assignment derived with the STRUCTURE algorithm (Pritchard et al., 2000), version v.2.2, from a set of 250 nuclear genetic polymorphisms (SNPs) as described in Manzano-Piedras et al. (2014).

Climate data

Present time

A set of eight bioclimatic variables relevant to the species' ecology were selected as model predictors: BIO1 (Annual Mean Temperature), BIO2 (Mean Diurnal Range), BIO3 (Isothermality), BIO4 (Temperature seasonality), BIO8 (Mean Temperature of Wettest Quarter), BIO12 (Annual precipitation), BIO15 (Precipitation seasonality) and BIO18 (Precipitation of Warmest Quarter) (see Table ST1 in Supplementary Information). These variables were derived from the Digital Climatic Atlas of the Iberian Peninsula (http://opengis.uab.es/wms/iberia/en_index.htm) using the *dismo* package in R (Hijmans et al., 2015). Data were accessed on February 19, 2015. Their pairwise degree of collinearity (Spearman's correlation coefficient) is lower than 0.7.

Future scenarios of climate change

We selected the 2070 RCP2.6 and RCP8.5 climate change scenarios (Moss et al., 2008), which represent the minimum and maximum trends in radiative forcing (van Vuuren et al., 2011) of the four RCP scenarios adopted by IPCC in its fifth Assessment Report (AR5) (IPCC, 2014) as modelled by four different climate change models: HadGEM2-ES (Met Office Hadley Centre, UK), MRI-CGCM3 (Meteorological Research Institute, Japan), MIROC-ESM (Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies), NorESM1-M (Norwegian Climate Centre, Norway). Data were downloaded from the WorldClim website (<http://www.worldclim.org>) on February 19, 2015. We reprojected them from their original WGS84 Lat/Long (EPSG:4326) projection into ETRS89/LAEA (EPSG: 3035) equal-area projection and resampled them from 30 seconds to 1 km resolution using bilinear interpolation. Finally, for each of the two RCP scenarios we averaged the four models to generate our GCC dataset for predictive purposes. We will refer to these simply as RCP2.6 and RCP8.5.

Modelling approach

We used two different statistical models, namely, a parametric beta regression and a non-parametric regression tree algorithm, as implemented in R packages *betareg* and *mvpart*, respectively (Cribari-Neto & Zeileis, 2010; Therneau & Atkinson, 2014). This combination of very different techniques allowed us to check for agreements and disagreements between them to make predictions more robust. We are assuming that populations are, at least, partially adapted to their local climate conditions and that their admixture setup is not exclusively due to demographic processes; a reasonable assumption for *Arabidopsis thaliana* (Picó et al., 2008;

Hancock et al., 2011; Fournier-Level et al., 2011). We modelled each genetic cluster separately, using their cluster membership percentages as dependent variables and the above mentioned climate variables as independent variables. No interaction terms were used. The model formula is as follows:

$$Y_{[GC1, GC2, GC3, GC4]} = \beta_0 + \beta_1 \cdot \text{BIO1} + \beta_2 \cdot \text{BIO2} + \beta_3 \cdot \text{BIO3} + \beta_4 \cdot \text{BIO4} + \beta_5 \cdot \text{BIO8} + \beta_6 \cdot \text{BIO12} + \beta_7 \cdot \text{BIO15} + \beta_8 \cdot \text{BIO18}$$

The dependent variable Y is the cluster membership coefficient for each of the four genetic clusters and can take values ranging from 0 to 1 which express the admixed degree of membership to the given genetic cluster. β_0 through β_8 are the regression coefficients and BIO1/2/3/4/8/12/15/18 are the climate predictors.

To validate models and prevent overfitting, we used 10-fold cross-validation and used the average of the mean absolute error and of the pseudo- R^2 of test folds as measures of predictive performance. Finally, we fitted the models again using the whole dataset and used these final models to make predictions. For each model, we calculated Moran's I at five distance intervals (2000, 4000, 6000, 8000 and 10 000 m) in order to assess residual spatial autocorrelation assessing its significant using 10000 randomizations (R package *ncf* (Bjornstad, 2013)). Spatial autocorrelation was calculated both at the variable level (vSAC) and at the models' residuals level (rSAC) in order to evaluate whether models managed to decrease already present vSAC.

Prediction under global climate change

At population level

We calculated the cluster membership percentage change as predicted by each modelling method for each population and genetic cluster. We then checked for each population and climate change scenario whether both modelling methods agreed on the direction of change; *i.e.* whether a given cluster membership value either increases or decreases. For those populations in which models did not agree we did not trust the predicted changes in genetic cluster membership percentages. For the rest of populations in which the models did agree we predicted the future genetic cluster membership value as the weighted average pseudo- R^2 of both modelling techniques. After that, we assessed the relative dominance of each genetic cluster in each population in order to highlight those populations for which their dominant genetic cluster changed. Finally, we assessed the degree of intraspecific composition populational change between current and future climate conditions by using the Pearson correlation coefficient between current and future genetic cluster membership percentages to give a global measure of population structure change, as in Jay et al. (2012).

At Iberian Peninsula level

We projected both models to the whole of the Iberian Peninsula at present time climate conditions and under both RCP2.6 and RCP8.5 climate change scenarios. As with the predictions at population level we used the weighted average pseudo- R^2 of both predictions to give a final map of potential genetic cluster membership suitability under climate change. We considered as unknown those areas similar to the environmental conditions of the populations for which models did not agree (greyed-out areas in Figure 3). We used the Gower's dissimilarity coefficient (Gower, 1971) as implemented in R package *StatMatch* (D'Orazio, 2015) and determined as unpredictable those areas with a value below 0.025.

Changes in potential distribution

In order to quantify the potentially suitable area for each genetic cluster and to compare its extent between present time and climate change scenarios, we used a threshold of 0.5 as a cut-off point for determining whether any given location is either suitable or unsuitable for each genetic cluster. We then calculated the potentially suitable area for each genetic cluster and scenario of climate change as well as additional measures of change such as the mean and median of the suitability scores.

RESULTS

Model performance and residual spatial autocorrelation

Overall, regression trees models showed consistent better predictive performance than beta regression ones for both mean absolute error (mae) and pseudo- R^2 (hereafter pr^2) metrics (Table ST2). For all genetic clusters, beta regression results in an average mae of 0.158 and average pr^2 of 0.263 while regression trees result in an average mae of 0.126 and average pr^2 of 0.320. The best predicted genetic cluster was GC2 for both regression trees (mae: 0.081 ± 0.013 , pr^2 : 0.561 ± 0.142) and beta regression (mae: 0.103 ± 0.016 , pr^2 : 0.468 ± 0.142). On the other hand, GC3 had the highest error in the case of beta regression models (mae: 0.205 ± 0.027 , pr^2 : 0.103 ± 0.102), while for regression trees it was GC1 according to mae (0.163 ± 0.025) and GC3 according to pr^2 (0.144 ± 0.116).

Tables ST3-1/4 and figures SF3-1/4 show results of the spatial autocorrelation (SAC) analysis. At the variable level (vSAC), it was always highest at the second distance class (4 000 m) for all genetic clusters with the exception of GC4, for which it was the third distance class (6 000 m). Except for 5 cases (GC1-Beta-4000, GC2-Beta-4000, GC2-RT-4000, GC4-Beta-8000 and GC4-RT-8000) residual SAC (rSAC) was always lower than vSAC, meaning that the

models were able to reduce inherent vSAC in most cases. For GC1 and GC4 rSAC was not significant in all models and distance classes except for one case each. GC3 rSAC was significant in two distance classes of the beta model and one distance class of the regression tree model. Beta regression for GC4 only had one class with non-significant rSAC but the regression tree model rSAC was only significant in two of them.

Predictions at the population level

Model comparison and prediction of change for genetic cluster membership percentages

Table 1 and Figure 1 summarize the degree of agreement between both modelling methods in determining the direction of change in genetic cluster membership percentages per population and climate change scenario. In most of the populations and for both climate change scenarios, both beta regression and regression tree models coincide in determining the direction of change, i.e. whether a given population increases or decreases its membership percentage for a given genetic cluster. For scenarios RCP2.6 and RCP8.5 the maximum coincidence was in the case of GC1, for which they only differed in 25 populations (9.1 %), while the minimum coincidence for RCP2.6 was in the case of both GC2 and GC4 with 50 populations each (18.2 %), and for RCP8.5 was in the case of GC4 with 63 populations (23.4 %). Considering only the populations for which models agreed in the direction of change, we can state that for scenario RCP2.6, GC1 had the biggest gain ($+0.169 \pm 0.100$) and GC3 the biggest loss (-0.456 ± 0.273), while for scenario RCP8.5 the biggest gain was for GC3 ($+0.214 \pm 0.172$) and the biggest loss also for GC3 (-0.434 ± 0.268). In figure 1 populations for which models do not agree are marked with a black diamond.

Genetic cluster turnover

Figure 2 shows for each climate change scenario and genetic cluster which populations are predicted to have their dominant genetic cluster changed. Numbers are calculated taking only into account those populations for which models agree (black dots in Figure 2), which is not applicable to present time. All genetic clusters but GC4 will suffer a switch in suitability for some of their dominant populations; i.e. their present time dominant cluster will be switched to another one. Overall, taking into account all shifts from every genetic cluster at present time, GC1 will go from 148 dominated populations to 132 in RCP2.6 and 60 in RCP8.5 with 16 populations of uncertain fate in RCP2.6 and 12 in RCP8.5. GC2 will go from its current 57 populations to 43 in RCP2.6 and 58 in RCP8.5 with 18 populations of uncertain fate in RCP2.6 and 15 in RCP8.5. GC3 will go from its current 38 populations to only 9 in RCP2.6 and 55 in RCP8.5 with no populations of uncertain fate in RCP2.6 and 3 in RCP8.5. Finally, GC4 will be

the only genetic cluster which will keep all its present time dominated populations in both climate change scenarios and will increase from its current 31 to 51 in RCP2.6 (with 5 of uncertain fate) and to 67 in RCP8.5 (with 4 of uncertain fate).

Figure SF4 and SF5 show the degree of genetic structure change that each population must undergo to be best suited for its new environment. Approximately a fifth of existing populations will have to undergo a major structural change: 17.3% in the case of RCP2.6 and 21% in the case of RCP8.5 (Table ST4).

Predictions at the level of the Iberian Peninsula

Table 2 shows the potentially suitable area, the mean and the median for each genetic cluster at present time and future scenario as a way to measure the overall change across the Iberian Peninsula (Figure 3) that each genetic cluster would undergo in each scenario of climate change. Genetic clusters GC1 and GC2 show a reduction in mean suitability. For RCP2.6 their means get reduced from 0.314 to 0.270 and from 0.208 to 0.204 and their potential distribution area gets reduced by 50.8% and 87.5%, respectively. On the other hand, GC3 and GC4 would undergo a general increase in their suitability to climate change. Although GC3 would decrease its mean from 0.225 to 0.217 for RCP2.6, for RCP8.5 it would increase up to 0.257. With respect to potential distribution area, it would increase by 835.0% for RCP2.6 and by 4887.4% for RCP8.5. Finally, land suitability for GC4 would also be substantially increased; from a mean value of 0.253 to 0.302 for RCP2.6 and up to 0.369 for RCP8.5 and an increase of potential distribution area of 175.2% for RCP2.6 and of 238.1% for RCP8.5.

The spatial distribution of these changes is shown in Figure 3 where there are clear trends northward. Higher values of suitability for GC1 get more constrained to the most northern part of the Iberian Peninsula both for RCP2.6 and RCP8.5. GC2 shows a shift towards the north-east of its climate suitability. GC3 moves towards central parts of the Iberian Peninsula and GC4 again shows a clear northerly expansion of climate suitability. Supplementary materials provide Figures SF6 and SF7-1/4 which show future predictions differentiated by modelling method.

DISCUSSION

Our results confirm the need for dealing with intraspecific genetic variation in order to understand and forecast the effects of climate change on species distributions as has been suggested (Hancock et al., 2011; Fournier-Level et al., 2011). We have shown that the impact of climate change is unevenly felt by the different genetic lineages of *Arabidopsis thaliana* in the Iberian Peninsula, an expected result given that these lineages are known to have local adaptation and their relative percentage mixture in populations is geographically structured in

the Iberian Peninsula (Picó et al., 2008). Moreover, it is possible to identify those populations which, for their particular genetic makeup, will need to undergo a major structural genetic change from those which may have a minor impact. This is a potentially useful outcome since it can help in optimizing scarce resources when undertaking conservation efforts.

Species distribution models should therefore consider genetic variation as has been suggested (Jump et al., 2009; Benito-Garzón et al., 2011; Yannic et al., 2016; Marcer et al., 2016). Although for the vast majority of species this kind of information is unavailable and traditional SDM techniques might be the only option, these results clearly suggest that, when possible, this should be the way to proceed. Also, given the fast pace in which genetics and genomics technologies progress (Hoffman et al., 2015; Tyler-Smith et al., 2015) it is not unthinkable that in a not so distant future this could be done for most species.

It is important to recognize that these kinds of analyses forget an important component, demography. As shown, we can give a snapshot of what might happen to a given species in the face of rising temperatures and shifting climate regimes. However, this is done as a single leap forward into the future without considering the population dynamics and resulting demographic changes which would occur if a sequence of many events in a real-time path to the projected 2070 year was considered. Although future research should try to incorporate demographic processes at the landscape level into the modelling process, this is not an easy task. The difficulty resides not only in model development but more importantly in the lack of sufficient quality data (Ehrlén & Morris, 2015) on species life history traits and populations and in the stochasticity, inherent to some processes like long-distance dispersal events (Pergl et al., 2011). In order to project spatially explicit models in time we need data on local demographic parameters over broad areas (Nathan 2006; Thuiller et al., 2014), which is usually unavailable. Even in the case of *Arabidopsis thaliana*, a model species and one of the better-known species, there is no reliable field-obtained data on important traits, such as population size, seed production, seedling survival or dispersal distance of natural populations over broad areas, which would be needed to build such models. A possibility is that of running simulations but this has its own caveats, too. Without actual data, their outcomes are very dependent on the value of the parameters that are fed to them and, thus, their results difficult to interpret.

The expected effect of climate change on the genetic lineages of *Arabidopsis thaliana* in the Iberian Peninsula is in accordance to the work of Marcer et al. (2016). Genetic clusters GC3 and GC4 will increase their potential distribution by expanding northward their distribution ranges. On the other hand, GC1 and GC2 will shrink their distribution ranges and their highest suitable areas will be more constrained to the north and north-east, respectively. At the population level, *Arabidopsis thaliana* will undergo a genetic turnover in many of its populations which will need

to shift from one dominant genetic cluster to another. Either dispersal, local re-adaptation or phenotypic plasticity must come to the rescue if most populations are to survive these changes induced by climate change. For instance, as shown by Picó et al. (2014), flowering time, which is mediated by temperature, appears to be the major life-history trait adjusting *Arabidopsis thaliana* to the different environmental conditions across the Iberian Peninsula. Plasticity in this and other traits may provide for some necessary time for populations to adapt (Donohue et al., 2005; Levin, 2009; Nicotra et al., 2010).

We have modelled genetic cluster membership percentages to infer potential future distributions at the intraspecific level rather than resorting to the use of thresholds to transform data into presence/absence and use SDM techniques. This work adds to the need for more research in this respect (Jay et al., 2012; Gotelli & Stanton-Geddes, 2015). We used two very different modelling approaches, a parametric beta regression and a non-parametric regression tree analysis, which showed a high degree of agreement in their predictions. Despite, their individual predictive power with a set of only-climate predictors was limited, their combined use allowed us to identify which predicted changes could be trusted and use this information to make predictions with highlighted populations and areas of uncertainty. The limited predictive power should not be attributed to the statistical modelling techniques themselves but it probably reflects the lack of important predictors such as land use and soil type (Marcer et al., 2016), a price currently needed to pay due to the lack of climate change models for these types of variables.

Finally, we would like to express the need for quality data if we are to understand the biological and environmental processes that drive species distributions and predict the fate of species in front of climate change. The building of quality and extensive datasets on natural populations such as the one used in this study require substantial resources in terms of time and funding. Yet, they are paramount to such undertakings. The lack of sufficient funding for such basic research and of community recognition for data building and curation is a major handicap which hinders advancement in this respect. The authors encourage funding agencies not to dismiss this pressing need.

ACKNOWLEDGEMENTS

A.M. and M.J.F were partially supported by the NEWFORESTS project (PIRSES-GA-2013-612645) from the European 7FP. F.X.P. and A.M. were partially supported by Ministerio de Economía y Competitividad of Spain (CGL2012-33220/) and the Agency for Management of University and Research Grants of the Generalitat de Catalunya (2014-SGR-913). F.X.P. acknowledges the Severo Ochoa Programme for Centres of Excellence in R+D+I (SEV-2012-0262) in Spain.

REFERENCES

- Alsos, I. G., Ehrich, D., Thuiller, W., Eidesen, P. B., Tribsch, A., Schönswetter, P., Lagaye, C., Taberlet, P. & Brochmann, C. (2012) Genetic consequences of climate change for northern plants. *Proceedings of the Royal Society B: Biological Sciences*, **279**(1735), 2042–2051.
- Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. & Curtis-McLane, S. (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**(1), 95-111.
- Araújo, M.B., Lobo, J.M., & Moreno, J.C. (2007) The effectiveness of Iberian protected areas in conserving terrestrial biodiversity. *Conservation Biology*, **21**, 1423–1432.
- Bakkenes, M., Alkemade, J. R. M., Ihle, F., Leemans, R. & Latour, J. B. (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, **8**(4), 390-407.
- Bálint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., Theissinger, K., Pauls, S.U. & Nowak, C. (2011) Cryptic biodiversity loss linked to global climate change. *Nature Climate Change*, **1**(6), 313-318.
- Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters*, **15**(4), 365-377.
- Benito Garzón, M., Alía, R., Robson, T. M. & Zavala, M.A. (2011) Intra-specific variability and plasticity influence potential tree species distributions under climate change. *Global Ecology and Biogeography*, **20**(5), 766-778.
- Bjornstad, O. N. (2013) *ncf*: spatial nonparametric covariance functions. R package version 1.1-5, <https://CRAN.R-project.org/package=ncf>
- Bouchabke, O., Chang, F., Simon, M., Voisin, R., Pelletier, G. & Durand-Tardif, M. (2008). Natural variation in *Arabidopsis thaliana* as a tool for highlighting differential drought responses. *PLoS ONE*, **3**(2), e1705
- Brennan A.C., Méndez-Vigo B., Haddioui 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.

- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P.,
Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M.,
Grace, J.B., Larigauderie, A., Srivastava, D.S. & Naeem, S. (2012). Biodiversity loss and its
impact on humanity. *Nature*, **489**(7415), 59-67
- Cribari-Neto, F. & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*
34(2), 1-24. URL <http://www.jstatsoft.org/v34/i02/>.
- Davis, M. B. & Shaw, R. G. (2001). Range shifts and adaptive responses to Quaternary
climate change. *Science*, **292**(5517), 673-679.
- Donohue K., Dorn L., Griffith C., Kim E., Aguilera A., Polisetty C. and Schmitt J. (2005)
The Evolutionary Ecology of Seed Germination of *Arabidopsis thaliana* : Variable Natural
Selection on Germination Timing. *Evolution*, **59**(4), 758–770.
- D’Orazio, M. (2015). StatMatch: Statistical Matching, R package version 1.2.3,
<https://CRAN.R-project.org/package=StatMatch>.
- Ehrlén J. & Morris W.F. (2015) Predicting changes in the distribution and abundance of
species under environmental change. *Ecology letters*, **18**: 303-314
- Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J. & Wilczek, A. M.
(2011) A Map of Local Adaptation in *Arabidopsis thaliana*. *Science*, **334**(6052), 86-89.
- Franklin, J., Davis, F. W., Ikegami, M., Syphard, A. D., Flint, L. E., Flint, A. L. & Hannah,
L. (2013) Modeling plant species distributions under future climates: How fine scale do climate
projections need to be? *Global Change Biology*, **19**(2), 473-483.
- Gomez, A., & Lunt D. (2006) Refugia within refugia: patterns of phylogeographic
concordance in the Iberian Peninsula, pp. 155–188 in *Phylogeography of Southern European
Refugia* (ed. by S. Weiss and N. Ferrand). Springer, Dordrecht, The Netherlands.
- Gotelli, N. J., & Stanton-Geddes, J. (2015) Climate change, genetic markers and species
distribution modelling. *Journal of Biogeography*, **42**(9), 1577-1585.
- Gower, J. C. (1971) A general coefficient of similarity and some of its
properties. *Biometrics*, **27**, 623–637.

- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.
- Hewitt, G. M. (2001) Speciation, hybrid zones and phylogeography -- or seeing genes in space and time. *Molecular Ecology*, **10**: 537-549.
- Hijmans, R., Phillips, S., Leathwick J. & Elith, J. (2015) *dismo*: Species Distribution Modeling, R package version 1.0-12. <https://CRAN.R-project.org/package=dismo>
- Hoffmann A., Griffin P., Dillon S., Catullo R., Rane R., Byrne M., Jordan R., Oakeshott J., Weeks A., Joseph L., Lockhart P., Borevitz J., & Sgrò C. (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, **2**(1):1-24.
- Jay, F., Manel, S., Alvarez, N., Durand, E. Y., Thuiller, W., Holderegger, R., Taberlet, P. & François, O. (2012). Forecasting changes in population genetic structure of alpine plants in response to global warming. *Molecular Ecology*, **21**(10), 2354-2368
- Jump, A. S., Marchant, R. & Peñuelas, J. (2009). Environmental change and the option value of genetic diversity. *Trends in Plant Science*, **14**(1), 51-58.
- Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., Toomajian, C., Roux, F. & Bergelson, J. (2011). Adaptation to Climate Across the Arabidopsis thaliana Genome. *Science*, **334**(6052), 83-86.
- IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. (ed by Core Writing Team, R.K. Pachauri & L.A. Meyer), 151 pp., IPCC, Geneva, Switzerland
- Levin D.A. (2009) Flowering-time plasticity facilitates niche shifts in adjacent populations. *New Phytologist*, **183**, 661-666.
- Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B. & Ackerly, D. D. (2009). The velocity of climate change. *Nature*, **462**(7276), 1052-1055.
- Manzano-Piedras, E., Marcer, A., Alonso-Blanco, C. & Picó, F. X. (2014). Deciphering the adjustment between environment and life history in annuals: Lessons from a geographically-explicit approach in Arabidopsis thaliana. *PLoS ONE*, **9**(2).

Marcer A., Méndez-Vigo B., Alonso-Blanco C. and Picó F.X. (2016). Tackling intraspecific genetic structure in distribution models better reflects species geographical range. *Ecology and Evolution*. doi: 10.1002/ece3.2010.

Moss, R., Babiker, S.B., Calvo, E. et al. (2008) Towards New Scenarios for Analysis of Emissions, Climate Change, Impacts, and Response Strategies. Intergovernmental Panel on Climate Change, Geneva, 132 pp.

Nathan, R. (2006) Long-Distance Dispersal of Plants. *Science*, **313**, 786–788.

Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F., & van Kleunen, M. (2010) Plant phenotypic plasticity in a changing climate. *Trends in Plant Science*, **15**, 684–692.

Pergl, J., Müllerová, J., Perglová, I., Herben, T. and Pysěk, P. (2011). The role of long-distance seed dispersal in the local population dynamics of an invasive plant species. *Diversity and Distributions*, **17**: 725-738

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.

Neiva, J., Assis, J., Coelho, N. C., Fernandes, F., Pearson, G. A. & Serrão, E. A. (2015) Genes Left Behind: Climate Change Threatens Cryptic Genetic Diversity in the Canopy-Forming Seaweed *Bifurcaria bifurcata*. *PLOS ONE*, **10**(7), e0131530.

Parmesan, C. & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**(6918), 37-42.

Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, **37**(1), 637-669.

Pauls, S. U., Nowak, C., Bálint, M. & Pfenninger, M. (2013) The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, **22**(4), 925-946.

Pfenninger, M., Bálint, M. & Pauls, S. U. (2012). Methodological framework for projecting the potential loss of intraspecific genetic diversity due to global climate change. *BMC Evolutionary Biology*, **12**:224

- Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**(2), 945-959.
- Razgour O., Salicini I., Ibáñez C., Randi E., & Juste J. (2015) Unravelling the evolutionary history and future prospects of endemic species restricted to former glacial refugia. *Molecular ecology*, **24**, 5267–5283.
- Sork, V.L., Davis, F.W., Westfall, R., Flint, A., Ikegami, M., Wang, H. & Grivet, D. (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Molecular Ecology*, **19**, 3806-3823.
- Therneau, T.M. & Atkinson, B. (2014). rpart, R port of rpart by Brian Ripley <ripley@stats.ox.ac.uk>. Some routines from vegan -- Jari Oksanen <jari.oksanen@oulu.fi> Extensions and adaptations of rpart to mvpart by Glenn De'ath. (2014). mvpart: Multivariate partitioning. R package version 1.6-2. <https://CRAN.R-project.org/package=mvpart>
- Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., Ferreira de Siqueira, M., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L. & Williams, S. E. (2004). Extinction risk from climate change. *Nature*, **427**(6970), 145-148.
- Thuiller, W., Lavorel, S., Araujo, M. B., Sykes, M. T. & Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences*, **102**(23), 8245-8250.
- Thuiller, W., Münkemüller, T., Schiffrers, K.H., Georges, D., Dullinger, S., Eckhart, V.M., Edwards, T.C., Gravel, D., Kunstler, G., Merow, C., Moore, K., Piedallu, C., Vissault, S., Zimmermann, N.E., Zurell, D., & Schurr, F.M. (2014) Does probability of occurrence relate to population dynamics? *Ecography*, 1155–1166.
- Tyler-Smith, C., Yang, H., Landweber, L.F., Dunham, I., Knoppers, B.M., Donnelly, P., Mardis, E.R., Snyder, M., & McVean, G. (2015) Where Next for Genetics and Genomics? *PLOS Biology*, **13**, e1002216.
- Valladares, F., Matesanz, S., Guilhaumon, F., Araújo, M.B., Balaguer, L., Benito-Garzón, M., Cornwell, W., Gianoli, E., van Kleunen, M., Naya, D.E., Nicotra, A.B., Poorter, H., & Zavala, M.A. (2014) The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters*, **17**, 1351–1364.

- 486 van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt,
487 G.C., Kram, T., Krey, V., Lamarque, J.F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith,
488 S.J., & Rose, S.K. (2011) The representative concentration pathways: An overview. *Climatic*
489 *Change*, **109**, 5–31.
- 490 Williams, P. H., Humphries, C., Araújo, M. B., Lampinen, R., Hagemeyer, W., Gasc J.-P.
491 & Mitchell-Jones, T. (2000). Endemism and important areas for representing European
492 biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian*
493 *Journal of Entomology* 2:21–46
- 494 Yannic, G., Pellissier, L., Ortego, J., Lecomte, N., Couturier, S., Cuyler, C., Dussault, C.,
495 Hundertmark, K. J., Irvine, R. J., Jenkins, D.A., Kolpashikov, L., Mager, K., Musiani, M.,
496 Parker, K. L., Roed, K. H., Sipko, T., Pórisson, S. G., Weckworth, B. V., Guisan, A.,
497 Bernatchez, L. & Côté, S. D. (2014) Genetic diversity in caribou linked to past and future
498 climate change. *Nature Climate Change*, **4**(12), 132-137.

TABLES

TABLE 1

Changes in genetic cluster membership at the population level.

Columns - gen.unit: Genetic cluster, rcp: Representative concentration pathway (climate change scenario), n.gain: Number of populations for which models agree on an increase in genetic cluster membership value, n.loss: Number of populations for which models agree on a decrease in genetic cluster membership value, unknown: Number of populations for which models do not agree, avg.gain: Average gain for populations on which models agree, sd.gain: Standard deviation of gain for populations on which models agree, avg.loss: Average loss for populations on which models agree, sd.loss: Standard deviation of loss for populations on which models agree

gen.unit	rcp	n.gain	n.loss	unknown	avg.gain	sd.gain	avg.loss	sd.loss
GC1	+2.6	93	156	25	0.169	0.100	-0.264	0.127
GC1	+8.5	76	173	25	0.138	0.084	-0.335	0.160
GC2	+2.6	157	67	50	0.117	0.078	-0.136	0.098
GC2	+8.5	141	72	61	0.122	0.089	-0.148	0.109
GC3	+2.6	183	50	41	0.135	0.074	-0.456	0.273
GC3	+8.5	200	46	28	0.214	0.172	-0.434	0.268
GC4	+2.6	161	63	50	0.129	0.110	-0.162	0.124
GC4	+8.5	173	38	63	0.200	0.138	-0.190	0.107

TABLE 2

Changes in land suitability in the Iberian Peninsula for each genetic cluster and time frame. Potential area is expressed in km² after applying a cut-off threshold of 0.5. Mean and Median are the mean and median predicted cluster membership percentage divided by 100, respectively. Present time shows the current values for each Genetic cluster and Statistic, and RCP +2.6 and RCP +8.5 show these values for each climate change scenario.

Genetic cluster	Statistic	Present time	RCP +2.6	RCP +8.5
GC1	Potential area	101900	50173	12713
GC1	Mean	0.314 +/- 0.148	0.270 +/- 0.113	0.194 +/- 0.088
GC1	Median	0.290	0.224	0.162
GC2	Potential area	60870	48371	42885
GC2	Mean	0.208 +/- 0.167	0.204 +/- 0.154	0.198 +/- 0.146
GC2	Median	0.162	0.178	0.171
GC3	Potential area	1033	8630	50487
GC3	Mean	0.225 +/- 0.062	0.217 +/- 0.080	0.257 +/- 0.132
GC3	Median	0.251	0.231	0.244
GC4	Potential area	89036	156021	212014
GC4	Mean	0.253 +/- 0.168	0.302 +/- 0.180	0.369 +/- 0.188
GC4	Median	0.154	0.200	0.300

FIGURES

FIGURE 1

Gains and losses in membership percentage to genetic clusters. Black diamonds show the populations for which models do not agree. Degree of change is expressed in membership percentage units divided by 100.

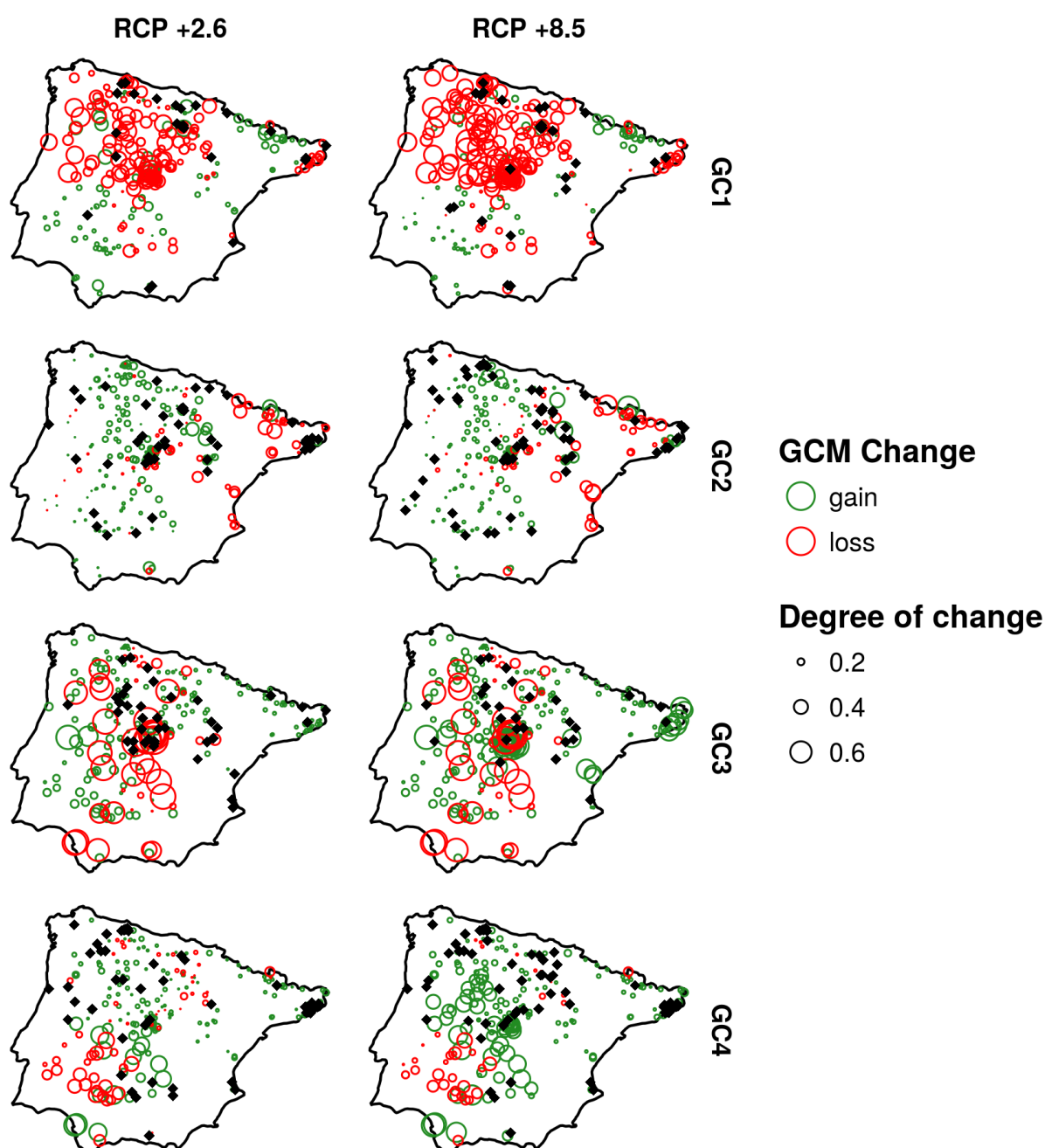


FIGURE 2

Genetic cluster turnover for populations under two scenarios of climate change in 2070. Numbers at bottom right of each map show the number of populations dominated per genetic cluster in each case. UNKN (unknown) is the number of populations for which models do not agree.

Genetic cluster ● GC1 ● GC2 ● GC3 ● GC4

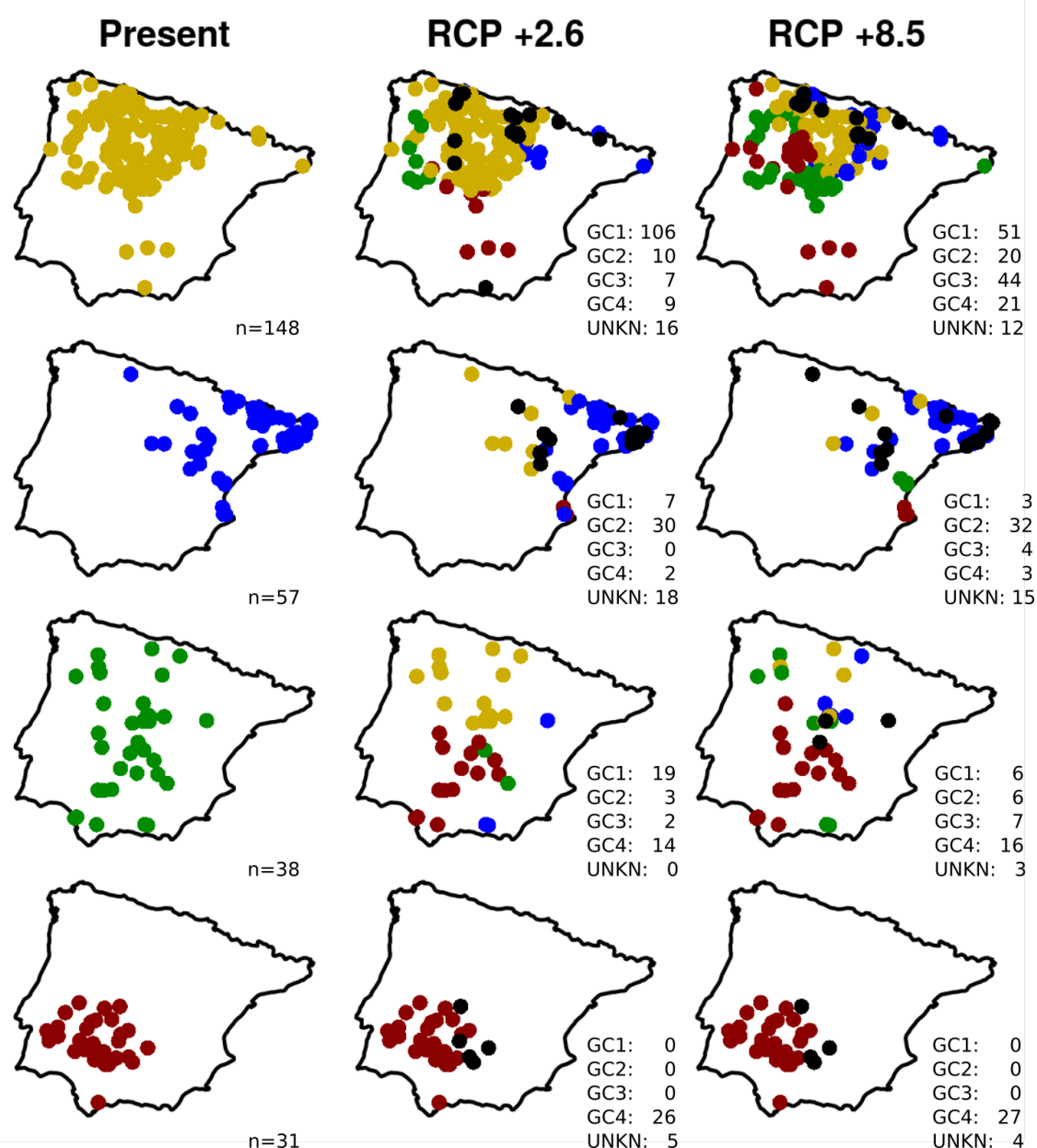


FIGURE 3

Projection of genetic cluster membership percentage under two scenarios of climate change (2070). Gray areas are zones environmentally similar to the environmental conditions of the populations for which models do not agree and thus cannot be predicted with confidence. Legends are expressed in membership percentage values divided by 100.

