Intraspecific niche models for the invasive ambrosia beetle Xylosandrus crassiusculus suggest contrasted responses to climate change

Teddy Urvois, Marie-Anne Auger-Rozenberg, Alain Roques, Carole Kerdelhué, J.-P. Rossi

Intraspecific niche models for the invasive ambrosia beetle *Xylosandrus crassiusculus* suggest contrasted responses to climate change

T. Urvois[1,2], M.-A. Auger-Rozenberg[1], A. Roques[1], C. Kerdelhué[2] & J.-P. Rossi[2]

[1] INRAE, URZF, 45075 Orléans, France

[2] UMR CBGP, INRAE, CIRAD, IRD, Institut Agro, Montpellier, France

Corresponding author: Jean-Pierre.Rossi@inrae.fr

Declaration of authorship

Author Contributions: CK, MAAR, JPR, and AR conceived and designed the study. TU collected the data. TU and JPR analyzed the data and made the figures. TU, JPR, and CK wrote the manuscript; other authors provided editorial advice.

Abstract

Xylosandrus crassiusculus is an invasive ambrosia beetle comprising two differentiated genetic lineages, named cluster 1 and cluster 2. These lineages invaded different parts of the world at different periods of time. We tested whether they exhibited different climatic niches using Schoener's D and Hellinger's I indices and modeled their current potential geographical ranges using the Maxent algorithm. The resulting models were projected according to future and recent past climate datasets for Europe and the Mediterranean region. The future projections were performed for the periods 2041-2070 and 2071-2100 using 3 SSPs and 5 GCMs. The genetic lineages exhibited different climate niches. Parts of Europe, the Americas, Sub-Saharan Africa, Asia, and Oceania were evaluated as suitable for cluster 1. Parts of Europe, South America, Central and South Africa, Asia, and Oceania were considered as suitable for cluster 2. Models projection under future climate scenarios indicated a decrease in climate suitability in Southern Europe and an increase in North Eastern Europe in 2071-2100. Most of Southern and Western Europe was evaluated as already suitable for both clusters in the early 20th century. Our results show that large climatically suitable regions still remain uncolonized and that climate change will affect the geographical distribution of climatically suitable areas. Climate conditions in Europe were favorable in the 20th century, suggesting that the recent colonization of Europe is rather due to an increase in propagule pressure via international trade than to recent environmental changes.

Keywords

Biological invasion, Species Distribution Modeling, Climate change, Genetic structure, Preparedness

Introduction

Biological invasions are a consequence of human population growth and the development of worldwide trade. Invasive species are responsible for considerable environmental and economic losses worldwide and the number of new invaders shows no sign of a decrease (Seebens et al. 2017). Global changes possibly facilitate new invasions by improving climatic suitability and exacerbating the impact of ongoing ones (Bradshaw et al. 2016). In that context, one important aspect of preparedness is the anticipation of areas at risk. Such anticipation can be reached using species distribution models (SDM) that assess species' potential range shifts or expansions under current or future climate conditions (Baquero et al. 2021; Rossi and Rasplus 2023).

In general, SDMs are calibrated at the species level, but this is increasingly debated because the resulting models are not always able to capture local adaptation (Pearman et al. 2010; Maguire et al. 2018; Banerjee et al. 2019; Smith et al. 2019). Subclade models are calibrated using datasets describing the geographical distribution of intraspecific lineages. Such data generally come from genetic analyses,

and the number of genotyped populations usually encompasses only a small subset of the occurrence records available in databases such as GBIF (Global Biodiversity Information Facility https://www.gbif.org/) in which information is mainly available at the species level. As a consequence, subclade models sometimes perform poorly, and deciding whether a species or a subclade-level model should be used is therefore a matter of data availability, model performance, or evidence of a niche divergence (Collart et al. 2020).

Xylosandrus crassiusculus is an ambrosia beetle native to Southeast Asia and invasive worldwide (Storer et al. 2017). During the last century, it reached most tropical and subtropical areas, as well as some countries in temperate regions. It was first detected in Madagascar more than a century ago (Schedl 1953). Later discovered in Hawaii in 1950 (Samuelson 1981) and North America in 1974 (Anderson 1974), it is now established in 31 states in the USA and one Canadian province. It was discovered in South America, specifically in Argentina, in 2001 (Kirkendall 2018) and in Australia in 2011 (Nahrung and Carnegie 2020). It reached Europe recently, as it was detected in Italy in 2003 (Pennachio et al. 2003), in France in 2014 (Roques et al. 2019), in Spain in 2016 (Gallego et al. 2017) and in Slovenia in 2017 (Kavčič 2018).

These recent detections have sparked considerable interest in the potential expansion of *X. crassiusculus* in Europe. In a first attempt at modeling the species potential distribution in Europe using species-level SDM, Urvois et al. (2021) failed and hypothesized that it could be due to the existence of differentiated genetic lineages exhibiting niche divergence. A preliminary study by Storer et al. (2017) suggested that *X. crassiusculus* is indeed divided into two differentiated subclades hereafter referred to as clusters, but these authors did not include specimens from Europe or South America. Interestingly, Ito and Kajimura (2009) documented a large genetic diversity in Japan and suggested that several subspecies could occur in this country. Unfortunately, because the two studies used different molecular markers and focused on different and hardly overlapping regions of the world, it was impossible to compare and synthesize the reported genetic structures. Using a comprehensive sampling and complementary molecular markers, Urvois et al. (2023) confirmed the existence of two genetic clusters that displayed different geographical distributions. They only co-occurred in Oahu Island in Hawaii, South Africa (Nel et al. 2020), Taiwan (Storer et al. 2017), Papua New Guinea, Australia (Tran et al. 2023), the Guangxi province in China, and Okinawa Island in Japan. Only cluster 2 was found in Europe. This study provided an occurrence dataset at the cluster level, which was lacking so far. As *X. crassiusculus* is a highly polyphagous species (Ranger et al. 2016), the availability of suitable hosts is probably not the main constraint upon its establishment or expansion but climate could be the decisive factor (Urvois et al. 2021).

The first goal of the present study was to assess the climatic niche differentiation between both clusters to evaluate whether they displayed different climatic preferences. Our second objective was to assess

their worldwide potential distribution according to a reference climate dataset (1979-2013), to identify (i) new areas where at least one cluster could establish, and (ii) areas where a geographical expansion is possible. Our third goal was to explore the effect of future climate change on both clusters' potential distributions in Europe. This would allow to identify the areas where cluster 2 could expand in the coming decades and, conversely, if some areas could become unsuitable (range shrink). In addition, it would allow us to assess whether suitability in Europe would increase for cluster 1, and identify potential areas where the species could establish. Finally, the fourth goal of the study was to test the hypothesis that European environmental conditions were unsuitable for *X. crassiusculus* in the past and that the continent became suitable only recently due to climate change, which could explain why *X. crassiusculus* invaded Europe only recently.

To achieve these goals, we compared the climatic niches of each genetic cluster and calibrated subclade SDM using the Maxent algorithm. A model was calibrated for each genetic cluster using reference climate conditions, and we projected the resulting models according to future climate scenarios in 2041-2070 and 2071-2100 and past climate conditions for each decade of the 20th century.

Methods

Data analysis and graphical displays were performed using the R Software v4.0.0 (R Core Team 2020).

Occurrence data

Records of *X. crassiusculus* specimens unambiguously assigned to each cluster were retrieved from Urvois et al. (2023) and Storer et al. (2017). We obtained 39 records for cluster 1 and 44 for cluster 2. We then removed the duplicated data by withdrawing all but one occurrence per pixel of the climate raster (see below), and obtained 38 and 44 records for cluster 1 and cluster 2, respectively (see Figure S1.1 in online resource 1 and online resources 2 and 3).

Environmental variables

We used the Chelsa dataset (version 1.2), which provides worldwide environmental layers with a 30 arc-second resolution (≈1 km² at the equator) for the past, reference (near current), and future climate conditions (Karger et al. 2017, 2020). The reference climate conditions corresponded to the period 1979-2013. We also used forecasts of future climate conditions for two periods, 2041-2070 and 2071-2100, and three Shared Socio-Economic Pathways (SSPs). SSPs represent different future scenarios named after their narrative (i.e. potential socio-economic development, ranging from 1 to 5) and radiative forcing value (i.e. change in energy flux, from 1.9 to 8.5 W/m²), ranging from the most optimistic SSP1-1.9 to the most pessimistic SSP5-8.5. The SSPs available from the Chelsa database for the period 2041-2070 and 2071-2100 were SSP1-2.6 (low greenhouse gas emission, estimated warming

in 2041-2060: 1.7 °C), SSP3-7.0 (high greenhouse gas emission, estimated warming in 2040-2060: 2.1 °C) and SSP5-8.5 (very high greenhouse gas emission, estimated warming in 2040-2060: 2.4 °C). For each SSP, we used five Global Circulation Models (GCMs). These are numerical models simulating the effect of changes in greenhouse gas concentrations on the climate and are named GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0, and UKESM1-0-LL. Past climate conditions were described for each decade from 1901-1910 to 1981-1990 based on the monthly data from CHELSAcruts (Karger and Zimmermann 2018) using the biovar function from the R package `dismo´ (Hijmans et al. 2017). We *a priori* selected seven climate descriptors assumed to be potential drivers of *X. crassiusculus*' distribution, to analyze both the relative niche of the clusters and their potential distributions. Six were associated with temperatures: the average temperatures of the warmest, coldest, wettest, and driest quarters (bio8, bio9, bio10, bio11, respectively), the maximum temperature of the warmest month (bio5) and the coldest temperature of the coldest month (bio6). The last variable was the precipitation seasonality (bio15) which can be considered as a proxy for water-related stress (Dannenberg et al. 2019), and thus increases tree susceptibility to *X. crassiusculus* (Ranger et al. 2015). *X. crassiusculus* spending most of its life buried in galleries, variables linked to monthly and quarterly precipitations were assumed to be less relevant and discarded.

Ecological niche divergence

We compared the realized niche of the two genetic clusters using the method proposed by Warren et al. (2008) implemented in the R package 'ecospat' (Broennimann et al. 2021). We computed Schoener's D (Schoener 1968) and Hellinger's based I indices (Guisan et al. 2017) and formally tested niche equivalency using 1000 randomizations. Two tests were run. The first one included all the climate descriptors involved in the model calibration and the second one was performed on the four bioclimatic variables that were selected in the course of the SDM calibration (see below): maximum temperature of the warmest month (bio5), mean temperature of the driest quarter (bio9), mean temperature of the coldest quarter (bio11) and the precipitation seasonality (bio15). The background environment used to calculate the clusters' niche overlap (i.e. all pixels of *X. crassiusculus*' distribution area) was characterized using six areas encompassing all occurrences used in this study (Figure S1.1 in online resource 1).

Species distribution modeling

Modeling

X. crassiusculus' invasion is still ongoing, and thus its geographical range is constantly changing. Such a non-equilibrium situation in the studied range makes it difficult to distinguish between locations where

153 the species is absent due to environmental conditions (i.e. true absences), and locations where it is
154 absent because of dispersal limitations. In such situations, presence-only algorithms are recommended
155 (Guisan et al. 2017). The Maxent algorithm uses presence data and background points (Phillips et al.
156 2006): the latter corresponds to randomly sampled points in the study area and provides information
157 about the environmental conditions across that area. It has been widely adopted in the last decade
158 because of its easy accessibility (Ahmed et al. 2015) and high-performance results (Elith et al. 2006).
159 We used the Maxent implementation available in the `MIAmaxent´ R package (Vollering et al. 2019),
160 which adapts Maxent with a different penalization method. Maxent penalizes the models' complexity
161 with lasso regularization, which keeps all predictors and transformations but shrinks their coefficients
162 to balance fit quality and model complexity. MIAmaxent, on the other hand, relies on subset selection
163 and performs a forward stepwise selection to either discard or retain variables and their
164 transformations. This leads to simpler models, which are therefore more easily interpretable, better
165 suited to small sample sizes, and more easily transferable to spatial or temporal projections (Elith et al.
166 2010; Moreno-Amat et al. 2015). This procedure additionally facilitates the management of collinearity
167 that may arise between environmental descriptors. MIAmaxent subset selection procedure only retains
168 two highly correlated variables when both account for a significant amount of variation.
169 Six transformation types were used on the environmental variables: linear, monotonous, deviation,
170 forward hinge, reverse hinge, and threshold. Because an infinity of transformations is possible for
171 spline-type transformations (i.e. forward hinge, reverse hinge, and threshold), the R package
172 MIAmaxent automatically identifies the ones that best explain the variation in the data and thus should
173 be involved in the selection procedure (see package documentation and Vollering et al. 2019). The
174 significance threshold in the subset selection was 0.05, and the best model for each cluster was
175 selected among the significant models ($p < 0.05$) based on the fraction of the deviance it explained
176 (D^2). The modeling approach was performed separately for each cluster. In each case, 10,000
177 background points were generated by randomly sampling locations in five and six areas encompassing
178 all occurrences of clusters 1 and 2, respectively (Figure S1.1 in online resource 1).
179
180 *Model evaluation*
181 Model performance was assessed using the Continuous Boyce Index (CBI) (Hirzel et al. 2006) and the
182 Area Under Curve (AUC). AUC typically relies on presence-absence data and is widely used (Ahmed et
183 al. 2015) even in presence-only modeling situations. AUC is provided here for comparison purposes
184 only. The CBI is a metric developed to evaluate presence-only models. This method involves dividing
185 the range of climatic suitability values into classes and calculating the frequency of occurrences falling
186 into each class (P) as well as the expected frequency of points falling into each class after random
187 reallocation (E). The CBI corresponds to the Spearman-ranked correlation between P/E and the

suitability classes (Hirzel et al. 2006). It ranges from -1 via 0 to 1, corresponding to counter-prediction, randomness, and perfect prediction, respectively. CBI was calculated using the `ecospat´ R package (Broennimann et al. 2021). Ideally, independent data points should be used to assess model quality, but these data are often not available. Here we had to compute the CBI using the dataset used for model calibration because we relied on a very limited number of genetically-characterized occurrences for each cluster. Following Hirzel et al. (2006), we identified climate suitability thresholds for the two clusters. These thresholds were used to reclassify climate suitability into three classes for cluster 1, namely unsuitable, marginal, and suitable climatic conditions, and four classes for cluster 2 (unsuitable, marginal, suitable, and optimal climatic conditions, see Results). The marginal areas correspond to climatic conditions nearing the species' tolerance limits and where the populations' growth rate is expected to be low. On the other hand, the optimal areas correspond to climatic conditions where the species can thrive. Finally, we computed the proportion of the area corresponding to each class worldwide and in the focal area including most of Europe and the Mediterranean region, between longitudes 20° W and 50° E and latitudes 27.5° N and 70° N.

Potential distribution under reference climate data (1979-2013)

The model was used to compute the climate suitability using the reference climate data (1979-2013). We reclassified these continuous values into 3 (respectively 4) discrete classes of suitability for cluster 1 (respectively 2). The classes were defined based on the thresholds resulting from the continuous Boyce analysis (see above).

Species level

We combined the areas considered as unsuitable for both clusters to derive a species-level projection and evaluated its accuracy using occurrences available from a previous survey for which no genetic information is available (Urvois et al. 2021).

Potential distribution under past and future climate conditions

The two models were projected in the focal area (parts of Europe and the Mediterranean) using the past and future climate datasets described in the *Environmental variables* section above.

Past. We built 11 maps for each cluster depicting the potential distribution for each decade from 1901-1910 to 1981-1990.

Future. We focused on 3 SSPs (SSP1-2.6, SSP3-7.0, and SSP5-8.5) projected in two periods (2041-2070 and 2071-2100) leading to six possible future situations. In each case, we computed the median of the projections associated with the five selected GCMs. The results are called consensus projections (Guisan et al. 2017). These maps were reclassified using the thresholds described above to produce

maps showing suitability classes for cluster 1 (unsuitable, marginal, and suitable) and cluster 2 (unsuitable, marginal, suitable, and optimal).

Results

Ecological niche comparison between clusters

Using the seven *a priori* selected variables, the ecological niche test indicated that the two clusters exhibited significantly different climatic niches ($p < 0.001$), with a Shoeners' D of 0.023 and a Hellingers' based I of 0.10 (Figure 1). The results obtained when running the same analyses with the four variables retained in the MIAmaxent models also showed a significant difference between the two clusters, with a Shoeners' D of 0.172 ($p = 0.040$) and a Hellingers' based I of 0.310 ($p < 0.01$) (Figure 1).

SDM under reference climate conditions

The raster source files (geotiff format) are available from Recherche Data Gouv at https://doi.org/10.57745/UB977K

cluster 1

MIAmaxent performed three selection rounds for cluster 1 and transformed the seven environmental variables into 62 transformed variables. The best model accounted for 6.8% of the null deviance ($p < 0.01$) and comprised the variables bio5, bio9, and bio15, accounting for 11.8%, 42.6%, and 45.6% of the total variation, respectively. The best model comprised four transformed variables, two of which included a deviation–type transformation with a parameter value of 1 and 2, bio9 and bio15, respectively. The variable bio9 had a threshold transformation for knot value of 11, and bio5 had a forward hinge transformation with a knot in the 15^{th} position. The AUC was 0.815, and the CBI was 0.9. The shape of the P/E curve allowed identifying two thresholds. The first threshold (th1 = 0.280) corresponded to the habitat suitability value for which P/E is lower than 1 (i.e. the model is predicting fewer occurrences than expected by chance) (see Figure S1.2 in online resource 1 for continuous Boyce index). The second threshold (th2 = 0.535) denoted the climate suitability for which the P/E value sharply increased. These thresholds were used to reclassify climate suitability into three categories corresponding to unsuitable (≤ th1), marginal (> th1 and ≤ th2), and suitable climatic conditions (> th2). The worldwide proportion of emerged lands corresponding to suitable, marginal, and unsuitable climate conditions were 10.6, 22.9, and 66.5% respectively when considering the extent of Figure 2 (A, B). High climate suitability was observed in Eastern and Western North America, South America (Brazil, Argentina), Southern Africa (South Africa, Botswana), Western Australia, and Southeastern China. Lower suitability, considered to reflect marginally suitable areas (Figure 2 A, B) was observed in Sub-Saharan Africa, Western Europe, Southeastern Asia, and South America from Guatemala to Brazil.

Around 16% of the focal area (parts of Europe and the Mediterranean area) were considered marginally suitable for *X. crassiusculus*' cluster 1 (Figure S1.3 and Table S1.1, in online resource 1). These marginal areas were mostly found in France, Spain, Italy, Greece, Turkey, and the Northern parts of Morocco, Algeria, and Tunisia (Figure 2 B; Figure S1.3 in online resource 1). The suitable areas represented 7% ~~of the focal area~~ and consisted of a few patches in Northwestern Spain, Brittany (France), the United Kingdom, Turkey, and the surroundings of the Azov Sea.

cluster 2

MIAmaxent performed four selection rounds for cluster 2 and transformed the seven environmental variables into 55 transformed variables. The best model accounted for 12.1% of the null deviance ($p < 0.0001$) and included variables bio5, bio11, and bio15, accounting for 28.1%, 37.2%, and 34.7% of the total variation, respectively. It comprised six transformed variables, three of which included a deviation–type transformation with a parameter value of 1, namely bio5, bio11, and bio15, and one (bio15) with a parameter of 5. The variable bio11 had a reverse hinge transformation with a knot in the 10^{th} position, and bio15 had a forward hinge transformation with a knot in the 7^{th} position. The AUC was 0.879, and the CBI was 0.901. The shape of the P/E curve allowed the identification of three thresholds. The first threshold (th1 = 0.144) corresponded to the habitat suitability value for which P/E is lower than 1 (Figure S1.2 in online resource 1). The second threshold (th2 = 0.424) denoted the climate suitability for which P/E increased over 1 (i.e. where the model started predicting more occurrences than expected by chance). The third threshold was placed where the habitat suitability value sharply increased (th3 = 0.720). These thresholds were used to reclassify climate suitability into four categories corresponding to unsuitable (≤ th1), marginal (> th1 and ≤ th2), suitable (> th2 and ≤ th3), and optimal climatic conditions (> th3). The category referred to as optimal was not found for cluster 1.

The worldwide proportion of emerged lands corresponding to optimal, suitable, marginal, and unsuitable climate conditions was 1.0, 3.9, 10.1, and 85.0% respectively (Figure 2 C, D). Optimal areas were found in the native range in China and Japan, and the invaded range in Argentina, Uruguay, South Africa, Southeastern and Southwestern Australia, and Europe. Marginal and suitable areas were distributed in China and Japan in the native area and Southwestern USA, Argentina, Central Africa, South Australia, and Europe in the invasive area (Figure 2 D).

Around 23% of the focal area was estimated to be at least marginally suitable (marginally suitable + suitable + optimal) for cluster 2 (Table S1.1 in online resource 1). Climate suitability was higher near the Mediterranean coast and generally decreased with increasing distance from the coast (Figure 2 C and Figure S1.3 in online resource 1). The optimal areas represented 2% of the surface and were distributed in Northwestern Spain, Northeastern Portugal, Southeastern France, Mediterranean coast from Almeria to Istanbul, and islands in the Eastern Mediterranean. Suitable areas represented 8% of

292 the focal area and were primarily found in Northern Spain, Central France, Italy, and the Balkans.
293 Marginally suitable areas were found in Southern Spain, from Northern France to the Netherlands, and
294 in the Balkans.
295
296 *Non-genetically assigned occurrences*
297 A total of 561 occurrences used in Urvois et al. (2021) were not genetically assigned to cluster 1 or
298 cluster 2 due to a lack of samples for genetic analyses. After removing duplicates, 420 such
299 observations remained worldwide. Among those, 65 (15.48%) fell into areas considered unsuitable for
300 both clusters (Figure 3). These records occurred on average ca. 56 km from the nearest suitable,
301 optimally or marginally suitable grid cells (see also Figure S1.4 in online resource 1 for details about the
302 distance separating these points from the nearest grid cell associated with suitable climate conditions).
303 Nearly fifty-seven percent of these points (56.9%) fell less than 25 km from the nearest suitable grid
304 cell.
305
306 *SDM projection under future conditions in Europe and the Mediterranean region*
307 *cluster 1*
308 Around 78 to 79% of the focal area were evaluated as unsuitable under SSP1-2.6 for the two time
309 periods. This value was 80% under SSP3-7.0 and SSP5-8.5 for 2041-2070 and 83-85% for 2071-2100
310 (Table S1.1, Figure 4, Figure S1.5 to Figure S1.8 in online resource 1).
311 Our results indicated a decrease in suitability in North Africa and Southern Spain and an increase in
312 suitability around the Netherlands and Northern Germany between the reference (1979-2013) and
313 future climate conditions (Figure 2 A, B, Figure 4, Figure S1.3 in online resource 1).
314 The projections obtained under SSP1-2.6, SSP3-7.0, and SSP5-8.5 showed similar situations with most
315 of Western Europe, Italy, Greece, and Turkey evaluated as at least marginally suitable in 2041-2070.
316 The suitability increased in northern Europe in 2071-2100 for SSP3-3.7 and SSP5-8.5 (Figure 4, Figure
317 S1.5-S1.8 in online resource 1).
318 *cluster 2*
319 The proportion of unsuitable areas decreased between the reference period and all future climate
320 conditions tested for cluster 2: it was 77.1% in 1979-2013 and ranged from 71.7% to 75.6% for the
321 future projections tested (Table S1.1 in online resource 1). The corresponding maps indicated a
322 decrease of suitability in the South of *X. crassiusculus*' potential distribution, and a range shift towards
323 Northern Europe, with suitable areas reaching Uppsala and Gävle (Sweden) for SSP5-8.5 for 2071-2100
324 (Figure 2 C, D, Figure 4 and Figure S1.5 to S1.8 in online resource 1). Optimal areas were mainly located
325 in Northern Spain and Western France in the six future conditions tested.
326

SDM under past conditions in Europe and the Mediterranean region

cluster 1

Around 25% of the focal area was projected to have been at least marginally suitable (i.e. suitable + marginally suitable surfaces) in 1901-1910, and this value ranged from 21.8 to 28.7% in the following decades (Table S1.1, Figure 5, Figures S1.9 to S1.11 in online resource 1). Despite variations between decades, the geographical distribution of suitable and marginally suitable areas corresponded to what is observed for the reference climate conditions (1979-2013, Figure 2 and Figures S1.3 in online resource 1).

cluster 2

The sum of marginally suitable, suitable, and optimal surfaces ranged from 18.8 to 24.3% of emerged lands in the focal area in the 20th century (Table S1.1, Figures S1.9 to S1.11 in online resource 1). As for cluster 1, the pattern of climate suitability observed during the 20th century did not differ strongly from what is observed for the reference climate conditions (1979-2013, Figure 2 and Figures S1.3 in online resource 1).

Discussion

Two genetic lineages with diverging climate niches

X. crassiusculus includes two highly divergent genetic clusters (Storer et al. 2017; Urvois et al. 2023). Even though genetic differentiation does not necessarily translate into ecological divergence (e.g. Andersen et al. (2012) showed complete overlap in resource utilization in two deeply diverging haplotypes in *X. morigerus*), our results revealed divergent, albeit partially overlapping, climatic niches between the lineages of *X. crassiusculus*. This is in line with the initial hypothesis stating that the clusters have different climatic requirements, and in addition to the genetic divergences, this suggests that *X. crassiusculus* may be composed of cryptic species rather than genetic lineages, although no morphological differences were observed (A. Cognato, comm. pers.). The evolution of secondary sexual characters – commonly used as morphological characters to distinguish Scolytinae species – is mainly driven by sexual selection. Thus, in species where siblings mate before dispersal, morphological evolution is expected to slow down due to a lack of sexual selection (Jordal et al. 2002), although the lack of outcrossing could help mutation fixation and thus morphological differences. Therefore, resolving the taxonomic status of the clusters will involve a revision of the whole taxonomic group (i.e., including all former species currently synonymized with *X. crassiusculus*) using molecular, morphological, ecological, and distribution data. In that context, our results constitute a first step in illustrating the existence of some ecological divergences between the genetically-identified clusters of *X. crassiusculus*.

Both clusters could invade new areas and widen their distribution in already-established areas

As we performed the SDM analyses at the lineage scale, we relied only on a limited number of occurrences for which genetic assignments were available. The resulting models performed well with high CBI and AUC, although the limited number of occurrences prevented model evaluation based on independent datasets. The area where cluster 1 could establish is evaluated as twofold the area suitable for cluster 2, the former including most of the latter. Regarding *X. crassiusculus*' native area, the models indicated suitable areas for both clusters in South-eastern China, Southeast Asian islands, and the Japanese archipelago. Concerning its invasive range, the areas estimated to be suitable for both species included Equatorial and South Africa, Southern and Western Australia, parts of the United States of America and South America, and Europe. A recent study actually showed that the two clusters are now indeed present in Australia (Tran et al. 2023). Urvois et al. (2023) showed that clusters 1 and 2 were mostly allopatric and co-occurred only in a few areas in the native and invasive ranges. Thus, the areas named above are at risk of being invaded by a second cluster, which could affect the invasion dynamic, host range, and damage, and could be treated as a new invasion.

The models also pointed towards areas where cluster 1 could establish but not cluster 2, and reciprocally. The areas optimal for cluster 2 were mainly in Australia, Japan, China, South Africa, Argentina, Uruguay, and Europe. Bark and ambrosia beetles are known to be easily transported and can be accidentally transferred over long distances as hitchhikers on traded plants (Raffa et al. 2015). As a consequence, dispersal is probably not a long-term limiting factor for this species, contrary to what was hypothesized in other cases (Monsimet et al. 2020) and all climatically suitable areas listed above are in fact at risk of invasion.

Our work has also allowed us to identify regions of the world where *X. crassiusculus* is unlikely to establish. As for *X. compactus* (Urvois et al. 2021), they correspond to either too cold or very hot regions, such as most of Northern North America, the highest regions of the Andean Mountains, hot and cold desert climates from Africa and the Middle East, desert, semi-desert and tropical regions of Australia, India, and most Northern Eurasia. Some occurrences from Urvois et al. (2021) for which genetic assignments were unavailable were found in areas estimated as unsuitable for both clusters. In addition, specimens from cluster 2 were identified in one locality in the Limpopo province in South Africa (Nel et al. 2020) not used in the models (because the dataset used for the present study was assembled earlier), which was considered unsuitable for cluster 2. Most of these occurrences were found in Central and Northeastern USA in the invaded area, while they were mainly found in Korea and Japan in the native area. Except for a few records in North America and South Korea, most occurrences falling in unsuitable areas fell within 25 km of the nearest climatically suitable areas. These mismatches could occur if the occurrences in unsuitable areas correspond to non-established or sink populations (Araújo and Peterson 2012) although this should not be the case since the occurrences were filtered to

keep only established populations of *X. crassiusculus,* according to the literature. Mismatches could also result from the existence of other genetic groups within *X. crassiusculus* with divergent ecological niches; indeed, Urvois et al (2023) showed that some individuals sampled in Japan could not be assigned to cluster 1 nor cluster 2, suggesting the existence of cryptic lineages that remain to be investigated. They could also correspond to misidentified specimens with similar morphology, such as *X. declivigranulatus,* previously synonymized as *X. crassiusculus* but recently "resurrected" by Smith et al. (2022). In the future, it will be crucial to increase the number of samples with genetic characterization and to include regions at the margins of the distribution ranges. Indeed, populations with local adaptations to certain climate conditions, for instance, at the margin of the native area, might have been overlooked and are known to potentially play an important role in invasion processes (Rey et al. 2012). From a methodological point of view, it would also be interesting to have a greater number of occurrence points and to better know the contours of the native range of each cluster in order to improve the performance of the models and possibly use algorithms requiring pseudo-absences (e.g. GLM or boosted regression trees).

Cluster 1 could invade Europe and cluster 2 could widen its distribution in Europe

Only cluster 2 has established in Europe yet, but the results indicate suitable climate conditions for cluster 1 in ca. 23% of the focal area, including Western France, most of Southwestern Europe, and a large part of the Mediterranean coast. Interestingly, suitable and marginal areas were distributed differently for the two clusters. Marginal areas were defined as nearing the clusters' tolerance limits and hence correspond to regions where the populations' growth rate could be limited and where populations are expected to be more susceptible to stochastic demographic processes. For cluster 2, suitable and optimal areas were mostly grouped and surrounded by marginal areas, while for cluster 1, the suitable areas consisted of small separated patches in a large marginal area. This suggests that the establishment of cluster 1 would be more difficult and that, once established, its dispersion would be more constrained and subject to demographic stochasticity. This could explain why cluster 1 is not yet present in Europe, even though some regions are climatically suitable.

The total suitable area for cluster 1 is expected to decrease between the reference climate and future climate projections, while it is expected to increase for cluster 2. The future model projections for the three SSPs showed a decrease in suitability in Southern Europe for both clusters and an increase towards the Netherlands and Denmark for cluster 1, and towards Northern Poland and Estonia for cluster 2. This northward shift would include some of the busiest ports in Europe, such as Amsterdam and Rotterdam, which are expected to be surrounded by suitable areas in the future for both clusters (SSP3-7.0 and SSP5-8.5). Climate suitability will also increase in Hamburg and Antwerp for both clusters. This pattern should increase the probability that *X. crassiusculus* enters Europe near climatically

suitable areas, hence the risk of new establishments in Northern Europe, possibly from different sources in the native area or in already invaded regions. These projected changes in the distribution of suitable areas in Europe could furthermore facilitate intra-continental dispersion.

Europe was already suitable for both clusters at the beginning of the 20th century

Cluster 2 invaded Africa, the Pacific Islands and North America between the beginning of the 20th century and 1974, while Europe was only recently colonized (between 2003 and 2019), probably with several independent introduction events (Urvois et al. 2023). This could be due to the fact that climate in Europe was not suitable during the 20th century, but became suitable and was actually invaded due to recent climate change. Our results did not support this hypothesis. On the contrary, the models showed that parts of the focal area were already suitable for both clusters during the 20th century, suggesting that both clusters could have established in Europe earlier. Thus, the late invasion of cluster 2 and lack of invasion of cluster 1 could result from (i) dispersal limitations or reduced propagule pressure, (ii) local climatic mismatches as parts of Europe associated with unsuitable climate conditions could have received specimens that failed to establish populations, (iii) stochastic processes leading to the extinction of the introduced specimens before reproduction, or a combination of the three. Indeed, *X. crassiusculus*' biology is known to limit the effects of mate-finding Allee effect (Gascoigne et al. 2009) and inbreeding depression (Peer and Taborsky 2005) but does not prevent environmental or demographic stochasticity. It is known that some invasion can take years before being detected so the invasion of *X. crassiusculus* in Europe could in fact be older than 2003. However, our models show that Europe had been suitable for decades, which probably predates *X. crassiusculus*' invasion, even if we account for this delayed detection.

Numerous studies showed that climate change is expected to have diverse effects depending on taxonomic groups, shrinking or expanding their potential distribution and favoring range shifts (Bellard et al. 2013; Pureswaran et al. 2018; Rahimi et al. 2021). Our results showed that climate change is expected to significantly affect *X. crassiusculus'* distribution in the future but is unlikely to have played a role in its late invasion of Europe. Using species distribution modeling with past climate data is a promising approach to decipher the impact of climate change on biological invasions' success, and to explicitly question whether current climate change has promoted recent invasions as frequently hypothesized.

Acknowledgments The authors would like to thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

References

Ahmed SE, McInerny G, O'Hara K, et al (2015) Scientists and software - surveying the species distribution modelling community. Divers Distrib 21:258–267

Andersen HF, Jordal BH, Kambestad M, Kirkendall LR (2012) Improbable but true: the invasive inbreeding ambrosia beetle *Xylosandrus morigerus* has generalist genotypes. Ecol Evol 2:247–57

Anderson DM (1974) First record of *Xyleborus semiopacus* in the continental United States (Coleoptera, Scolytidae). Cooperative Economic Insect Report 24:863–864

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. Ecology 93:1527–39

Banerjee AK, Mukherjee A, Guo W, et al (2019) Combining ecological niche modeling with genetic lineage information to predict potential distribution of *Mikania micrantha* Kunth in South and Southeast Asia under predicted climate change. Global Ecol Conserv 20:e00800

Baquero RA, Barbosa AM, Ayllón D, et al (2021) Potential distributions of invasive vertebrates in the Iberian Peninsula under projected changes in climate extreme events. Divers Distrib 27:2262–2276

Bellard C, Thuiller W, Leroy B, et al (2013) Will climate change promote future invasions? Global Change Biol 19:3740–8

Bradshaw CJ, Leroy B, Bellard C, et al (2016) Massive yet grossly underestimated global costs of invasive insects. Nat Commun 7:12986

Broennimann O, Di Cola V, Guisan A (2021) ecospat: Spatial Ecology Miscellaneous Methods. R package version 3.2.

Collart F, Hedenäs L, Broennimann O, et al (2020) Intraspecific differentiation: Implications for niche and distribution modelling. J Biogeogr 48:415–426

Dannenberg MP, Wise EK, Smith WK (2019) Reduced tree growth in the semiarid United States due to asymmetric responses to intensifying precipitation extremes. Sci Adv 5:eaaw0667

Elith J, H. Graham C, P. Anderson R, et al (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. Methods Ecol Evol 1:330–342

Gallego D, Lencina JL, Mas H, et al (2017) First record of the granulate ambrosia beetle, *Xylosandrus crassiusculus* (Coleoptera: Curculionidae, Scolytinae), in the Iberian Peninsula. Zootaxa 4273:431–434

Gascoigne J, Berec L, Gregory S, Courchamp F (2009) Dangerously few liaisons: a review of mate-finding Allee effects. Popul Ecol 51:355–372

Guisan A, Thuiller W, Zimmermann NE (2017) Habitat suitability and distribution models with applications in R. Cambridge University Press, Cambridge.

Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) dismo: Species Distribution Modeling. R package version 1.1-4

Hirzel AH, Le Lay G, Helfer V, et al (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecol Model 199:142–152

Ito M, Kajimura H (2009) Phylogeography of an ambrosia beetle, *Xylosandrus crassiusculus* (Motschulsky) (Coleoptera: Curculionidae: Scolytinae), in Japan. Appl Entomol Zool 44:549–559

Jordal BH, Normark BB, Farrell BD, Kirkendall LR (2002) Extraordinary haplotype diversity in haplodiploid inbreeders: phylogenetics and evolution of the bark beetle genus Coccotrypes. Mol Phylogenet Evol 23:171–188

Karger DN, Conrad O, Bohner J, et al (2017) Climatologies at high resolution for the earth's land surface areas. Sci Data 4:170122

Karger DN, Schmatz DR, Dettling G, Zimmermann NE (2020) High-resolution monthly precipitation and temperature time series from 2006 to 2100. Sci Data 7:248

Karger DN, Zimmermann (2018) CHELSAcruts - High resolution temperature and precipitation timeseries for the 20th century and beyond. EnviDat. doi:10.16904/envidat.159.

Kavčič A (2018) First record of the Asian ambrosia beetle, *Xylosandrus crassiusculus* (Motschulsky) (Coleoptera: Curculionidae, Scolytinae), in Slovenia. Zootaxa 4483:191–193

Kirkendall LR (2018) Invasive bark beetles (Coleoptera, Curculionidae, Scolytinae) in Chile and Argentina, including two species new for South America, and the correct identity of the Orthotomicus species in Chile and Argentina. Diversity 10:40

Maguire KC, Shinneman DJ, Potter KM, Hipkins VD (2018) Intraspecific niche models for Ponderosa Pine (*Pinus ponderosa*) suggest potential variability in population-level response to climate change. Syst Biol 67:965–978

Monsimet J, Devineau O, Petillon J, Lafage D (2020) Explicit integration of dispersal-related metrics improves predictions of SDM in predatory arthropods. Sci Rep 10:16668

Moreno-Amat E, Mateo RG, Nieto-Lugilde D, et al (2015) Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. Ecol Model 312:308–317

Nahrung HF, Carnegie AJ (2020) Non-native Forest Insects and Pathogens in Australia: Establishment, Spread, and Impact. Front For Glob Change 3:37. https://doi.org/10.3389/ffgc.2020.00037

Nel WJ, De Beer ZW, Wingfield MJ, Duong TA (2020) The granulate ambrosia beetle, *Xylosandrus crassiusculus* (Coleoptera: Curculionidae, Scolytinae), and its fungal symbiont found in South Africa. Zootaxa 4838(3):zootaxa.4838.3.7.

Pearman PB, D'Amen M, Graham CH, et al (2010) Within-taxon niche structure: niche conservatism, divergence and predicted effects of climate change. Ecography 33:990–1003

537 Peer K, Taborsky M (2005) Outbreeding depression, but no inbreeding depression in haplodiploid
538 ambrosia beetles with regular sibling mating. Evolution 59:317–323
539 Pennachio F, Roversi PF, Francardi V, Gatti E (2003) *Xylosandrus crassiusculus* (Motschulsky) a bark
540 beetle new to Europe (Coleoptera Scolytidae). Redia 86:77–80
541 Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic
542 distributions. Ecol Model 190:231–259
543 Pureswaran DS, Roques A, Battisti A (2018) Forest insects and climate change. Curr For Rep 4:35–50
544 R Core Team (2020) R: a language and environment for statistical computing. R Foundation for
545 Statistical Computing, Vienna, Austria
546 Raffa KF, Grégoire J-C, Lindgren BS (2015) Chapter 1 - Natural History and Ecology of Bark Beetles. In:
547 Vega FE, Hofstetter RW (eds) Bark Beetles. Academic Press, San Diego, pp 1–40
548 Rahimi E, Barghjelveh S, Dong P (2021) Estimating potential range shift of some wild bees in response
549 to climate change scenarios in northwestern regions of Iran. J Ecology Environ 45:14
550 Ranger CM, Reding ME, Schultz PB, et al (2016) Biology, ecology, and management of nonnative
551 ambrosia beetles (Coleoptera: Curculionidae: Scolytinae) in ornamental plant nurseries. J Integr Pest
552 Manage 7:1–23
553 Ranger CM, Schultz PB, Frank SD, et al (2015) Non-native ambrosia beetles as opportunistic exploiters
554 of living but weakened trees. PLoS One 10:e0131496
555 Rey O, Estoup A, Vonshak M, et al (2012) Where do adaptive shifts occur during invasion? A
556 multidisciplinary approach to unravelling cold adaptation in a tropical ant species invading the
557 Mediterranean area. Ecol Lett 15:1266–1275
558 Roques A, Bellanger R, Daubrée JB, et al (2019) Les scolytes exotiques : une menace pour le maquis.
559 Phytoma 727:16–20
560 Rossi JP, Rasplus JY (2023) Climate change and the potential distribution of the glassy-winged
561 sharpshooter (*Homalodisca vitripennis*), an insect vector of *Xylella fastidiosa*. Sci Total Environ
562 860:160375
563 Samuelson GA (1981) A synopsis of Hawaiian Xyleborini (Coleoptera: Scolytidae). Pac Insects 23:50–92
564 Schedl KE (1953) Fauna Madagascariensis - III. Mémoires de l'Institut Scientifique de Madagascar - Série
565 E - Tome III 67–106
566 Schoener TW (1968) Anolis lizards of Bimini: resource partitioning in a complex fauna. Ecology 49:704–
567 7226
568 Seebens H, Blackburn TM, Dyer EE, et al (2017) No saturation in the accumulation of alien species
569 worldwide. Nat Comm 8:14435
570 Smith AB, Godsoe W, Rodriguez-Sanchez F, et al (2019) Niche estimation above and below the species
571 level. TREE 34:260–273

572 Smith SM, Urvois T, Roques A, Cognato AI (2022) Recognition of the pseudocryptic species *Xylosandrus*
573 *declivigranulatus* (Schedl) as distinct from *Xylosandrus crassiusculus* (Motschulsky) (Coleoptera:
574 Curculionidae: Scolytinae: Xyleborini). Coleopt Bull 76: 367-374
575 Storer C, Payton A, McDaniel S, et al (2017) Cryptic genetic variation in an inbreeding and cosmopolitan
576 pest, *Xylosandrus crassiusculus*, revealed using ddRADseq. Ecol Evol 7:10974–10986
577 Tran HX, Doland Nichols J, Li D, et al (2023) Seasonal flight and genetic distinction among *Xylosandrus*
578 *crassiusculus* populations invasive in Australia. Australian Forestry 85:224–231.
579 Urvois T, Auger-Rozenberg M-A, Roques A, et al (2021) Climate change impact on the potential
580 geographical distribution of two invading Xylosandrus ambrosia beetles. Sci Rep 11:1339
581 Urvois T, Perrier C, Roques A, et al (2023) The worldwide invasion history of a pest ambrosia beetle
582 inferred using population genomics. Mol Ecol https://doi.org/10.1111/mec.16993
583 Vollering J, Halvorsen R, Mazzoni S (2019) The MIAmaxent R package: Variable transformation and
584 model selection for species distribution models. Ecol Evol 9:12051–12068
585 Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism:
586 quantitative approaches to niche evolution. Evolution 62:2868–2883
587
588

Figure legends

Figure 1. Bivariate plots of the realized ecological niche of cluster 1 (A, C) and cluster 2 (B, D). A, B: PCA based on the four climate descriptors after model the selection procedure (bio5, bio9, bio11 and bio15); C, D: PCA based on the seven climate descriptors used to calibrate the models (bio5, bio6, bio8, bio9, bio10, bio11 and bio15). The color gradient shows the density of the occurrences of the clusters. Solid lines indicate the 95% of the available (background) environment. For a given cluster, dashed line indicates the envelope of the occurrences of the other cluster.

Figure 2. Worldwide climate suitability for two genetic lineages of *Xylosandrus crassiusculus* according to reference climate conditions (1979-2013). A. Climate suitability for cluster 1. B. Reclassified climate suitability for cluster 1. C. Climate suitability for cluster 2. D. Reclassified climate suitability for cluster 2. Reclassifed maps are based on the continuous Boyce index approach (Figure S1.2 in online resource 1). Projection: EPSG 4326.

Figure 3: Map showing areas estimated as climatically unsuitable for both clusters 1 and 2 of *Xylosandrus crassiusculus*. Neither cluster 1 nor cluster 2 is expected to occur in these regions. Red crosses depict occurrence records from Urvois et al. (2021) falling into unsuitable areas. Projection: EPSG 4326.

Figure 4: Potential distribution of two genetic lineages of *Xylosandrus crassiusculus* in the period 2071-2100 according to the shared socio-economic pathways SSP1-2.6 and SSP-8.5 in parts of Europe and the Mediterranean region. Maps depict the consensus derived from the median of the model projected using five GCM for each SSP. A. Climate suitability for cluster 1 in 2071-2100 under SSP1-2.6. B. Climate suitability for cluster 2 in 2071-2100 under SSP1-2.6. C. Climate suitability for cluster 1 in 2071-2100 under SSP5-8.5. D. Climate suitability for cluster 2 in 2071-2100 under SSP5-8.5. Projection: EPSG 4326.

Figure 5. Potential distribution of two genetic lineages of *Xylosandrus crassiusculus* in the period 1941-1950 5 in parts of Europe and the Mediterranean region. A cluster 1. B cluster 2. Projection: EPSG 4326.

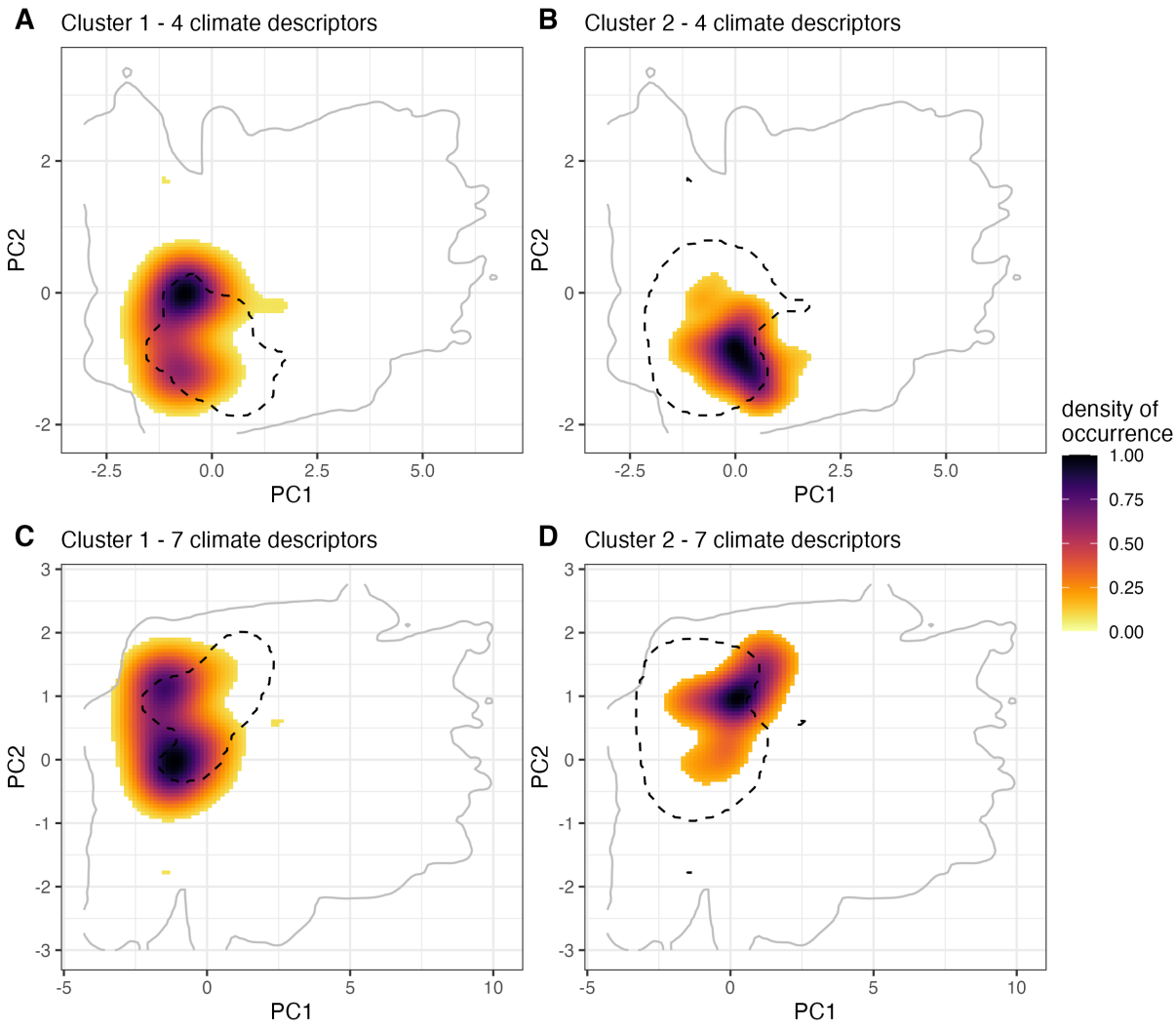

A
Cluster 1 - 4 climate descriptors
B
Cluster 2 - 4 climate descriptors
C
Cluster 1 - 7 climate descriptors
D
Cluster 2 - 7 climate descriptors
PC1
PC2
density of occurrence
1.00
0.75
0.50
0.25
0.00

Figure 1

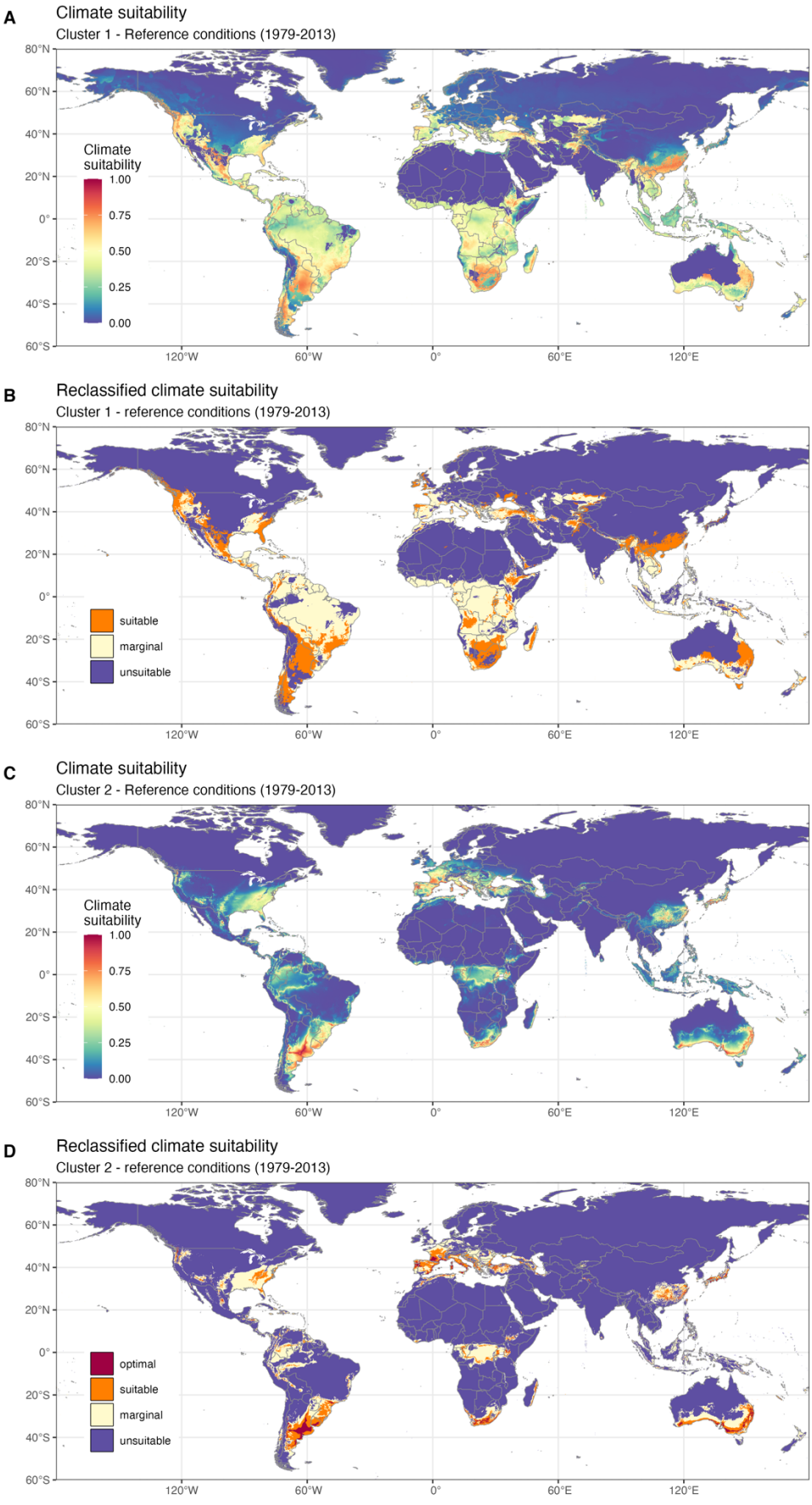
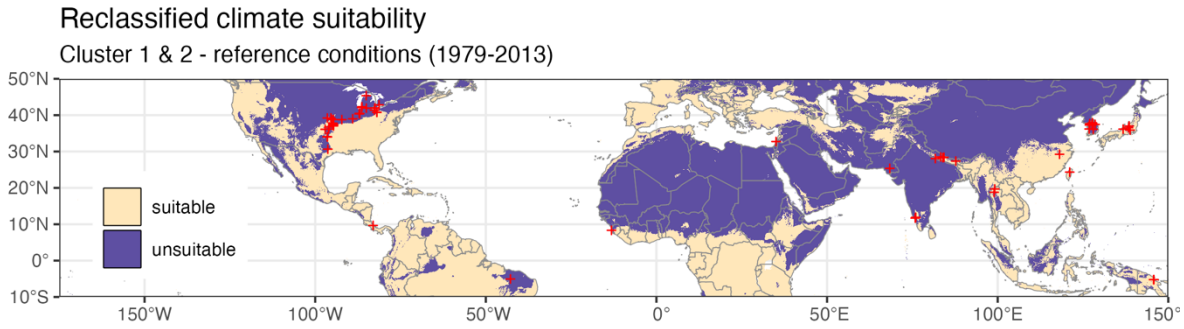

A
Climate suitability
Cluster 1 - Reference conditions (1979-2013)
Climate suitability
1.00
0.75
0.50
0.25
0.00
B
Reclassified climate suitability
Cluster 1 - reference conditions (1979-2013)
suitable
marginal
unsuitable
C
Climate suitability
Cluster 2 - Reference conditions (1979-2013)
Climate suitability
1.00
0.75
0.50
0.25
0.00
D
Reclassified climate suitability
Cluster 2 - reference conditions (1979-2013)
optimal
suitable
marginal
unsuitable
80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
120°W
60°W
0°
60°E
120°E

Figure 2



628

629 Figure 3

630

631

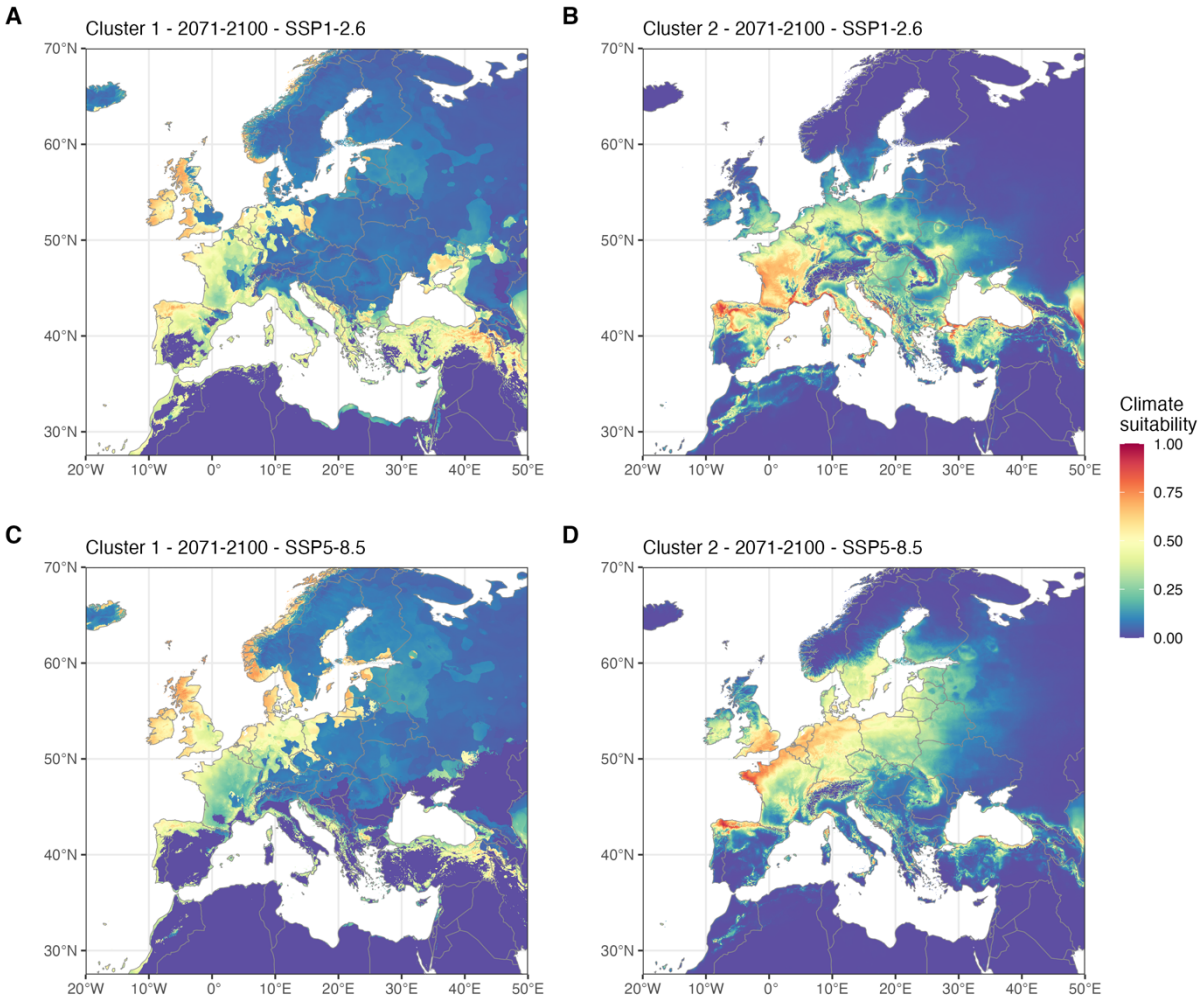
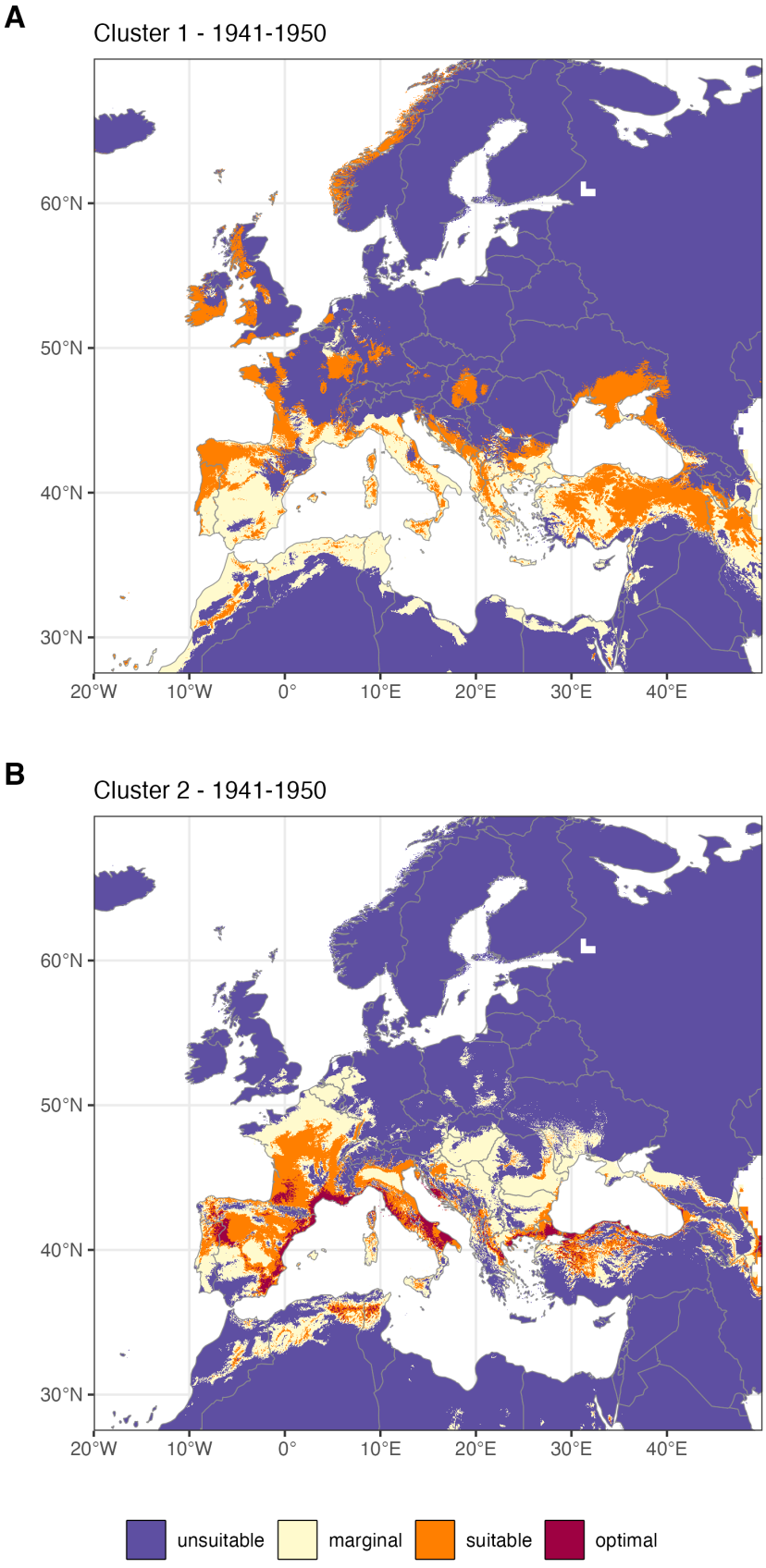

A
Cluster 1 - 2071-2100 - SSP1-2.6
B
Cluster 2 - 2071-2100 - SSP1-2.6
C
Cluster 1 - 2071-2100 - SSP5-8.5
D
Cluster 2 - 2071-2100 - SSP5-8.5
70°N
60°N
50°N
40°N
30°N
20°W
10°W
0°
10°E
20°E
30°E
40°E
50°E
Climate suitability
1.00
0.75
0.50
0.25
0.00

Figure 4



635

636 Figure 5